

**FLAVONOID PATHWAY GENE DISCOVERIES IN
Boesenbergia rotunda THROUGH RNA-SEQ
TRANSCRIPTOME PROFILING OF CELL SUSPENSION
CULTURES IN RESPONSE TO PHENYLALANINE**

NOOR DIYANA BINTI MD MUSTAFA

**FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR**

2017

**FLAVONOID PATHWAY GENE DISCOVERIES IN
Boesenbergia rotunda THROUGH RNA-SEQ
TRANSCRIPTOME PROFILING OF CELL SUSPENSION
CULTURES IN RESPONSE TO PHENYLALANINE**

NOOR DIYANA BINTI MD MUSTAFA

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

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

2017

**UNIVERSITY OF MALAYA
ORIGINAL LITERARY WORK DECLARATION**

Name of Candidate: NOOR DIYANA BINTI MD MUSTAFA

Registration/Matric No: SHC080079

Name of Degree: DOCTOR OF PHILOSOPHY

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):

FLAVONOID PATHWAY GENE DISCOVERIES IN *Boesenbergia rotunda*
THROUGH RNA-SEQ TRANSCRIPTOME PROFILING OF CELL SUSPENSION
CULTURES IN RESPONSE TO PHENYLALANINE

Field of Study: PLANT BIOTECHNOLOGY

I do solemnly and sincerely declare that:

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

Candidate's Signature

Date:

Subscribed and solemnly declared before,

Witness's Signature

Date:

Name: PROF. DR. ROFINA YASMIN BINTI OTHMAN

Designation:

FLAVONOID PATHWAY GENE DISCOVERIES IN *Boesenbergia rotunda* THROUGH RNA-SEQ TRANSCRIPTOME PROFILING OF CELL SUSPENSION CULTURES IN RESPONSE TO PHENYLALANINE

ABSTRACT

Panduratin A extracted from *Boesenbergia rotunda* is a flavonoid reported to possess a range of medicinal properties which include anti-dengue, anti-HIV, anti-cancer, antioxidant and anti-inflammatory activities. *B. rotunda* is a plant from the Zingiberaceae family commonly used as a food ingredient and traditional medicine in Southeast Asia and China. Reports on the health benefits of secondary metabolites extracted from *B. rotunda* over the last few years have increased demands for panduratin A. However, large scale extraction has been hindered by low abundance of the compound in nature and limited knowledge of its biosynthetic pathway. Experiments showed an increase in panduratin A production after 14 days post treatment with exogenous phenylalanine, an aromatic amino acid derived from the shikimic acid pathway. Transcriptome sequencing and digital gene expression (DGE) analysis of untreated and phenylalanine treated *B. rotunda* cell suspension cultures were carried out to elucidate the key genes differentially expressed in the panduratin A biosynthetic pathway. Total RNA of untreated and 14 days post-phenylalanine treated cell suspension cultures were extracted and sequenced using next generation sequencing technology employing an Illumina-Solexa platform. The transcriptome data generated 101,043 unigenes with 50,932 (50.41%) successfully annotated in the public protein databases; including 49.93% (50,447) in the non-redundant (NR) database, 34.63% (34,989) in Swiss-Prot, 24.07% (24,316) in Kyoto Encyclopedia of Genes and Genomes (KEGG) and 16.26% (16,426) in Clusters of Orthologous Groups (COG). Through DGE analysis, it was found that 14,644 unigenes were up-regulated and 14,379 unigenes down-regulated in response to exogenous phenylalanine treatment. In the flavonoid pathway leading to the proposed panduratin A

production, 2 unigenes encoded for phenylalanine ammonia-lyase (PAL), 3 for 4-coumaroyl:coenzyme A ligase (4CL) and 1 for chalcone synthase (CHS) were found up-regulated. In this thesis, a flavonoid pathway leading to panduratin A biosynthesis was proposed. In addition, two enzymes, namely flavonoid O-methyltransferase and prenyltransferase, were suggested to be key enzymes for panduratin A biosynthesis. In the transcriptome data, Unigene891_All, Unigene21507_All were predicted to encode flavonoid O-methyltransferase whereas, Unigene31983_All was predicted to encode prenyltransferase. At the gene regulatory level, transcripts for MYB transcription factors in the transcriptome database were further analysed by transcriptome-wide R2R3 MYB transcription factor analysis. As a result, transcripts for three MYB transcription factors namely *BrMYB1*, *BrMYB2* and *BrMYB3* were successfully sequenced and characterized. To date, this is the first report of *B. rotunda de novo* transcriptome data that could serve as a reference for gene or enzyme functional studies in the Zingiberaceae family. Although enzymes that are directly involved in the panduratin A biosynthetic pathway were not completely elucidated, the data provide an overall picture of gene regulation patterns leading to panduratin A production.

ABSTRAK

Panduratin A yang diekstrak daripada *Boesenbergia rotunda* adalah kompaun flavonoid yang dilaporkan mempunyai pelbagai faedah dari segi perubatan termasuk aktiviti anti-denggi, anti-HIV, anti-kanser, antioksidan dan anti-radang. *B. rotunda* adalah tumbuhan dari keluarga Zingiberaceae yang biasa digunakan sebagai makanan dan ubat-ubatan tradisional di Asia Tenggara dan China. Laporan mengenai manfaat kesihatan metabolit sekunder yang diekstrak daripada *B. rotunda* sejak beberapa tahun yang lalu telah menyebabkan permintaan yang semakin meningkat untuk panduratin A. Walau bagaimanapun pengeluaran berskala besar tidak berjaya dihasilkan kerana banyak faktor termasuk penghasilan panduratin A secara semulajadi yang sangat rendah dan pengetahuan laluan biosintesis panduratin A yang terhad. Berdasarkan kajian terdahulu, penghasilan panduratin A dapat ditingkatkan dengan penambahan fenilalanin di dalam ampaiian kultur selama 14 hari di mana fenilalanin adalah asid amino aromatik yang diperoleh melalui laluan asid shikimik. Penjujukan seluruh transkrip dan analisis ekspresi gen secara digital (DGE) daripada ampaiian kultur *B. rotunda* yang dirangsang dan tanpa rangsangan oleh fenilalanin telah dijalankan untuk mengenal pasti gen utama yang terlibat dalam penghasilan panduratin A. Seluruh RNA daripada ampaiian kultur yg tidak dirangsang dan dirangsang oleh fenilalanin selama 14 hari telah diekstrak dan penjujukan transkrip telah dilakukan melalui kaedah teknologi penjujukan generasi baru yang menggunakan platform Illumina-Solexa. Keseluruhannya, data transkrip telah menjana 101,043 unigen dengan 50,932 (50.41 %) berjaya dipadankan dalam pangkalan data protein awam; termasuk 49,93 % (50,447) dalam pangkalan data yang tidak bertindih (NR), 34.63 % (34,989), di Swiss-Prot , 24.07 % (24 ,316), dalam Ensiklopedia Kyoto daripada Gen dan Genom (KEGG) dan 16.26 % (16,426) dalam Kelompok Kumpulan Orthologous (COG). Melalui analisis DGE, didapati bahawa ekspresi 14,644 unigenes meningkat dan ekspresi 14,379 unigenes menurun disebabkan oleh rangsangan daripada

fenilalanin. Dalam laluan secara langsung flavonoid yang membawa kepada penghasilan panduratin A, didapati bahawa ekspresi 2 unigen yang merekodkan fenilalanin ammonia-lyase (PAL), 3 unigen untuk coumaroyl:koenzim A ligase (4CL) meningkat dan 1 unigen untuk chalcone sintase (CHS) meningkat. Dalam tesis ini, laluan flavonoid dalam penghasilan panduratin A telah dicadangkan. Selain itu, terdapat dua jenis enzim penting iaitu flavonoid O-methyltransferase dan prenyltransferase yang turut dicadangkan terlibat dalam penghasilan panduratin A. Daripada data transcriptom ini, Unigene891_All dan Unigene21507_All telah diramalkan berfungsi sebagai flavonoid O-methyltransferase. Manakala Unigene31983_All pula diramalkan merekodkan prenyltransferase. Di peringkat pengawalaturan gen, analisis yang lebih mendalam terhadap transkrip untuk MYB faktor transkripsi telah dilakukan melalui transkripsi mendalam (*transcriptome-wide*) R2R3 MYB faktor transkripsi analisis. Selain daripada itu, penjujukan transkrip untuk tiga MYB faktor transkripsi telah berjaya dilakukan, iaitu *BrMYB1*, *BrMYB2* and *BrMYB3*. Ini merupakan laporan pertama data transkrip *B. rotunda* yang boleh dijadikan sebagai rujukan untuk kajian masa depan yang berkenaan fungsi gen dan enzim dalam keluarga Zingiberaceae secara amnya. Walaupun enzim yang terlibat secara langsung dalam penghasilan panduratin A di laluan biosintesisnya tidak difahami sepenuhnya, data ini memberikan gambaran keseluruhan corak transkripsi gen yang terlibat secara tidak langsung ke arah penghasilan panduratin A.

ACKNOWLEDGEMENTS

Alhamdulillah praise to Allah SWT, the Most Merciful and the Most Gracious for giving me the strength mentally and physically to complete the degree of Doctor of Philosophy. I wish to express my deepest gratitude to Prof. Dr. Rofina Yasmin for giving me an opportunity to be her student, providing me well equipped laboratory facilities and making this research project possible by the collaboration with Beijing Genome Institute, Shenzhen, China. Thank you for your valuable guidance, supervision, suggestions and helps. The appreciation is also extended to Prof. Dr. Norzulaani and Dr. Yusmin for their constant encouragement, motivation, supervision, and guidance. In addition, I would like to thank National Science Fellowship (NSF) scholarship from Ministry of Science, Technology and Innovation Malaysia for funding my study.

Along the journey to complete this research project, I have met wonderful people, whom I shall remember in my life with fondness. A special thanks to all my friends in BGM1 lab, PBRL lab and Transcriptome lab – especially Kak Marina, Kak Maria, Kak Azma, Pei See, Sher Ming, Nadia, Akmar, Ain and Mariam for their full supports, motivation, suggestions and helps.

I am forever indebted to my parents, Md Mustafa bin Ab Samad and Noor Hayati binti Salim. Thank you for your endless supports financially and spiritually. To my mother, thank you for being there for me and keep me on track whenever I feel demotivated. I also wish to thank my husband Mohamed Bahtiar bin Kasim for his encouragement and sacrifices. To my lovely daughter Noor Khaliesa Zahra binti Mohamed Bahtiar, you are my strength to complete this study. Last but not least, I wish to extend my gratitude to my siblings: Noor Izni, Noor Qistina and Noor Ally Nadia. Thank you for your full supports.

TABLE OF CONTENTS

Abstract.....	iii
Abstrak.....	v
Acknowledgements.....	vii
Table of Contents.....	viii
List of Figures.....	xiv
List of Tables.....	xx
List of Symbols and Abbreviations.....	xxv
List of Appendices.....	xxxiii
CHAPTER 1: INTRODUCTION.....	1
1.1 Background.....	1
1.2 Objectives.....	4
1.3 Research Question.....	4
1.4 Hypothesis.....	4
CHAPTER 2: LITERATURE REVIEW.....	5
2.1 Introduction of <i>Boesenbergia rotunda</i>	5
2.2 Primary and secondary metabolites in plants.....	7
2.3 Plant phenolics.....	10
2.4 Plant flavonoids.....	14
2.5 Secondary metabolite pathways in plants.....	19
2.6 Secondary metabolites in <i>B. rotunda</i>	23
2.6.1 Bioactive compounds in <i>B. rotunda</i>	25
2.6.2 Panduratin A and 4-hydroxypanduratin A.....	30
2.7 Prenyltransferase.....	31
2.7.1 Prenyltransferase classification.....	31

2.7.2	Plant flavonoid prenyltransferases.....	34
2.7.3	Prenylated flavonoids in plants.....	40
2.7.4	Prenylated flavonoid in <i>B. rotunda</i>	43
2.8	Plant MYB transcription factors.....	45
2.8.1	MYB transcription factors classification in plants.....	45
2.8.2	MYB transcription factor functions in plants.....	49
2.8.3	MYB transcription factors in secondary metabolic pathway in plant.....	52
2.9	Transcriptome and RNA-Seq.....	59
2.10	Public protein databases.....	60
2.11	Applications of transcriptome profiling.....	62
CHAPTER 3: METHODOLOGY.....		67
3.1	Plant materials.....	67
3.1.1	<i>B. rotunda</i> plant.....	67
3.1.2	Establishment of <i>B. rotunda</i> cell suspension cultures.....	67
3.1.3	Phenylalanine treatment.....	68
3.2	General molecular technique.....	68
3.2.1	DNA extraction using CTAB method.....	68
3.2.2	RNA extraction from cell suspension cultures using CTAB method.....	70
3.2.3	RNA extraction from cell suspension cultures using Easy Spin RNA extraction kit method.....	71
3.2.4	Gel electrophoresis.....	72
3.2.5	Preparation of cDNA.....	73
3.2.6	Polymerase chain reaction (PCR).....	73
3.2.6.1	PCR reaction mixture.....	73
3.2.6.2	PCR parameters.....	74
3.2.6.3	Gel electrophoresis.....	74

3.2.7	Real-time or quantitative polymerase chain reaction (qPCR)....	74
3.3	Transcriptome profiling sequencing.....	75
3.3.1	Library preparation and sequencing.....	75
3.3.2	Transcript assembly and annotation.....	75
3.4	Transcriptome profiling analysis.....	77
3.4.1	Functional annotation.....	77
3.4.2	CDS prediction.....	77
3.4.3	Digital Gene Expression analysis for elucidating differentially expressed genes (DEGs).....	78
3.4.4	KEGG pathway enrichment analysis.....	80
3.4.5	Identification of genes in phenylpropanoid and flavonoid pathway.....	81
3.4.6	Analysis on genes involved in panduratin A proposed pathway...	81
3.4.6.1	Gene identification and analysis.....	81
3.4.6.2	Multiple sequence alignments.....	81
3.4.6.3	Phylogenetic tree analysis.....	82
3.4.7	Identification and classification of transcription factors and transcription regulators.....	82
3.4.8	Analysis on R2R3 MYB transcription factors.....	83
3.4.8.1	Identification and analysis of R2R3 MYB transcription factors.....	83
3.4.8.2	Multiple sequence alignments.....	83
3.4.8.3	Protein motif identification.....	84
3.4.8.4	Phylogenetic tree analysis.....	84
3.5	Validation and expression pattern analysis.....	84
3.5.1	Primer design.....	85
3.5.2	qPCR analysis.....	86
3.6	Sequencing of MYB transcription factor in <i>B. rotunda</i>	87
3.6.1	Primer design.....	87
3.6.2	Putative <i>BrMYB1</i> , <i>BrMYB2</i> and <i>BrMYB3</i> genes amplifications...	89

3.6.3	Gene cloning.....	89
3.6.3.1	Purification of PCR fragments.....	89
3.6.3.2	Ligation into pCR4-TOPO vector.....	90
3.6.3.3	Transformation One Shot TOP10 competent cells.....	91
3.6.3.4	Bacteria colony screening.....	91
3.6.3.5	Propagating the positive transformants.....	92
3.6.3.6	Plasmid Purification.....	92
3.6.3.7	Preparation of plasmid DNA for sequencing.....	93
3.6.4	Sequence analysis.....	93
3.7	Putative <i>BrMYB2</i> transcription factor in response to phenylalanine treatment.....	95
3.7.1	Primer design of reference genes for qPCR analysis.....	95
3.7.2	Primer design of putative <i>BrMYB2</i> gene primer design for qPCR analysis.....	96
3.7.3	qPCR analysis.....	97
CHAPTER 4: RESULTS.....		98
4.1	Establishment of <i>B. rotunda</i> cell suspension cultures.....	98
4.2	Total RNA assessment.....	99
4.3	Transcriptome profile analysis.....	100
4.3.1	RNA-Seq quality control.....	100
4.3.2	Short-read <i>de novo</i> sequencing and assembly.....	106
4.3.3	Functional annotation and gene ontology classification.....	108
4.3.4	Differentially expressed unigenes analysis.....	111
4.4	Pathway analysis.....	113
4.4.1	Representation of genes regulation in phenylpropanoid pathway and flavonoid pathway.....	117
4.4.2	Representation of genes regulation in proposed pathway.....	125
4.4.2.1	Flavonoid O-methyltransferase analysis.....	126

4.4.2.2	Prenyltransferase analysis.....	133
4.5	Experimental validation.....	138
4.6	Analysis of transcription factors and transcription regulators.....	139
4.7	R2R3 MYB transcription factor analysis.....	142
4.8	Molecular cloning and characterization of putative <i>BrMYB1</i> , <i>BrMYB2</i> and <i>BrMYB3</i> genes in <i>B. rotunda</i>	148
4.8.1	Genomic DNA extraction and total RNA extraction.....	148
4.8.2	PCR verification.....	149
4.8.3	Putative <i>BrMYB1</i> , <i>BrMYB2</i> and <i>BrMYB3</i> genes amplification....	149
4.8.4	Gel purification and plasmid prep.....	151
4.8.5	Sequence analysis of <i>BrMYB1</i> , <i>BrMYB2</i> and <i>BrMYB3</i> genes....	153
4.8.5.1	Putative <i>BrMYB1</i> sequence features.....	153
4.8.5.2	Putative <i>BrMYB2</i> sequence features.....	160
4.8.5.3	Putative <i>BrMYB3</i> sequence features.....	170
4.9	Gene expression study of <i>BrMYB2</i>	181
4.9.1	Total RNA extraction.....	182
4.9.2	qPCR analysis.....	183
CHAPTER 5: DISCUSSION.....		185
5.1	General transcriptome analysis.....	185
5.2	Transcription factors.....	189
5.3	Enzymes that are involved in the panduratin A pathway.....	203
5.4	Prenyltransferase.....	206
5.5	Flavonoid O-methyltransferase.....	208
5.6	Bioactive compounds proposed pathway.....	211
5.7	Proposed prenylation mechanism.....	215
5.8	Approaches to increase panduratin A production in the future.....	219

CHAPTER 6: CONCLUSION.....	224
References.....	226
List of Publications and Papers Presented.....	264
Appendix.....	265

University of Malaya

LIST OF FIGURES

Figure 2.1: <i>Boesenbergia rotunda</i> plant.....	6
Figure 2.2: Biosynthetic pathways of alkaloids, phenolics and terpenoids derived from primary metabolism (Großkinsky et al., 2012).....	8
Figure 2.3: The differences in dynamics between primary and secondary pathways.....	10
Figure 2.4: The relationship between shikimic acid pathway and biosynthesis of major plant phenolic groups adapted from Bowsher et al. (2008)..	13
Figure 2.5: a) Basic flavonoid skeleton structure showing its origins in the shikimate pathway and malonate pathways (Bowsher et al., 2008). b) The flavonoid numbering system of the carbon atoms on the three rings A, B and C (Kumar & Pandey, 2013). c) The chalcone numbering system of the carbon atoms on the ring A and C (Buckingham & Munasinghe, 2015).....	15
Figure 2.6: Overview of simplified pathways leading to major flavonoids classification: chalcones, aurones, flavanones, dihydrochalcones, dihydroxyflavonols, isoflavones, flavones, flavonols, phlobaphenes leucoanthocyanidins, anthocyanins and proanthocyanidins (Bowsher et al., 2008; Mierziak et al., 2014; Winkel-Shirley, 2001).....	17
Figure 2.7: Proposed biosynthetic pathway of gingerol and curcumin that are exclusively present in ginger and turmeric, denoted as dotted arrows (Koo et al., 2013).....	20
Figure 2.8: Established biosynthetic pathway of 8-dimethylallylnaringenin from <i>Sophora flavescens</i> (Sasaki et al., 2008; Yamamoto et al., 2000).....	21
Figure 2.9: Biosynthetic pathways of α -bitter acid, β -bitter acid and xanthohumol in <i>Humulus lupulus</i> (Li et al., 2015).....	22
Figure 2.10: Bioactive compounds extracted from <i>B. rotunda</i> which include flavanones such as pinostrobin and pinocembrin; and chalcones such as cardamonin, panduratin A, 4-hydroxypanduratin A and isopanduratin A.....	26
Figure 2.11: Coupling reaction of the shikimate/polyketide and isoprenoid pathways.....	32
Figure 2.12: The flavonoid prenyltransferases from the Leguminosae (Li et al., 2014).....	39

Figure 2.13: Phylogenetic tree relationship between plant aromatic prenyltransferases with homogentisate prenyltransferases generated by Wang et al. (2014).....	40
Figure 2.14: The prenyl side chain (Alhassan et al., 2014).....	41
Figure 2.15: Prenylation patterns encountered on flavonoids (Yang et al., 2015).	42
Figure 2.16: The structure of prenylated flavonoids found in <i>B. rotunda</i> (Morikawa et al., 2008; Win et al., 2007).....	45
Figure 2.17: Schematic diagram of plant MYB transcription factor classes.....	47
Figure 2.18: Regulation of flavonoid biosynthetic pathway by R2R3 MYB transcription factors subgroup 4, 6 and 7 in Arabidopsis (Dubos et al., 2010).....	53
Figure 3.1: Poisson distribution calculation to identify differentially expressed genes between control and phenylalanine treated samples.....	78
Figure 3.2: The probability formula to calculate the expressed gene equally between two samples.....	79
Figure 3.3: Reads per kb per Million reads (RPKM) equation to calculate the transcripts level.....	79
Figure 3.4: The probability formula to identify significantly enriched metabolic pathways or signal transduction pathways in DEGs comparing to the genome background.....	80
Figure 3.5: The position of a) putative <i>BrMYB1</i> , b) putative <i>BrMYB2</i> and c) putative <i>BrMYB3</i> primers relative to the Unigene14544_All, Unigene51701_All and Unigene19149_All, respectively.....	88
Figure 3.6: The position putative <i>BrMYB2</i> primers for qPCR analysis.....	97
Figure 4.1: Establishment of <i>B. rotunda</i> cell suspension cultures.....	98
Figure 4.2: Total RNA extraction of <i>B. rotunda</i> callus using modified CTAB method.....	99
Figure 4.3: Quality scores across all bases of RNA-Seq data from paired-end reads of control (a and b) and phenylalanine-treated (c and d) samples.....	101
Figure 4.4: Quality scores distribution over all sequences of RNA-Seq data from paired-end reads of control (a and b) and phenylalanine-treated (c and d) samples.....	102

Figure 4.5: Sequence content of all bases in RNA-Seq data from paired-end reads of control (a and b) and phenylalanine-treated (c and d) samples.....	103
Figure 4.6: N content across all bases in RNA-Seq data from paired-end reads of control (a and b) and phenylalanine-treated (c and d) samples....	104
Figure 4.7: Adaptor content in RNA-Seq data from paired-end reads of control (a and b) and phenylalanine-treated (c and d) samples.....	105
Figure 4.8: COG functional annotation of <i>B. rotunda</i> transcripts.....	109
Figure 4.9: Histogram presentation of unigene distributions in GO functional classification.....	110
Figure 4.10: RPKM calculations for Unigene51701_All from control (a) and phenylalanine treated (b) samples.....	112
Figure 4.11: Fold changes calculation of Unigene51701_All in phenylalanine treated sample compared to control.....	112
Figure 4.12: Expression level of unigenes that show significant differentially expressed in control (TC) and phenylalanine treated (TT) samples..	113
Figure 4.13: General flavonoid biosynthetic pathway (adapted from Bowsher et al., 2008).....	120
Figure 4.14: Proposed pathway leading to cardamonin, alpinetin, pinocembrin, pinostrobin, 4-hydroxypanduratin A and panduratin A production (adapted from Bowsher et al., 2008).....	125
Figure 4.15: Comparison of the deduced amino acid sequence of Unigene891_All and Unigene25107_All with other flavonoid O-methyltransferases from other plants such as <i>Oryza sativa</i> (OsNOMT), <i>Medicago sativa</i> (ChOMT), <i>Hordeum vulgare</i> subsp. <i>vulgare</i> (Fl-OMT) and <i>Musa acuminata</i> subsp. <i>malaccensis</i> (MaFl-OMT1-like_1, MaFl-OMT1-like_2 and MaFl-OMT1-like_3).....	131
Figure 4.16: Phylogenetic relationship of Unigene891_All and Unigene25107_All from <i>B. rotunda</i> with flavonoid O-methyltransferase (FOMT) proteins from other plants.....	132
Figure 4.17: Multiple sequence alignment of Unigene31983_All with other flavonoid prenyltransferases from other plants such as <i>Sophora flavescens</i> (SfLDT, SfN8DT1, SfN8DT2, SfN8DT3, SfFPT and SfG6DT), <i>Glycine max</i> (GmG4DT), <i>Lupinus albus</i> (LaPT1) and <i>Glycyrrhiza uralensis</i> (GuA6DT).....	136
Figure 4.18: Phylogenetic tree relationship between Unigene31983_All with plant aromatic and homogentisate prenyltransferases.....	138

Figure 4.19: Expression pattern validation of selected unigenes by qPCR.....	139
Figure 4.20: Multiple sequence alignment of 13 R2R3 MYB transcription factors ORF.....	144
Figure 4.21: R2 and R3 MYB repeats were highly conserved in selected 11 R2R3 MYB proteins in <i>B. rotunda</i>	145
Figure 4.22: Phylogenetic relationship of 13 R2R3 MYB transcription factor proteins from <i>B. rotunda</i> with other characterized plant R2R3 MYB proteins.....	146
Figure 4.23: DNA and RNA extraction of <i>B. rotunda</i> callus.....	148
Figure 4.24: PCR verification of total RNA extracted using putative <i>BrMYB1</i> partial gene primers.....	149
Figure 4.25: MYB open reading frame gene amplification of a) <i>BrMYB1</i> , b) <i>BrMYB2</i> and c) <i>BrMYB3</i>	151
Figure 4.26: Gel purification of <i>BrMYB1</i> , <i>BrMYB2</i> and <i>BrMYB3</i> fragments.....	152
Figure 4.27: Plasmid isolation for a) <i>BrMYB1</i> , b) <i>BrMYB2</i> and c) <i>BrMYB3</i> sequencing.....	152
Figure 4.28: Nucleotide sequencing alignment results of transcriptomic generated Unigene14544_All sequence, putative <i>BrMYB1</i> sequence from DNA template and putative <i>BrMYB1</i> sequence from cDNA template.....	155
Figure 4.29: Phylogenetic tree analysis of putative BrMYB1 protein with other R3-type MYB in plants.....	158
Figure 4.30: Putative BrMYB1 protein multiple sequence alignment with other MYB-related proteins from other plant species.....	159
Figure 4.31: Nucleotide sequencing alignment results of transcriptomic generated Unigene51701_All sequence, putative <i>BrMYB2</i> sequence from DNA template and putative <i>BrMYB2</i> sequence from cDNA template.....	163
Figure 4.32: Phylogenetic tree analysis of putative BrMYB2 protein with other C2 repressor R2R3 MYB proteins in plants.....	167
Figure 4.33: Putative BrMYB2 protein multiple sequence alignment with other C2 repressor R2R3 proteins from other plant species.....	168
Figure 4.34: Nucleotide sequencing alignment results of transcriptomic generated Unigene19149_All sequence, putative <i>BrMYB3</i> sequence from DNA template and putative <i>BrMYB3</i> sequence from cDNA template.....	173

Figure 4.35: Phylogenetic tree analysis of putative BrMYB3 protein with other R2R3 MYB proteins in plants that harbouring bHLH interaction motif which include proteins from subgroup 4, 5, 6 15 and 27.....	177
Figure 4.36: Putative BrMYB3 protein multiple sequence alignment with other R2R3 proteins from other plant species that harbouring bHLH interaction motif which include proteins from subgroup 4, 5, 6 15 and 27.....	178
Figure 4.37: Schematic diagram of <i>BrMYB2</i> and <i>BrMYB3</i> putative genes in <i>B. rotunda</i>	181
Figure 4.38: Total RNA extraction of control and phenylalanine treated <i>B. rotunda</i> callus after a) 1 day; b) 7 days and c) 14 days treatment....	182
Figure 4.39: Expression pattern of putative <i>BrMYB2</i> in <i>B. rotunda</i> cell suspension culture after 1, 7 and 14 days post treated with phenylalanine.....	184
Figure 5.1: Hypothesis on metabolic flux in the proposed flavonoid pathway leading to cardamonin, alpinetin, pinocembrin, pinostrobin, 4-hydroxypanduratin A and panduratin A production. Established flavonoid pathway was adapted from Bowsher et al., 2008.....	214
Figure 5.2: Proposed mechanism of prenylated chalcone by prenyltransferase through C- α (a) and C- β (b) prenylation.....	218
Figure 5.3: Suggestions on changing the metabolic flux in the proposed flavonoid pathway leading to increase panduratin A production. Established flavonoid pathway was adapted from Bowsher et al., 2008.....	221
Figure A.1: a) 1kb Plus DNA ladder and b) RiboRuler High Range RNA Ladder.....	280
Figure A.2: RNA assessment results using Agilent 2100 Bioanalyzer prior to sending RNA samples for sequencing of control and phenylalanine treated samples.....	282
Figure A.3: RNA assessment results from Beijing Genome Institute, China prior to whole transcriptome sequencing of a) control and b) phenylalanine treated samples.....	283
Figure A.4: GC distribution over all sequences in RNA-Seq data from paired-end reads of control (a and b) and phenylalanine-treated (c and d) samples.....	285
Figure A.5: Distribution of sequence length over all sequences in RNA-Seq data from paired-end reads of control (a and b) and phenylalanine-treated (c and d) samples.....	286

Figure A.6: Sequence duplication levels in RNA-Seq data from paired-end reads of control (a and b) and phenylalanine-treated (c and d) samples.....	287
Figure A.7: Kmer content in RNA-Seq data from paired-end reads of control (a and b) and phenylalanine-treated (c and d) samples.....	288
Figure A.8: KEGG phenylpropanoid pathway containing gene expression patterns.....	294
Figure A.9: KEGG flavonoid pathway containing gene expression patterns.....	295
Figure A.10: Gradient PCR between 45 – 55°C for a) <i>BrMYB1</i> , b) <i>BrMYB2</i> and c) <i>BrMYB3</i>	296
Figure A.11: PCR library colony screening of a) <i>BrMYB1</i> , b) <i>BrMYB2</i> and c) <i>BrMYB3</i> that amplified from DNA template.....	297
Figure A.12: PCR library colony screening of a) <i>BrMYB1</i> , b) <i>BrMYB2</i> and c) <i>BrMYB3</i> that amplified from cDNA template.....	298
Figure A.13: Illustration of pCR4-TOPO vector.....	299
Figure A.14: Dissociation curves of 9 selected unigenes for transcriptome data validation.....	334
Figure A.15: Dissociation curve of <i>EF1α</i> , <i>β-Tub</i> and <i>BrMYB2</i>	335
Figure A.16: Endogenous control profile of <i>EF1α</i> , <i>β-Tub</i> in control and phenylalanine treated callus after 0, 7 and 14 days treatment.....	335

LIST OF TABLES

Table 2.1: Examples of flavonoids based on the structure backbone of the flavonoid groups, such as flavones, flavonols, flavanones, flavanonols, isoflavones and flavan-3-ols (Kumar & Pandey, 2013).....	18
Table 2.2: Some secondary metabolites isolated from <i>B. rotunda</i>	23
Table 2.3: Summary of the presence of some metabolite compounds across the plant families and plant species (KNapSack Metabolite Information).....	24
Table 2.4: The summary of medicinal properties in <i>B. rotunda</i> which include anti-cancer, anti-dengue, anti-inflammatory, anti-HIV, anti-aging, antioxidant, antibacterial, anti-tumor and antimutagenic.....	27
Table 2.5: Plant flavonoid prenyltransferases in NCBI database.....	36
Table 2.6: Summary of molecular characterization of plant aromatic prenyltransferases.....	37
Table 2.7: Distribution MYB transcription factor subfamilies within selected plant species.....	49
Table 2.8: MYB transcription factors that are involved in the regulation of the phenylpropanoid pathway. This table was modified from Liu et al., 2015.....	54
Table 3.1: Forward and reverse primer sequences used to validate transcriptome data which consist of 4 random up-regulated unigenes, 5 flavonoid unigenes and one reference unigene (<i>EF1α</i>).....	86
Table 3.2: Forward and reverse primer sequences to amplify the putative <i>BrMYB1</i> , <i>BrMYB2</i> and <i>BrMYB3</i>	89
Table 3.3: Forward and reverse primer sequences for reference genes in qPCR analysis.....	96
Table 3.4: Forward and reverse primer sequences for <i>BrMYB2</i> gene in qPCR analysis.....	97
Table 4.1: Concentration and purity of total RNA from control and phenylalanine treated samples.....	99
Table 4.2: Summary of reads assembly generated by SOAPdenovo from control and phenylalanine treated <i>B. rotunda</i> callus.....	106
Table 4.3: Function annotation of <i>B. rotunda</i> transcriptome data in four public protein databases.....	108

Table 4.4: Summary of DEGs expression levels in <i>B. rotunda</i> transcriptome data.....	113
Table 4.5: Distributions of all unigenes and DEGs in KEGG database classification.....	114
Table 4.6: Summary of unigene distributions in KEGG pathways that has significant differential expression of genes.....	115
Table 4.7: Unigenes potentially related to panduratin A biosynthesis in phenylpropanoid pathway.....	118
Table 4.8: Unigenes potentially related to panduratin A biosynthesis in flavonoid pathway.....	119
Table 4.9: Gene regulation patterns in flavonoid pathway.....	123
Table 4.10: Summary of homology search using Protein blast of two unigenes annotated as flavonoid O-methyltransferase.....	128
Table 4.11: Summary of homology search using Protein blast of Unigene31983_All annotated as prenyltransferase.....	134
Table 4.12: Transcription factors (TFs) identified in <i>B. rotunda</i> by mapping to transcription factors in rice database using iTAK software.....	140
Table 4.13: Transcription regulators identified in <i>B. rotunda</i> using transcription regulators in rice database using iTAK software.....	141
Table 4.14: Summary of manually analysed R2R3 MYB transcription factor in <i>B. rotunda</i> transcriptome data, which included the total number of R2R3 MYB transcription factor annotation, unigene that had complete nucleotide sequence without gaps and the unigene that contained the domain region.....	143
Table 4.15: Summary of phylogenetic tree analysis of 13 <i>B. rotunda</i> R2R3 MYB proteins, which classify the R2R3 proteins into 7 subgroups with the predicted functions.....	147
Table 4.16: Concentration and purity of DNA and total RNA of <i>B. rotunda</i>	148
Table 4.17: The percentage identity of putative <i>BrMYB1</i> cDNA with other plant MYB transcription factor at the nucleotide level over the entire sequence using NCBI blastx.....	156
Table 4.18: The percentage identity of putative BrMYB1 protein sequence (ORF sequence) with other plant MYB transcription factor at the protein level over the entire sequence using NCBI blastp.....	157

Table 4.19: The percentage identity of putative <i>BrMYB2</i> cDNA with other plant MYB transcription factor at the nucleotide level over the entire sequence using NCBI blastx.....	165
Table 4.20: The percentage identity of putative BrMYB2 protein sequence (ORF sequence) with other plant MYB transcription factor at the protein level over the entire sequence using NCBI blastp.....	166
Table 4.21: The percentage identity of putative <i>BrMYB3</i> cDNA with other plant MYB transcription factor at the nucleotide level over the entire sequence using NCBI blastx.....	175
Table 4.22: The percentage identity of putative BrMYB3 protein sequence (ORF sequence) with other plant MYB transcription factor at the protein level over the entire sequence using NCBI blastp.....	176
Table 4.23: Concentration and purity of total RNA from control and phenylalanine treated samples after 1, 7 and 14 days post-phenylalanine treatment.....	182
Table 5.1: Proposed prenylation of prenylated flavonoids found in <i>B. rotunda</i> , which include the type of prenyl moiety, type of aromatic compound, C-prenylation location and involvement of further modification after prenylation.....	217
Table A.1: CTAB buffer composition in 200 ml solution.....	268
Table A.2: CTAB buffer composition in 200 ml solution.....	270
Table A.3: DNase I reaction mixtures in 13 μ l total reaction volume.....	270
Table A.4: PCR reaction mixtures in 100 μ l total reaction volume.....	271
Table A.5: PCR reaction mixtures in 100 μ l total reaction volume.....	272
Table A.6: Reverse Transcriptase PCR reaction mixtures in 20 μ l total reaction volume.....	272
Table A.7: qPCR reaction mixtures in 100 μ l total reaction volume.....	273
Table A.8: List of primers for <i>B. rotunda</i> MYB transcription factors sequencing	273
Table A.9: List of primers in qPCR study.....	274
Table A.10: TBE buffer composition in 1000 ml solution.....	279
Table A.11: Murashige and Skoog compositions in 1000 ml media preparation	281
Table A.12: The reports on control (TC) and phenylalanine treated (TT) samples from Beijing Genome Institute, China.....	284

Table A.13: Kmer sequences in RNA-Seq data from paired-end reads of control and phenylalanine-treated samples.....	289
Table A.14: Overrepresented sequences in RNA-Seq data from control file 2...	289
Table A.15: The length distribution of unigene of control sample, phenylalanine treated sample and All Unigene.....	290
Table A.16: The gap distribution of unigene of control sample, phenylalanine treated sample and All Unigene.....	290
Table A.17: The number of unigenes that were assigned in the COG functional categories.....	291
Table A.18: Unigene that assigned to GO-terms which classified under biological process, cellular component and molecular function.....	292
Table A.19: List of plant R2R3 MYB proteins with their respective NCBI accession numbers to construct phylogenetic tree.....	313
Table A.20: List of plant flavonoid O-methyltransferase proteins with their respective NCBI accession numbers to construct phylogenetic tree analysis.....	315
Table A.21: List of plant flavonoid O-methyltransferase proteins with their respective NCBI accession numbers to construct multiple sequence alignment.....	316
Table A.22: List of plant flavonoid prenyltransferase proteins with their respective NCBI accession numbers to construct phylogenetic tree...	316
Table A.23: List of plant flavonoid prenyltransferase proteins with their respective NCBI accession numbers to construct protein multiple sequence alignment.....	317
Table A.24: List of plant R3-type MYB proteins with their respective NCBI accession numbers that were used to construct phylogenetic tree for BrMYB1 analysis.....	317
Table A.25: List of plant R3-type MYB proteins with their respective NCBI accession numbers that were used to construct multiple sequence alignment for BrMYB1 analysis.....	318
Table A.26: List of plant subgroup 4 R2R3 MYB proteins with their respective NCBI accession numbers that were used to construct phylogenetic tree for BrMYB2 analysis.....	318
Table A.27: List of plant subgroup 4 R2R3 MYB proteins with their respective NCBI accession numbers that were used to construct multiple sequence alignment for BrMYB2 analysis.....	319

Table A.28: List of plant bHLH dependent R2R3-type MYB proteins with their respective NCBI accession numbers that were used to construct phylogenetic tree for BrMYB3 analysis.....	320
Table A.29: List of plant bHLH dependent R2R3-type MYB proteins with their respective NCBI accession numbers that were used to multiple sequence alignment for BrMYB3 analysis.....	321
Table A.30: Data generated from QuantStudio 12 K Flex software for random up-regulated unigenes with four technical replicates.....	331
Table A.31: Data generated from QuantStudio 12 K Flex software for selected phenylpropanoid and flavonoid unigenes with four technical replicates.....	332
Table A.32: Data generated from QuantStudio 12 K Flex software for <i>BrMYB2</i> gene with four technical replicates.....	333

University of Malaya

LIST OF SYMBOLS AND ABBREVIATIONS

18S	:	18 Svedberg
DXR	:	1-deoxy-D-xylulose-5-phosphate reductoisomerase
DXS	:	1-deoxy-D-xylulose-5-phosphate synthase
2,4-D	:	2,4-dichlorophenoxyacetic acid
28S	:	28 Svedberg
MCS	:	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase
MCT	:	2-C-methyl-D-erythritol 4-phosphate cytidyltransferase
HMGR	:	3-hydroxy-3-methylglutaryl-coenzyme A reductase
4CL	:	4-coumaroyl:CoA ligase
CMK	:	4-diphosphocytidyl-2-C-methyl-D-erythritol kinase
IDS	:	4-hydroxy-3-methylbut-2-enyl diphosphate reductase (isopentenyl/dimethylallyl diphosphate synthase)
HDS	:	4-hydroxy-3-methylbut-2-enyl diphosphate synthase
BAP	:	6 benzylaminopurine
AACT	:	acetyl-CoA acetyltransferase
ATPase	:	adenylpyrophosphatase
α	:	Alpha
NH ₄ NO ₃	:	ammonium nitrate
ANR	:	anthocyanidin reductase.
ANS	:	anthocyanidin synthase
bp	:	basepair
β	:	Beta
H ₃ BO ₃	:	boric acid
MCF-7	:	breast cancer cell

CCOMT	:	caffeoyl-CoA O-methyltransferase
CaCl ₂ .2H ₂ O	:	calcium chloride
CTAB	:	cetylmethylammonium bromide
CHI	:	chalcone isomerase
CHS	:	chalcone synthase
C4H	:	cinnamate-4-hydroxylase
CAD	:	cinnamyl alcohol dehydrogenase
COGs	:	Clusters of Orthologous Groups
CoCl ₂ .6H ₂ O	:	cobalt chloride
CDS	:	codon sequence
cDNA	:	complementary DNA
CuSO ₄ .5H ₂ O	:	copper sulphate
CS	:	cucurbitadienol synthase
Ct	:	cycle threshold
CAS	:	cycloartenol synthase
CYP450	:	cytochrome P450
P450	:	cytochrome P450
°C	:	degrees Celsius
Δ	:	delta
dNTP	:	deoxynucleotide
DNA	:	deoxyribonucleic acid
DEPC	:	diethyl pyrocarbonate
DEGs	:	differentially expressed genes
DGE	:	digital gene expression
DFR	:	dihydroflavonol 4-reductase
DKS1	:	diketide synthase 1

DMAPP	:	dimethylallyl diphosphate
MVD	:	diphosphomevalonate decarboxylase
EBGs	:	early biosynthetic genes
EF1a	:	elongation factor 1 alpha
EC	:	Enzyme Commission
=	:	equals
EAR	:	ethylene response factor-associated amphiphilic repression
EDTA	:	ethylenediaminetetraacetic acid
Na ₂ EDTA.2H ₂ O	:	ethylenediaminetetraacetic acid, disodium
FDR	:	False Discovery Rate
FPP	:	farnesyl diphosphate
FPS	:	farnesyl diphosphate synthase/farnesyl pyrophosphate synthetase
FeSO ₄ .7H ₂ O	:	Ferum sulphate
F3H	:	flavanone-3-hydroxylase
FS1	:	flavone synthase 1
FS2	:	flavone synthase 2
FOMT	:	flavonoid O-methyltransferase
F3'H	:	flavonol 3'-hydroxylase
F3H	:	flavonol 3-hydroxylase
FLS	:	flavonol synthase
FLS	:	flavonol synthase
GO	:	Gene Ontology
GPP	:	geranyl diphosphate
GPS	:	geranyl diphosphate synthase
C6	:	glioma cells
g	:	gram

>	: greater than
HTH	: helix-turn-helix
H4IIE	: hepatoma cell
HG	: Homogentisate
HT-29	: human colon cancer cell
HIV	: human immunodeficiency virus
PANC-1	: human pancreatic cancer cell
DU145	: human prostate cancer cell
PC-3	: human prostate cancer cell
PrEC	: human prostate epithelial cells
HCl	: hydrochloric acid
HMGS	: hydroxymethylglutaryl-CoA synthase
IAA	: Indole-3-acetic acid
IFS	: isoflavone synthase
IPP	: isopentenyl diphosphate
IPI:	: isopentenyl-diphosphate delta-isomerase
KNO ₃	: kalium nitrate
kb	: Kilobases
KEGG	: Kyoto Encyclopedia of Genes and Genomes
LBGs	: late biosynthetic genes
≤	: less than or equal to
LDOX	: leucoanthocyanidin dioxygenase
LAR	: leucoanthocyanidin reductase
L	: Litre
LB	: Luria-Bertani
Mg	: magnesium

MgCl	:	magnesium chloride
MgSO ₄	:	magnesium sulphate
MgSO ₄ .7H ₂ O	:	magnesium sulphate
MnSO ₄ .4H ₂ O	:	manganase sulphate
MPSS	:	massively parallel signature sequencing
MMP1	:	matrix metalloproteinase-1
mRNA	:	messenger RNA
MRCNS	:	Methicillin-Resistant Coagulase-Negative Staphylococci
MRSA	:	Methicillin-Resistant <i>Staphylococcus aureus</i>
MSSA	:	Methicillin-Sensitive <i>Staphylococcus aureus</i>
MSCNS	:	Methicillin-Susceptible Coagulase-Negative Staphylococci
MEP	:	methyl erythritol phosphate
MT	:	methyltransferase
MVA	:	Mevalonate
MK	:	mevalonate kinase
μ	:	micro
μg	:	microgram
μg/ml	:	microgram per millilitre
μl	:	microlitre
μM	:	micromolar
μmol/m ² /s	:	micromoles per square meter per second
mg	:	milligram
mg/l	:	milligram per litre
mg/ml	:	milligram per millilitre
ml	:	millilitre
mm	:	millimetre

M	: molar
mol	: mole
MS	: Murashige and Skoog
ng	: nanogram
ng/μl	: nanogram per microlitre
nm	: Nanometer
NAA	: naphthalene acetic acid
NCBI	: National Center for Biotechnology Information
NGS	: next-generation sequencing
NO	: nitric oxide
NR	: non-redundant
OMT	: O-methyltransferase
ORF	: open reading frame
A2780	: ovarian cancer cell
C3'H	: p-coumaroyl 5-O-shikimate 3'-hydroxylase
CST	: p-coumaroyl shikimate transferase
%	: percent
PAL	: phenylalanine ammonia lyase
PMK	: phosphomevalonate kinase
PHB	: p-hydroxybenzoate
PKS	: polyketide synthase
PCR	: polymerase chain reaction
PVP	: polyvinylpyrrolidone
KCl	: potassium chloride
KI	: potassium Iodide
KH ₂ PO ₄	: potassium phosphthate

pH	: potential of hydrogen
PT	: prenyltransferase
PGE2	: prostaglandin E2
qPCR	: quantitative PCR
qRT-PCR	: quantitative real-time PCR
ROS	: reactive oxygen species
RPKM	: Reads per kb per Million reads
RQ	: relative quantification
RNA	: Ribonucleic acid
rRNA	: ribosomal RNA
RIN	: RNA integrity number
rpm	: rotations per minute
SAHC	: S-adenosylhomocysteine
SAMS	: S-adenosylmethionine synthetase
s	: seconds
SAGE	: serial analysis of gene expression
NaoAc	: sodium acetate
NaCl	: sodium chloride
Na ₂ MnO ₄ .2H ₂ O	: sodium manganate
SQE	: squalene epoxidase
SQS	: squalene synthetase
SD	: standard deviation
TFs	: transcription factors
TRs	: transcription regulators
tRNA	: transfer RNA
TBE	: tris-borate-EDTA

TNF- α	:	tumour necrosis factor-alpha
CURS	:	turmeric curcumin synthase
UDPG	:	UDP-glucosyltransferase
UDPG	:	UDP-glucosyltransferase
UV	:	Ultraviolet
v/v	:	volume per volume
WDR	:	WD40-repeat
ZnSO ₄ .7H ₂ O	:	zinc sulphate

University of Malaya

LIST OF APPENDICES

Appendix A: Solution preparation.....	265
Appendix B: Compositions of Murashige and Skoog (MS).....	281
Appendix C: RNA assessment results using Agilent 2100 Bioanalyzer.....	282
Appendix D: Transcriptome data.....	285
Appendix E: Gradient PCR for <i>BrMYB1</i> , <i>BrMYB2</i> and <i>BrMYB3</i>	296
Appendix F: PCR library colony screening.....	297
Appendix G: Map of pCRTM4-TOPO® vector.....	299
Appendix H: Predicted open reading frames.....	300
Appendix I: Conserved domain search using Protein blast.....	309
Appendix J: NCBI accession number for phylogenetic tree construction and multiple alignment analysis.....	313
Appendix K: Sequencing results.....	322
Appendix L: Blastx homology alignment search.....	328
Appendix M: qPCR data.....	331

CHAPTER 1: INTRODUCTION

1.1 Background

Boesenbergia rotunda (Linnaeus) Mansfield, Kulturpflanze (Perry & Metzger, 1980) is believed to have originated from the Indian, Southern China and Southeast Asia regions (Burkill et al., 1966). It is a traditional medicinal plant known locally in Malaysia and Indonesia as temu kunci, merkunci, dekunci or temu kecil (Burkill et al., 1966), in Thailand as kra-chai (Burkill et al., 1966), in China as Chinese ginger or Chinese keys, while its English name is finger root ginger.

B. rotunda is a perennial herb belonging to the Zingiberaceae family. It is a small herbaceous plant with short, slender rhizomes (Ching et al., 2007). The rhizomes are widely used in Southeast Asia as an edible spice or vegetable and in ethnomedicine as an ingredient for the treatment of aphthous ulcers, dry mouth, stomach discomforts, leucorrhoea, dysentery, inflammation, rheumatism and muscular pains (Larsen, 1996; Saralamp et al., 1996). Traditionally, their rhizomes are eaten raw to treat mouth ulcers (Wijayakusuma, 2001) or prepared together with other medicinal plant rhizomes as a tonic for postnatal treatment to restore blood circulation and to rejuvenate the body (Perry & Metzger, 1980; Wijayakusuma, 2001). Crushed rhizomes are used externally to release stomach gas, improve appetite, improve digestion and treat rheumatism (Perry & Metzger, 1980; Wijayakusuma, 2001).

The major bioactive constituents in *B. rotunda* are flavonoids. To date, more than 20 flavonoids have been isolated from *B. rotunda* and are classified into two main groups, flavanones and chalcones. Based on their flavonoid carbon skeleton structure, compounds that can be classified as flavanones include pinocembrin, pinostrobin, alpinetin, rotundaflavone I and rotundaflavone II, while cardamonin, 4-hydroxy panduratin A, panduratin A, isopanduratin A, boesenbergin A, krachaizin A and krachaizin B are

classified as chalcones (Hwang et al., 1982; Mongkolsuk & Dean, 1964; Morikawa et al., 2008; Trakoontivakorn et al., 2001; Tuntiwachwuttikul et al., 1984).

Among isolated secondary metabolites from *B. rotunda*, panduratin A has been shown to possess various medicinal properties which include anti-dengue (Kiat et al., 2006), anti-cancer (Cheah et al., 2011; Kirana et al., 2007; Win et al., 2007; Yun et al., 2006; Yun et al., 2005), anti-inflammatory (Tewtrakul et al., 2009; Tuchinda et al., 2002; Yanti et al., 2009a; Yanti et al., 2009b; Yun et al., 2003), anti-HIV-1 protease (Cheenpracha et al., 2006), antibacterial (Rukayadi et al., 2009; Yanti et al., 2009c), anti-aging (Shim et al., 2008a), antioxidant (Shim et al., 2008b; Shindo et al., 2006; Sohn et al., 2005) and anti-obesity properties (Kim et al., 2012).

Despite the extensive reports on the potential use of panduratin A, only limited amounts of panduratin A that can be extracted from their natural source. This has resulted in unmet market demands when high quantities of panduratin A are required. Harvesting of mature rhizomes require almost a one year planting cycle for *B. rotunda*. In addition, extraction of panduratin A from 10 kilograms of dried *B. rotunda* rhizome using a solvent extraction method only yielded approximately 715.2 mg of panduratin A (Tewtrakul et al., 2009). Although chemically synthesized panduratin A has been reported, the economics of the procedures continues to hinder large-scale production of panduratin A (Chee et al., 2010). Alternatively the enhancement of panduratin A production through genetic manipulation of its secondary metabolic pathways is a potential strategy for panduratin A yield improvement and this would require knowledge of its biosynthetic pathway which at present remains unclear.

Panduratin A production has been shown in a published report from this laboratory to be enhanced by the addition of exogenous phenylalanine into *B. rotunda* cell suspension cultures (Tan et al., 2012). Phenylalanine is an aromatic amino acid produced from the

shikimic acid pathway (Herrmann & Weaver, 1999). It provides the essential 6-carbon ring and 3-carbon side chain that is central to all phenylpropanoids. Phenylalanine is also the precursor for the production of cinnamic acid, the first phenylpropanoid in the phenylpropanoid pathway, which is eventually channelled into the production of most flavonoids in plants including panduratin A.

For elucidation of the genes that are involved in the panduratin A biosynthetic pathway, transcriptome profiles that were derived from phenylalanine treated and untreated (control) *B. rotunda* cell suspension cultures were sequenced, and compared. *De novo* transcriptome of *B. rotunda* was done by combining both transcripts from control and treated samples to generate longer sequences. Subsequently, gene regulation patterns between the control and phenylalanine treated cell suspension cultures were analysed using DGE analysis by mapping both transcriptome profiles to the *de novo* transcriptome database. The focus of the research was to resolve the gene regulation patterns in the phenylpropanoid pathway that leads to panduratin A biosynthesis in *B. rotunda* cell suspension cultures in response to exogenous phenylalanine. Additionally the *de novo* transcriptome data would also enrich the plant database and eventually serve as reference sequences for other Zingiberaceae family plant species.

1.2 Objectives

The objectives of this study are:

1. To sequence the total RNA in *Boesenbergia rotunda* in response to exogenous phenylalanine treatment through RNA-Seq technology
2. To compare the transcriptome data between the phenylalanine treated and untreated *B. rotunda* cell cultures
3. To analyse regulation patterns of genes in the flavonoid pathway that leads to panduratin A biosynthesis in *B. rotunda* cell suspension cultures in response to exogenous phenylalanine
4. To identify and characterize the R2R3 MYB transcription factors in *B. rotunda*.

1.3 Research Question

Can genes encoding for regulatory proteins and enzymes involved in flavonoid biosynthetic pathway be elucidated by comparing the transcriptome profile of control and phenylalanine-treated *B. rotunda* callus?

1.4 Hypothesis

Genes encoding for regulatory proteins and enzymes involved in flavonoid biosynthetic pathway can be elucidated by comparing the transcriptome profile of control and phenylalanine-treated *B. rotunda* callus.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction of *Boesenbergia rotunda*

Boesenbergia rotunda (Linnaeus) Mansfield, Kulturpflanze (Larsen, 1996) is synonym to *Gastrochilus panduratum* Ridley (Burkill et al., 1966), *Boesenbergia pandurata* (Roxb.), *Kaempferia pandurata* Roxb. and *Gastrochilus panduratus* (Roxb.) Ridl. (Perry & Metzger, 1980). It is believed to have originated from India, Southern China and Southeast Asia region (Perry & Metzger, 1980). It is a traditional medicinal plant that belongs to Zingiberaceae family. Locally, *B. rotunda* is known as temu kunci, merkunci, dekunci, temu kecil in Malaysia and Indonesia (Burkill et al., 1966). In Thailand, it is locally known as kra-chai (Saralamp et al., 1996). In China, *B. rotunda* is known as Chinese ginger or Chinese keys or fingerroot ginger.

Tuchinda et al. (2002), reported that *B. pandurata* which is *B. rotunda*'s synonym has four rhizome varieties including yellow, black, white and red rhizomes (Tuchinda et al., 2002). In this thesis, *B. rotunda* from yellow rhizome variety was used. The height of the plant is around 50 – 70 cm with underground rhizomes. The leaves are green, broad ovate-oblong in shape that emerge on stems rising from the red base of the plant (Figure 2.1 a). It has pink flowers and the underground part contains yellow rhizomes with flashy rootlets hanging from it (Figure 2.1 b-d). The rootlets resemble a bunch of keys or finger-like shape with strong aromatic smell (Figure 2.1 e). In plant tissue culture, *B. rotunda* callus formation was initiated from shoot bud explants (Yusuf et al., 2011a; Yusuf et al., 2011b). There are three types of calli formed i. e. friable, mixed and compact callus. In this study, mixed callus was separated from explants and further propagated (Figure 2.1 f). Mixed callus in the liquid media is referred to as cell suspension cultures (Figure 2.1 g).

Generally, both leaves and rhizomes of *B. rotunda* are widely used in Malaysia, Indonesia and Thailand as both spices in food ingredient and traditional medicine. In

traditional medicine, rhizomes and roots are used in the treatment of dry mouths, ulcers and coughs (Saralamp et al., 1996). The crushed rhizomes and roots are used externally for the treatment of rheumatism, to dispel flatulence, improve the appetite, and treatments for diarrhoea and dysentery (Burkill et al., 1966). Post-partum tonic mixture and paste are prepared from *B. rotunda* roots for women after giving birth and applied externally to the body to ease muscle pains (Burkill et al., 1966).

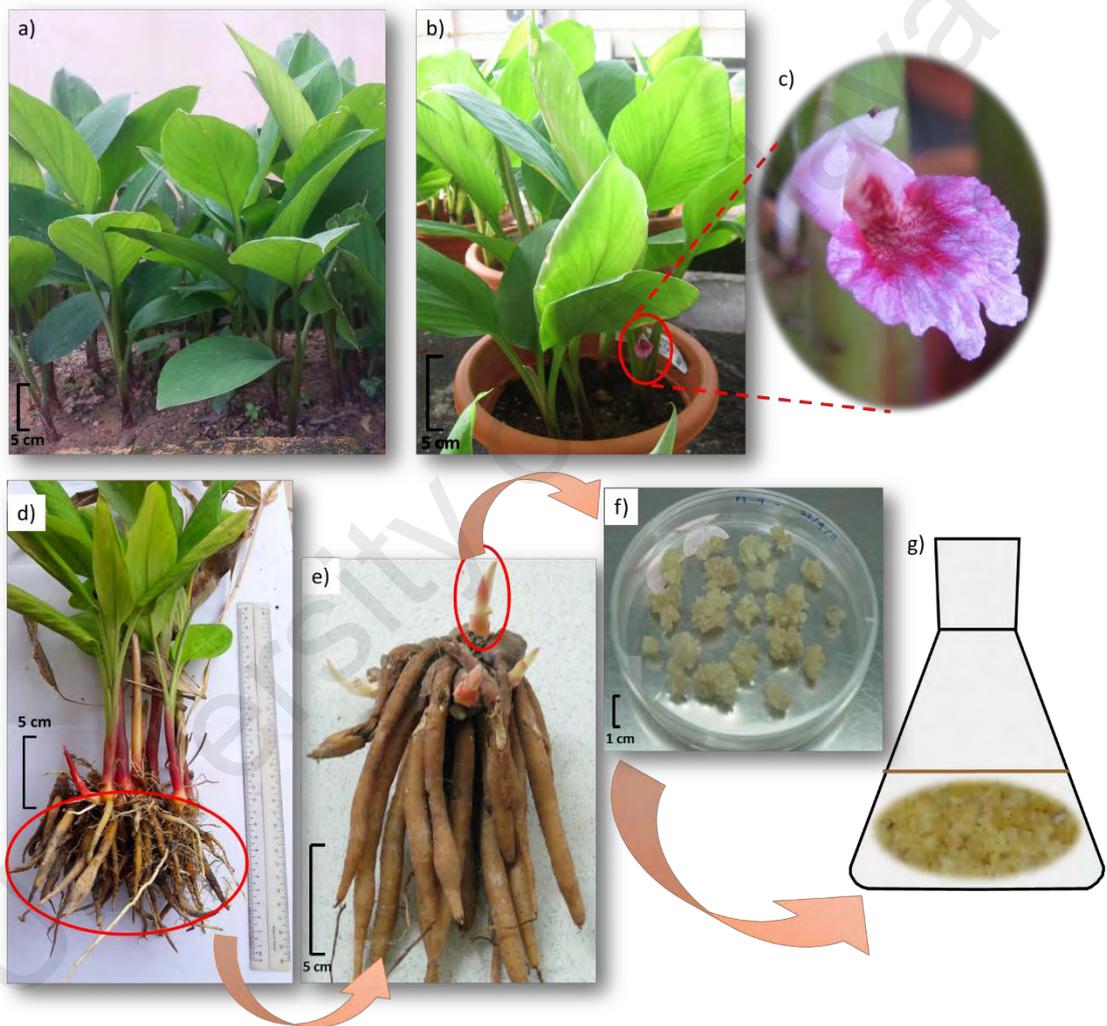


Figure 2.1: *Boesenbergia rotunda* plant.

a) *B. rotunda* planted on the ground; b) *B. rotunda* planted in the pot, showing the flower protruding between the stems, c) Flower d) *B. rotunda* plant showing the underground rhizomes; e) The shape of rhizomes resemble a bunch of keys or finger-like. Shoot bud is indicated by the red circle f) Yellow mixed callus initiated from shoot bud (red circle) and g) Mixed callus propagated in liquid media was referred as cell suspension cultures.

2.2 Primary and secondary metabolites in plants

Plants produce vast and diverse organic compounds that include primary and secondary metabolites. The characteristics of plant primary metabolites are uniform, universal and conservative, while secondary metabolites are unique, diverse and adaptive (Hartmann, 1996). Primary metabolites are essential for plant growth, development and reproduction while the absence of secondary metabolites may cause long-term impairment of plant survivability (Iriti & Faoro, 2009). The common examples of primary metabolites are amino acids, nucleotides, lipids and sugars, while secondary metabolites are flower pigments and scents.

There are over 100, 000 different secondary metabolites discovered in plants which can be classified under three main groups which include alkaloids, phenolics and terpenoids (Iriti & Faoro, 2009). In general, the products of primary metabolism are the precursors in the secondary metabolic pathways (Figure 2.2). For example, alkaloids are nitrogen-containing secondary metabolites derived from aromatic amino acids as end product of shikimate pathway (phenylalanine, tryptophan and tyrosine) and aliphatic amino acids derived from citric acid cycle (lysine and ornithine) (De Luca & St Pierre, 2000). Phenolic compounds such as phenylpropanoids flavonoids and lignins are derived from shikimate pathway and malonate pathway (Bowsher et al., 2008). Typically, phenolic compounds can be recognized by at least one aromatic ring structure with one or more hydroxyl groups. Terpenoids or isoprenoids such as monoterpenes, diterpenes, triterpenes, sesquiterpenes and polyterpenes that contain one or more C₅ units, are derived from the cytosolic mevalonate pathway or the plastidial methylerythritol phosphate pathway. The basic structure of terpenoids derived from five-carbon precursors of isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) (Nagegowda, 2010).

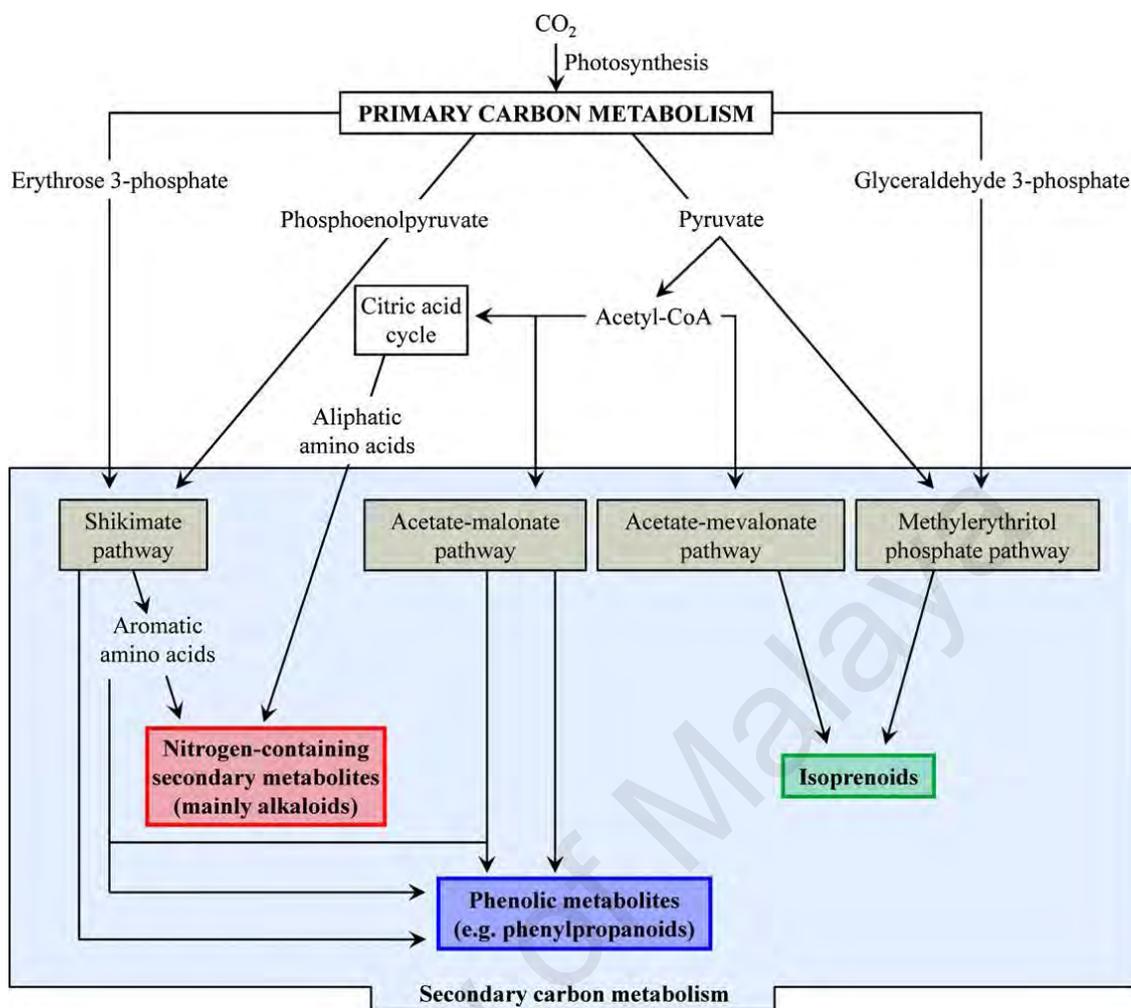


Figure 2.2: Biosynthetic pathways of alkaloids, phenolics and terpenoids derived from primary metabolism (Großkinsky et al., 2012).

The functions of plant secondary metabolites are defined mainly by their interactions with the environment which include protection against biotic and abiotic stresses such as defense against herbivores and pathogens, and protection against extreme climate changes and environmental stresses such as UV light, drought, flood and high saline salts (Hartmann, 1996). Additionally, secondary metabolites are also attractants (colours and scents) to flowers and fruits for pollination and seed dispersal (Iriti & Faoro, 2009).

Most of the alkaloids are poisonous and they serve major function in deterring herbivores. For instance, *Senecio jacobaea*, yellow-flowered ragwort containing pyrrolizidine alkaloids are toxic to cattle and horses (Cortinovis & Caloni, 2015).

Herbivores usually avoid the toxic plants high in alkaloids due to their bitter taste (Bowsher et al., 2008). Anthocyanins are important flavonoids that are responsible for the different colours of flowers and fruits in plants. For example, pelargonidin anthocyanins provide orange, pink and red hues in *Lupinus podophyllus*, while cyanidin anthocyanins provide the red colours in *Papaver rhaeas*, and delphinidin anthocyanins give blue colors in *Delphinium occidentale* (Bowsher et al., 2008). Flavonols are another important flavonoid that play a role as nectar guides which draw the pollinator towards the pollen and nectar by absorbing the ultraviolet light (Thompson et al., 1972). Additionally, flower scents also serve as tracking signals for insects. Depending on species, some scents are attractant and some are repellent due to their toxicity. For instance, volatile monoterpene myrcene is an attractive scent to a number of species, while β -pinene is generally a repellent (Bowsher et al., 2008).

Due to different environmental exposure and interactions with other living things, plants produce mixtures of secondary metabolites that vary at different times, space and developmental stages, which in turn lead to great structural diversity, restricted occurrence and high intraspecific variability (Hartmann, 1996). Hence, secondary metabolites are differentially distributed within the plant kingdom and the production is species-specific (Hartmann, 1996).

The structurally diverse secondary metabolites were produced through basic biosynthetic pathways leading to one or a few key metabolites that were later diversified by highly specific enzymatic transformations such as specific hydroxylation, O-methylation, dehydrogenation, esterification and glycosylation (Hartmann, 1996). One of the largest groups of secondary metabolites are phenolic compounds. Figure 2.3 shows the differences between primary and secondary metabolic pathways. Unlike primary metabolites, secondary metabolites do not undergo rapid turnover and their formation is

characterized by low steady state dynamics and thus low specific enzyme activities involved (Hartmann, 1996).

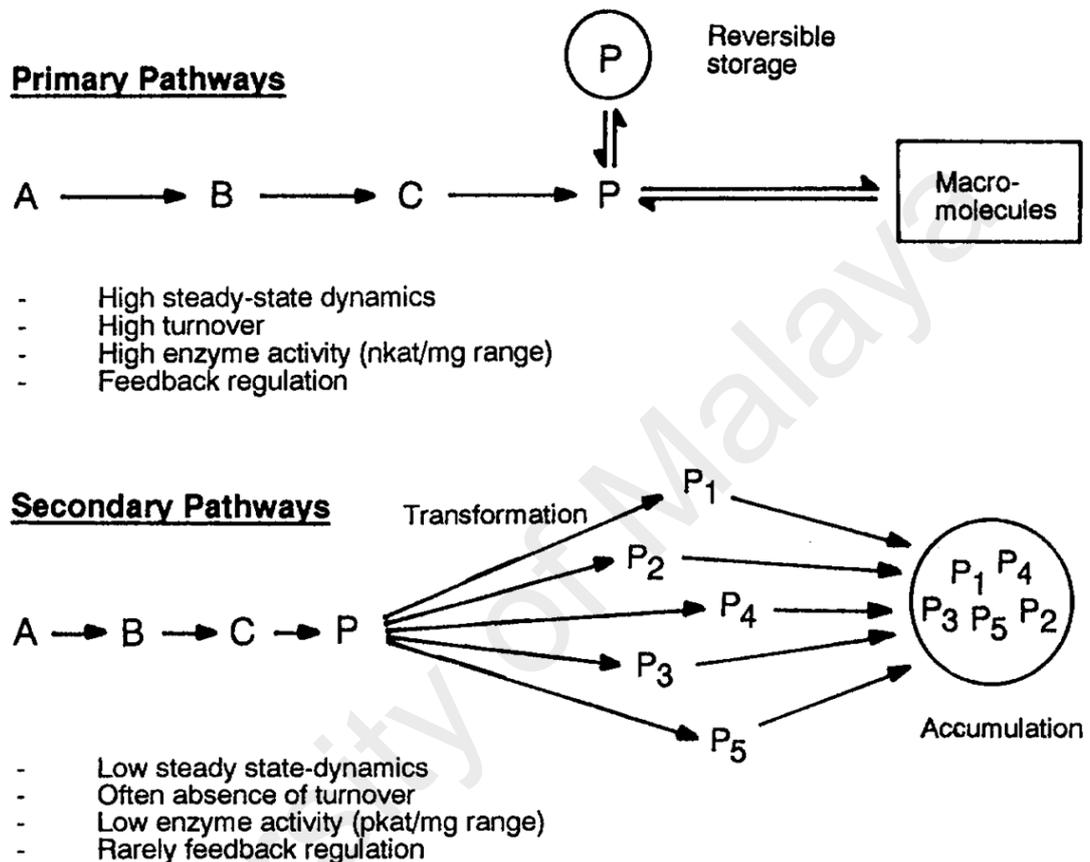


Figure 2.3: The differences in dynamics between primary and secondary pathways.

A to C is the intermediates of basic pathways; P is the product/key-intermediate; P₁ to P₅ is the transformation products (Hartmann, 1996).

2.3 Plant phenolics

Plants can produce a very wide range of phenolic compounds with diverse structures and properties. About 10,000 different plant phenolics have been identified, which share a common component, an aromatic hydrocarbon ring either phenyl or benzyl that is attached to at least one hydroxyl group (Bowsher et al., 2008). In plants, chorismic acid in the shikimate pathway serves as the substrate for primary metabolites leading to

biosynthesis of tryptophan, phenylalanine and tyrosine, which are eventually precursors of secondary metabolites (Weaver & Herrmann, 1997). As shown in Figure 2.4, the starting point of all plant phenolics is through the phenylalanine produced from the shikimic acid pathway (Herrmann & Weaver, 1999). Phenylalanine is an aromatic amino acid that provides the essential 6-carbon ring and 3-carbon side chain to all phenylpropanoids (Bowsher et al., 2008). The most abundant plant phenolic groups can be divided into four groups based on the carbon skeleton namely: simple phenolics, lignin, flavonoids and proanthocyanidins.

Plant phenolics that are derived from phenylpropanoid pathway are also referred to as phenylpropanoids. Simple phenolics consist of three main groups that include simple phenylpropanoids, benzoic derivatives and coumarins. The basic skeleton of simple phenylpropanoids consist of six-carbon phenyl ring attached to a three carbon side chain (C_6-C_3), while benzoic derivatives basic skeleton consist of a phenyl ring attached to one carbon side chain (C_6-C_1). Coumarins also have the basic C_6-C_3 like simple phenylpropanoid skeleton, but instead of a linear side chain, coumarins side chain is cyclized to form a ring.

Simple phenolics are volatile and serve as mobile signals for pollinators' attractants or herbivores' deterrents (Boeckler et al., 2011; Heil, 2011; Tahvanainen et al., 1985). Simple phenolics also serves as a phytoalexin and an allelopathics to protect plants from pathogens and competition from other neighbouring plants (Lattanzio et al., 1996; Li et al., 2010).

The general phenylpropanoid pathway branches into two major downstream pathways, lignin and flavonoid pathways (Figure 2.4). Lignin is a very complex polymer that is formed from monolignol (phenylpropanols) subunits oxidatively coupled through ether and carbon-carbon linkages (Ferrer et al., 2008). Lignin polymers consist of

polymerization of C₆-C₃ basic skeleton. The complex structures of lignin is important for plant structural support (Whetten & Sederoff, 1995). Different plants consist of different lignin subunit compositions and hence, a variety of different structural properties can be found in the plant kingdom (Whetten et al., 1998). In addition to structural support, lignin plays an important role in water transport in plants (Whetten & Sederoff, 1995).

Flavonoids are one of the largest groups of phenolic compounds found in plants (Ferrer et al., 2008). Flavonoids have additionally a six-carbon prenyl ring (C₆-C₃-C₆) that is derived from condensed three units of malonyl Co-A. Similar to lignin, proanthocyanidins are also phenolic polymers. Proanthocyanidins or condensed tannins are derived from flavonoids. Therefore, their basic structure consists of polymerization of C₆-C₃-C₆. In plants, condensed tannins serve as feeding deterrents and antifungal agents (Bryant et al., 1991; Latté & Kolodziej, 2000).

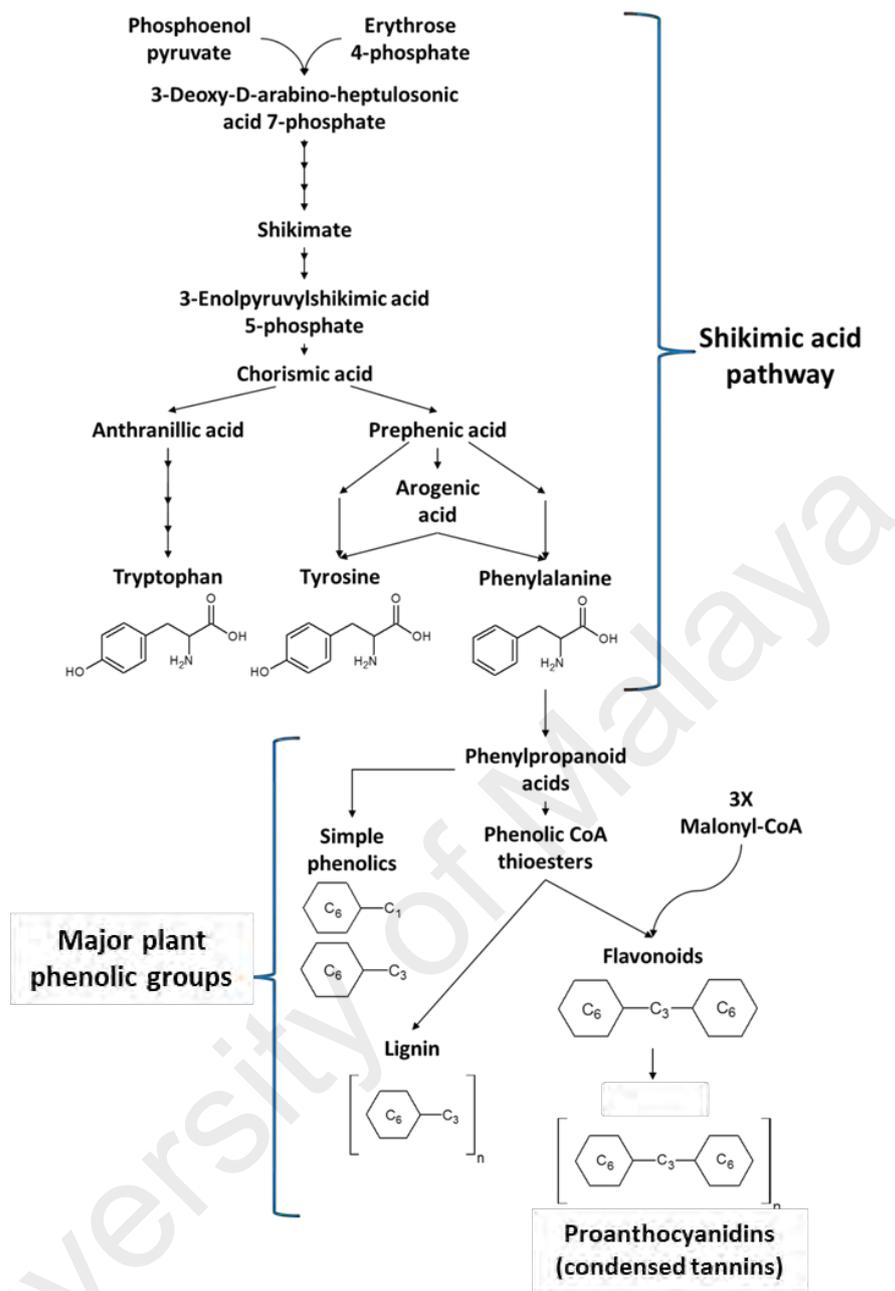


Figure 2.4: The relationship between shikimic acid pathway and biosynthesis of major plant phenolic groups adapted from Bowsher et al. (2008).

2.4 Plant flavonoids

Flavonoids are a very large and diverse group of phenolics. About 6000 flavonoids have been isolated and identified from thousands of plant species (Bowsher et al., 2008; Forkmann & Martens, 2001). Apart from medicinal plants, flavonoids also can be found in dietary foods. For instance, catechin and epicatechin can be found in tea; genistien and diadzin in soybean; kaempferol and quercetin in broccoli; naringenin and taxifolin in citrus fruits; and luteolin in red pepper (Heim et al., 2002).

Flavonoids have a wide range of biological functions in plants such as flowers, fruits and seed pigmentation; protection against UV light, defense against phytopathogens such as pathogenic microorganisms, insects and animals, acting as regulators in plant-microbe interactions, allelopathy in plant-plant interactions and nectar guide and pollinators' attractant for plants' reproduction (Gill & Tuteja, 2010; Heil, 2011; Mierziak et al., 2014). The classic example of flavonoids are anthocyanins that give red, purple, or blue pigments to flowers and fruits, depending on their pH and structures. Blueberries, cranberries, strawberries and grapes are rich in anthocyanins.

In flavonoid biosynthesis, phenylalanine from the shikimate pathway was converted into simple phenylpropanoids in phenylpropanoid pathways producing p-coumaroyl CoA, which provides the ring B together with three-carbon bridge that later formed ring C in flavonoid structure (Figure 2.5 a). Initially, cinnamic acid, the first phenylpropanoid acid was produced through deamination of phenylalanine catalysed by phenylalanine ammonia lyase (PAL). Then, cinnamic acid is hydroxylated to p-coumaric acid by cinnamate-4-hydroxylase (C4H) and subsequently, converted to p-coumaroyl-CoA by 4-coumaroyl:CoA ligase (4CL) through attachment of CoA to a phenolic compound (Figure 2.6). Subsequently, the pathway is divided into two major pathways; the flavonoid and lignin biosynthetic pathway (Du et al., 2009).

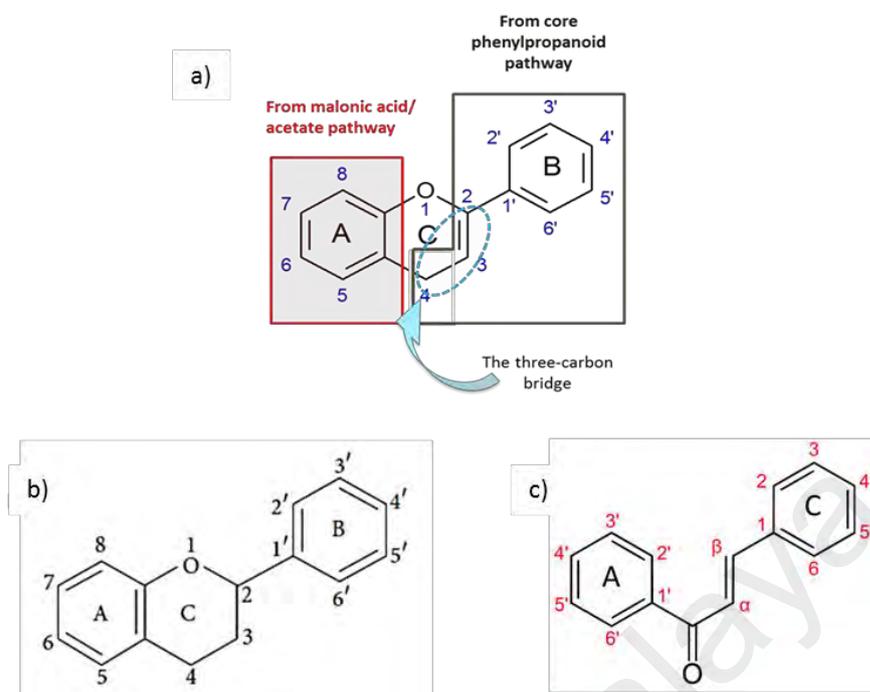


Figure 2.5: a) Basic flavonoid skeleton structure showing its origins in the shikimate pathway and malonate pathways (Bowsher et al., 2008). b) The flavonoid numbering system of the carbon atoms on the three rings A, B and C (Kumar & Pandey, 2013). c) The chalcone numbering system of the carbon atoms on the ring A and C (Buckingham & Munasinghe, 2015).

The first enzyme in the flavonoid pathway is chalcone synthase (CHS) which is responsible for production of chalcone by condensation of three molecules of malonyl CoA from malonate pathway with a molecule of p-coumaroyl CoA. As shown in the simplified flavonoids biosynthetic pathway, chalcones serve as the precursors for all flavonoids (Figure 2.6). All flavonoids share the same basic skeleton; consisting of two six-carbon rings (ring A and ring B) linked by a three-carbon bridge that usually forms the third ring, ring C (Figure 2.5 a).

Unlike chalcones, flavonoids are close-chain $C_6-C_3-C_6$ compounds. Therefore the numbering system of flavonoid constituents differ from open-ring chalcones (Figure 2.5 b and c). For flavonoids, the carbon numbers in ring A and ring C are un-primed, while for chalcones, the carbon numbers in ring C are un-primed. Besides, the location of ring C for both groups also differs.

As shown in Figure 2.6, flavonoids are classified into several groups based on modifications on rings A, B or C which include chalcones, aurones, dihydrochalcones, flavanones, flavanols, flavones, isoflavones, flavanonols, flavones, leucoanthocyanidins, anthocyanins and proanthocyanidins. Chalcones and dihydrochalcones are distinguished from other flavonoids due to their open-chain C₆-C₃-C₆ structures (Buckingham & Munasinghe, 2015). In flavonoid pathway, chalcones are converted to flavanones by isomerization catalysed by chalcone isomerase (CHI). Flavanones serve as important intermediates to produce other major flavonoid groups such as flavanols, flavones, isoflavones, dihydroxyflavonols, flavones, leucoanthocyanidins, anthocyanins and proanthocyanidins. Reductases (DFR, LAR and ANR), hydroxylases (F3H) and oxydoreductases (FS1, FS2, FLS, IFS and ANS) are enzymes that are responsible for the flavonoid biosynthesis. Polymerization of flavan-3-ols produce proanthocyanidins or condensed tannins, while polymerization of flavan-4-ols produce phlobaphenes.

In general, the flavonoid groups differ in the level of oxidation and ring C substitution pattern, while individual compounds within a group differ in the substitution pattern on ring A and B (Kumar & Pandey, 2013). Luteolin, apigenin and chrysin have hydroxyl group at the same position on ring A, but differ in hydroxyl group position on ring B. Luteolin has a hydroxyl group at C-3' and C-4' on ring B, while apigenin only has the hydroxyl group at C-4' and none hydroxyl group on ring B of chrysin (Table 2.1). Hence, further flavonoid modifications on the flavonoid backbone structures such as O-methylation, hydroxylation, glycosylation, acylation and prenylation leading towards further diverse complex flavonoid derivatives (Forkmann & Martens, 2001). Besides, flavonoids biological activities are structure dependent and the modifications often alter their solubility, reactivity and stability (Kumar & Pandey, 2013; Mierziak et al., 2014).

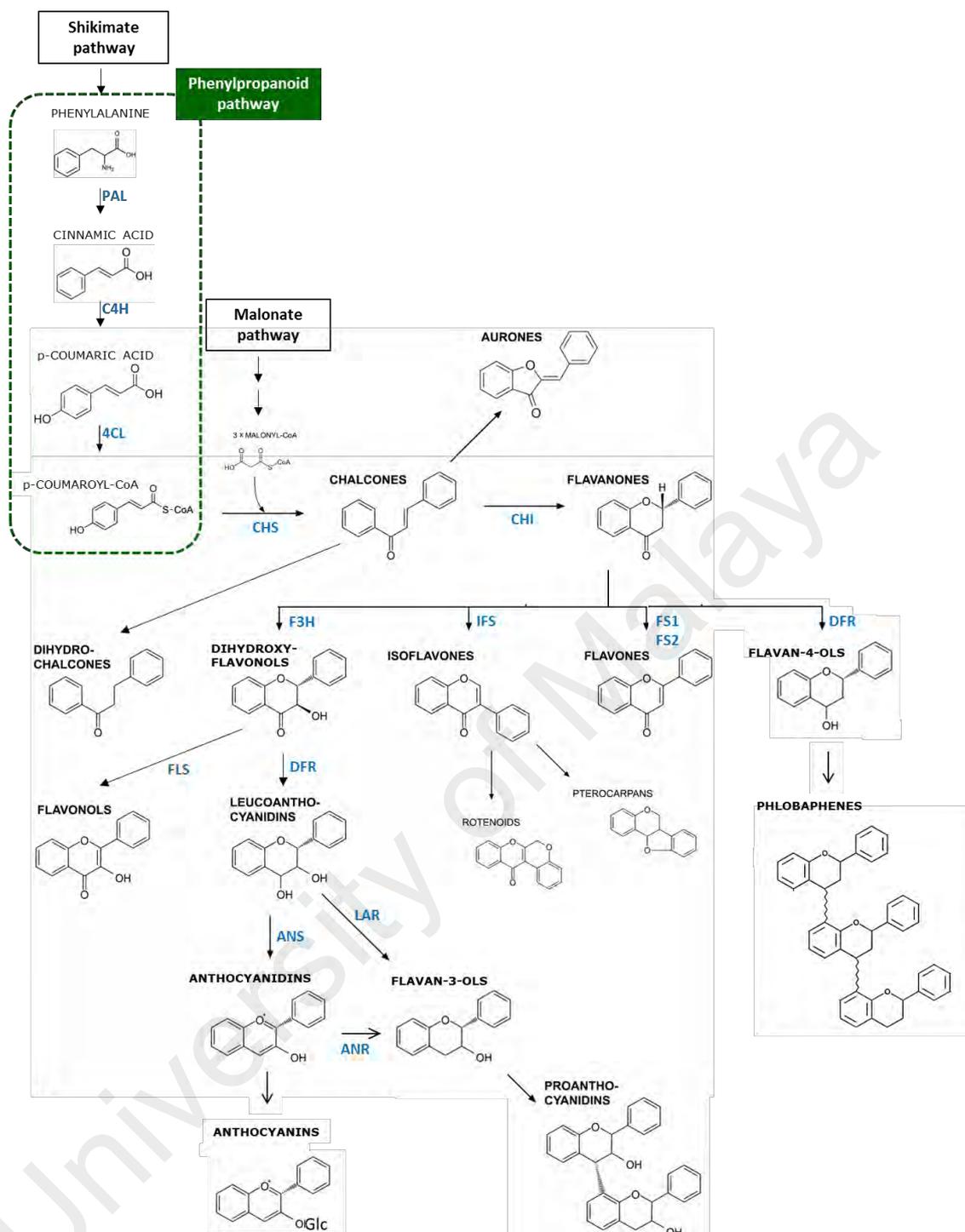


Figure 2.6: Overview of simplified pathways leading to major flavonoids classification: chalcones, aurones, flavanones, dihydrochalcones, dihydroxyflavonols, isoflavones, flavones, flavonols, phlobaphenes leucoanthocyanidins, anthocyanins and proanthocyanidins (Bowsher et al., 2008; Mierziak et al., 2014; Winkel-Shirley, 2001).

Enzyme abbreviations: PAL, phenylalanine ammonia lyase; C4H, cinnamate-4-hydroxylase; 4CL, 4-coumaroyl:coenzyme A ligase; CHS, chalcone synthase; CHI, chalcone isomerase; FS1/FS2, flavone synthase 1 and 2; IFS, isoflavone synthase; DFR, dihydroflavonol 4-reductase; F3H, flavanone-3-hydroxylase; FLS, flavonol synthase; ANS, anthocyanidin synthase; LAR, leucoanthocyanidin reductase and ANR, anthocyanidin reductase.

Table 2.1: Examples of flavonoids based on the structure backbone of the flavonoid groups, such as flavones, flavonols, flavanones, flavanonols, isoflavones and flavan-3-ols (Kumar & Pandey, 2013).

Group of flavonoid	Structure backbone	Examples
Flavones		 Luteolin Apigenin Chrysin
Flavonols		 Quercetin Kaempferol Galangin
Flavanones		 Hesperetin Naringenin
Flavanonol		 Taxifolin
Isoflavones		 Genistein Daidzein
Flavan-3-ols		 Catechin Epicatechin

The biological activities of flavonoids benefit human as they act as natural antioxidants and can be used to treat diseases and improve human health (Mierziak et al., 2014). The antioxidant properties of flavonoids are dependent on the ability of flavonoids to scavenge reactive oxygen species (ROS) and chelate metal catalyst (Gill & Tuteja, 2010). Flavonoids also inhibit the lipid peroxidation that can cause membrane damage (Ratty & Das, 1988). Flavonoids have also been recognized to exert anti-inflammatory, anti-allergic, anti-diabetic, anti-carcinogenic, antibacterial and anti-viral activities (Babu et

al., 2013; García-Lafuente et al., 2009; Kawai et al., 2007; Liu et al., 2008; Middleton Jr et al., 2000; Seelinger et al., 2008; Xu & Lee, 2001). Hence, flavonoids are considered as the major plant bioactive secondary metabolite due to their wide range of medicinal properties.

2.5 Secondary metabolite pathways in plants

Each plant species has its unique sets of secondary metabolites (Hartmann, 1996). Hence, some secondary metabolite production can only be found exclusively within a specific plant family, or plant genera or plant species. Secondary metabolites are produced through basic biosynthetic pathways leading to one or few key metabolites that are later diversified by highly specific enzymatic transformation, which eventually activate the exclusive pathways in plants (Hartmann, 1996). Although many secondary metabolites have been isolated and identified in plants, most of the specific biosynthetic pathways remain unresolved.

For instance, even though ginger and turmeric are from the same Zingiberaceae family, they have different attributes and secondary metabolite compositions. Ginger has a yellowish interior rhizome, while turmeric has a bright orange interior rhizome. The main compound of ginger (*Zingiber officinale*) is gingerols, while the main compound in turmeric (*Curcuma longa*) is curcuminoids. Figure 2.7 shows the proposed pathway of gingerols and curcumin in ginger and turmeric. The basic biosynthetic pathway was indicated by solid arrows and the conversions have been demonstrated in other plant species (Koo et al., 2013). The highly specific enzymes such as polyketide synthases (PKSs), reductases, hydroxylases and O-methyltransferases were proposed to be directly involved in gingerols and curcumin biosynthetic pathways. According to Koo et al. (2013), diketide synthases and curcuminoid/gingerol synthases are two distinct classes of

polyketide synthases that worked in tandem. The production of different compounds in these plants is dependent on the different combinations of these two classes.

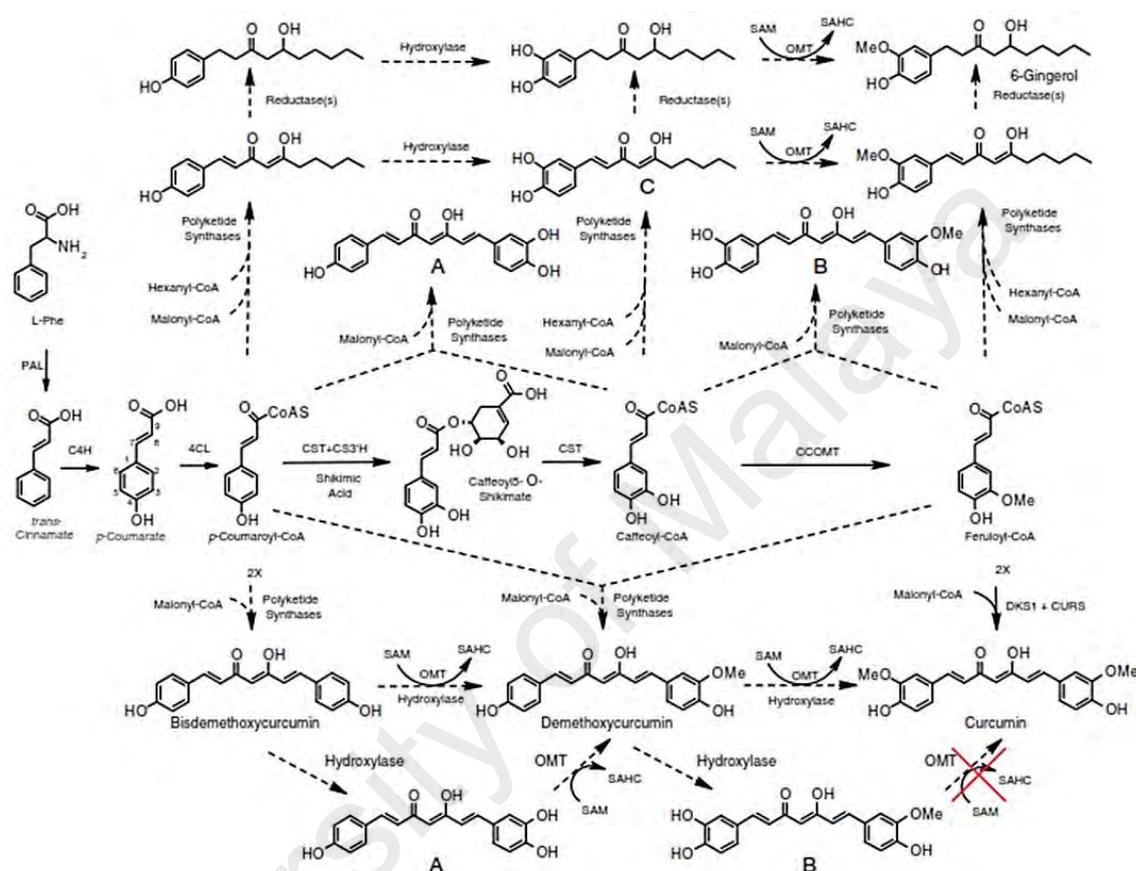


Figure 2.7: Proposed biosynthetic pathway of gingerol and curcumin that are exclusively present in ginger and turmeric, denoted as dotted arrows (Koo et al., 2013).

Enzyme abbreviations: PAL: phenylalanine ammonia lyase; C4H: cinnamate 4-hydroxylase; 4CL: 4-coumarate:CoA ligase; CST: p-coumaroyl shikimate transferase; C3'H: p-coumaroyl 5-O-shikimate 3'-hydroxylase; OMT: O-methyltransferase; CCOMT: caffeoyl-CoA O-methyltransferase; SAMs: S-adenosylmethionine synthetase; SAHC: S-adenosylhomocysteine; DKS1; diketide synthase 1 and CURS; turmeric curcumin synthase. The large red X and the solid arrows associated with the DKS1 + CURS reactions indicate that formation of curcumin has been demonstrated to proceed directly from feruloyl-CoA, and not through the orthodiol intermediate B. Compounds A and B, therefore, are not likely to be intermediates in the pathway to curcumin but instead are likely to be products of a different pair of PKS enzymes.

In order to validate the proposed pathway, a series of the basic procedures have been reported including isolating the cDNA of enzymes that are involved in the proposed

pathway, followed by cloning into an expression vector and enzyme assay analysis. Several secondary metabolite biosynthetic pathways were successfully elucidated through these methods. For instance, 8-dimethylallylnaringenin from *Sophora flavescens* was catalysed by prenyltransferase enzyme, SfN8DT-1 (Figure 2.8). 8-dimethylallylnaringenin, an active estrogenic compound is not exclusively present in *Sophora flavescens* but also can be found in *Humulus lupulus* and other plant species (Keiler et al., 2013).

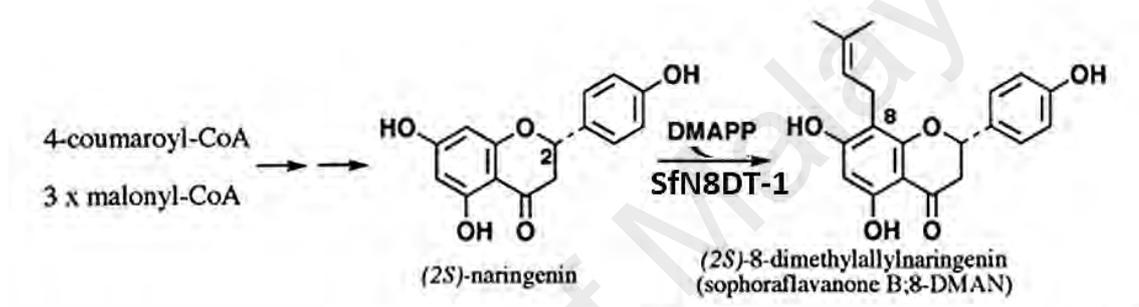


Figure 2.8: Established biosynthetic pathway of 8-dimethylallylnaringenin from *Sophora flavescens* (Sasaki et al., 2008; Yamamoto et al., 2000).

Biosynthetic pathways of α -bitter acid, β -bitter acid and xanthohumol in *Humulus lupulus* can be considered as more complicated pathways as more specific enzymes are involved compared to the 8-dimethylallylnaringenin pathway (Figure 2.9). The establishment of the pathways were almost completed with elucidation of 8 enzymes in the pathways through enzymatic assay and molecular functional characterizations approach. Chalcone synthase, H1CHS and valetophenone synthase, H1VPS from *Humulus lupulus* were the first enzymes characterized in the pathway (Okada et al., 2004); followed by carboxyl CoA ligases, H1CCL1, H1CCL2 and H1CCL4 (Xu et al., 2013) and prenyltransferases, H1PT1, H1PT1-L and H1PT2 (Li et al., 2015; Tsurumaru et al., 2012).

However, in *B. rotunda*, although many bioactive compounds have been isolated and identified, the specific enzymes involved in their biosynthetic pathways are still unknown. Especially the compounds that are exclusively present in *B. rotunda* such as panduratin A, isopanduratin A and 4-hydroxypanduratin A. Hence, it is crucial to elucidate and functionally characterize the highly specific enzymes involved in the pathway to establish this unique pathway in *B. rotunda*.

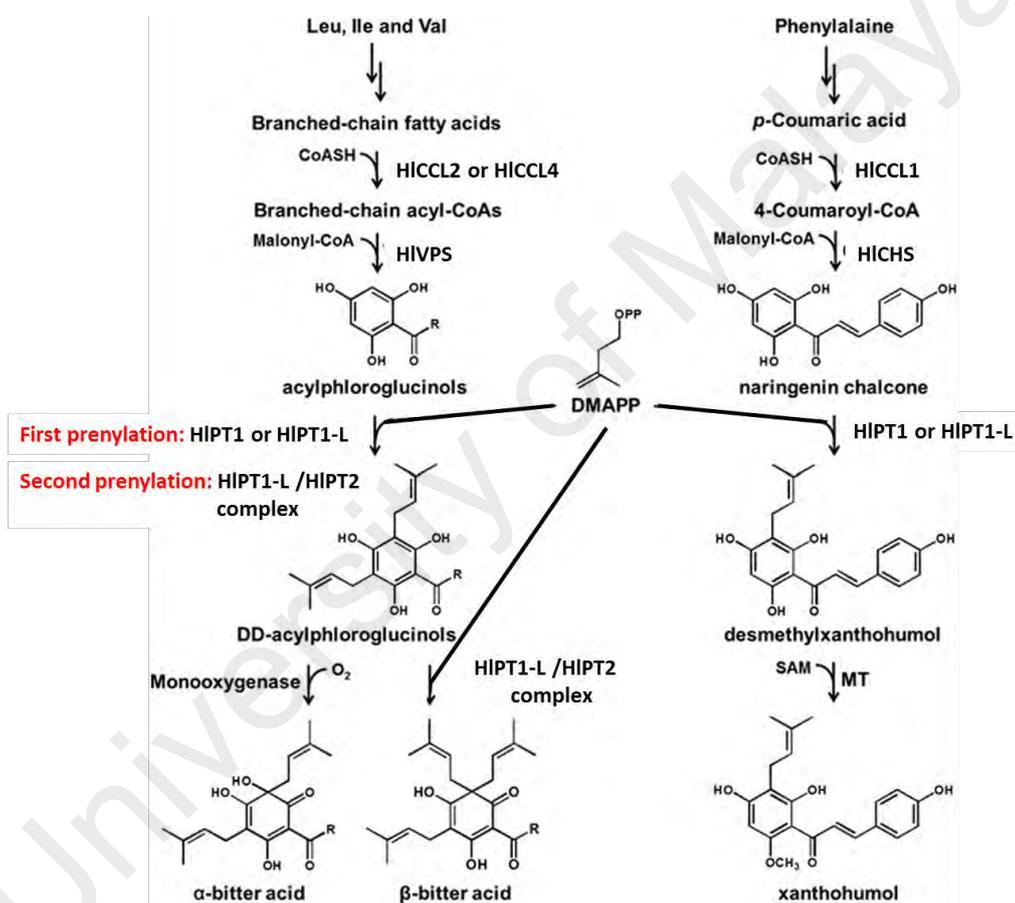


Figure 2.9: Biosynthetic pathways of α -bitter acid, β -bitter acid and xanthohumol in *Humulus lupulus* (Li et al., 2015).

The R group in acylphloroglucinols, DD-acylphloroglucinols and bitter acids are isobutyryl, isopropyl and butan-2-yl groups.

Enzyme abbreviations: HICCL: *Humulus lupulus* carboxyl CoA ligase; HICHS: *Humulus lupulus* chalcone synthase; HIVPS: *Humulus lupulus* valetophenone synthase; HIPT: *Humulus lupulus* prenyltransferase and MT: methyltransferase.

2.6 Secondary metabolites in *B. rotunda*

Flavonoids are major constituents in *B. rotunda* extracts as presented in Table 2.2, which include flavanones and chalcones. The flavonoid pathway is a common secondary metabolite pathway in plants. However, some of the flavonoids are present only in certain plant families or plant species. For example, alpinetin, pinocembrin and pinostrobin are found in more than one plant family, including Zingiberaceae as presented in Table 2.3. Nevertheless, not all plant species in the Zingiberaceae family contain these compounds. Unlike alpinetin, which can be found in *Alpinia mutica*, *Alpinia pinnanensis*, *Alpinia* spp. and *B. rotunda*, pinocembrin and pinostrobin are only present in *B. rotunda*. Besides, chalcones such as boesenbergin A and boesenbergin B; and cyclohexanyl chalcones such as panduratin A, isopanduratin A, 4-hydroxypanduratin A and 2-hydroxypanduratin A are found exclusively in *B. rotunda*. This suggests that *B. rotunda* consists of specific pathways that lead to the biosynthesis of these compounds that are not present in the other plant species.

Table 2.2: Some secondary metabolites isolated from *B. rotunda*.

Secondary metabolite compounds	Reference
Pinostrobin	(Mongkolsuk & Dean, 1964)
Alpinetin	(Mongkolsuk & Dean, 1964)
Boesenbergin A	(Jaipetch et al., 1982)
Boesenbergin B	(Mahidol et al., 1984)
Panduratin A	(Tuntiwachwuttikul et al., 1984)
Panduratin B	(Pancharoen et al., 1987)
Panduratin C	(Cheenpracha et al., 2006)
Panduratin D – I	(Win et al., 2008)
Cardamonin	(Jaipetch et al., 1982)
Pinocembrin	(Jaipetch et al., 1982)
Pinocembrin chalcone	(Trakoontivakorn et al., 2001)
Pinostrobin chalcone	(Jaipetch et al., 1982)
Isopanduratin A	(Hwang et al., 2004)
4-hydroxypanduratin A	(Trakoontivakorn et al., 2001)
Rotundaflavone I	(Morikawa et al., 2008)

Table 2.2, continued

Rotundaflavone II	(Morikawa et al., 2008)
Krachaizin A	(Morikawa et al., 2008)
Krachaizin B	(Morikawa et al., 2008)
geranyl-2,4-dihydroxy-6-phenethylbenzoate	(Win et al., 2007)
3'-geranylcardamonin or 2',4'-dihydroxy-3'-(1''-geranyl)-6'-methoxychalcone	(Win et al., 2007)
(1' R,2' S,6' R)-2-hydroxyisopanduratin A	(Win et al., 2007)
8-geranylpinostrobin	(Win et al., 2007)

Table 2.3: Summary of the presence of some metabolite compounds across the plant families and plant species (KNapSack Metabolite Information).

Secondary metabolite Compounds	Family	Species
Alpinetin	Asteraceae	<i>Helichrysum forskahlii</i> , <i>Helichrysum spp.</i> , <i>Mikania micrantha</i>
	Betulaceae	<i>Alnus spp.</i>
	Combretaceae	<i>Combretum albopunstatum</i> Suesseng
	Fabaceae	<i>Dalbergia parviflora</i> , <i>Dalea scandens</i> ,
	Labiatae	<i>Scutellaria amabilis</i> HARA, <i>Scullaria spp.</i>
	Myrtaceae	<i>Eucalyptus spp.</i>
	Piperaceae	<i>Piper spp.</i>
	Zingiberaceae	<i>Alpinia mutica</i> , <i>Alpinia pinnanensis</i> , <i>Alpinia spp.</i> , <i>Boesenbergia rotunda</i>
Pinocembrin	Annonaceae	<i>Anomianthus dulcis</i> , <i>Uvaria chamae</i>
	Asteraceae	<i>Flourensia hirsuta</i> , <i>Flourensia ilicifolia</i> , <i>Flourensia retinophylla</i> , <i>Helichrysum forskahlii</i> , <i>Helichrysum spp.</i> , <i>Phonus arborescens</i>
	Betulaceae	<i>Alnus spp.</i>
	Combretaceae	<i>Combretum albopunstatum</i> Suesseng
	Fabaceae	<i>Glycyrrhiza glabra</i>
	Labiatae	<i>Scullaria spp.</i>
	Myrtaceae	<i>Eucalyptus spp.</i>
	Pinaceae	<i>Pinus cembra</i> , <i>Pinus spp.</i> , <i>Pseudotsuga wilsoniana</i>
	Piperaceae	<i>Piper gaudichaudianum</i> , <i>Piper hostmannianum</i>
	Rosaceae	<i>Citrus spp.</i>
	Salicaceae	<i>Populus spp.</i>
	Santalaceae	<i>Viscum coloratum</i>
	Zingiberaceae	<i>Boesenbergia rotunda</i>

Table 2.3, continued

Pinostrobin	Annonaceae	<i>Uvaria chamae</i>
	Asteraceae	<i>Helichrysum spp.</i>
	Betulaceae	<i>Alnus sp.</i>
	Labiatae	<i>Scullaria spp.</i>
	Lauraceae	<i>Aniba sp.</i>
	Pinaceae	<i>Larix sp., Pinus sp., Pinus strobus</i>
	Polygonaceae	<i>Polygonum ferrugineum</i>
	Rosaceae	<i>Prunus sp.</i>
	Salicaceae	<i>Populus sp.</i>
	Zingiberaceae	<i>Boesenbergia rotunda</i>
Cardamonin	Asteraceae	<i>Helichrysum forskahlii, Helichrysum spp.</i>
	Lauraceae	<i>Lindera umbellata</i>
	Piperaceae	<i>Piper aduncum, Piper hispidium,</i>
	Pteridaceae	<i>Pityrogramma chrysophylla, Pityrogramma tartarea</i>
	Salicaceae	<i>Populus spp.</i>
	Woodsiaceae/ Dryopteridaceae	<i>Woodsia scopulina</i>
	Zingiberaceae	<i>Alpinia mutica, Boesenbergia rotunda</i>
Pinostrobin chalcone	Asteraceae	<i>Helichrysum spp.</i>
	Lauraceae	<i>Lindera umbellata</i>
	Piperaceae	<i>Piper aduncum, Piper hispidium,</i>
	Pteridaceae	<i>Onychium siliculosum, Pityrogramma chrysophylla, Pityrogramma tartarea</i>
	Salicaceae	<i>Populus spp.</i>
	Zingiberaceae	<i>Boesenbergia rotunda</i>
Panduratin A	Zingiberaceae	<i>Boesenbergia rotunda</i>
Isopanduratin A	Zingiberaceae	<i>Boesenbergia rotunda</i>
4-hydroxypanduratin A	Zingiberaceae	<i>Boesenbergia rotunda</i>
2-hydroxypanduratin A	Zingiberaceae	<i>Boesenbergia rotunda</i>
Boesenbergin A	Zingiberaceae	<i>Boesenbergia rotunda</i>
Boesenbergin B	Zingiberaceae	<i>Boesenbergia rotunda</i>

2.6.1 Bioactive compounds in *B. rotunda*

Secondary metabolites are beneficial to plants and human. Secondary metabolites that possess medicinal properties are defined as bioactive compounds. Examples include taxol, an antileukemic and antitumor agent from *Taxus brevifolia* (Wani et al., 1971) and xanthohumol, anti-HIV-1 agent purified from Hops *Humulus lupulus* (Wang et al., 2004).

In general, solvent extraction methods were used to screen bioactive compounds in medicinal plants such as leaves, whole plants, fruit, roots or rhizomes. Secondary metabolite compounds were extracted based on polarity and solubility characteristics. Chloroform, methanol, ethyl acetate and hexane are the commonly used solvents. The solvent extracts containing mixture of compounds were then tested for various medicinal properties. In *B. rotunda*, there have been several bioactive compounds identified, which include pinostrobin, pinocembrin; cardamonin, panduratin A, 4-hydroxypanduratin A and isopanduratin A (Figure 2.10). The summary of medicinal properties of solvent extracted and bioactive compounds in *B. rotunda* which include anti-cancer, anti-dengue, anti-inflammatory, anti-HIV, anti-aging, antioxidant, antibacterial, anti-tumor and antimutagenic are represented in Table 2.4. Among the bioactive compounds mentioned above, panduratin A and 4-hydroxypanduratin A has been very often reported to possess medicinal properties.

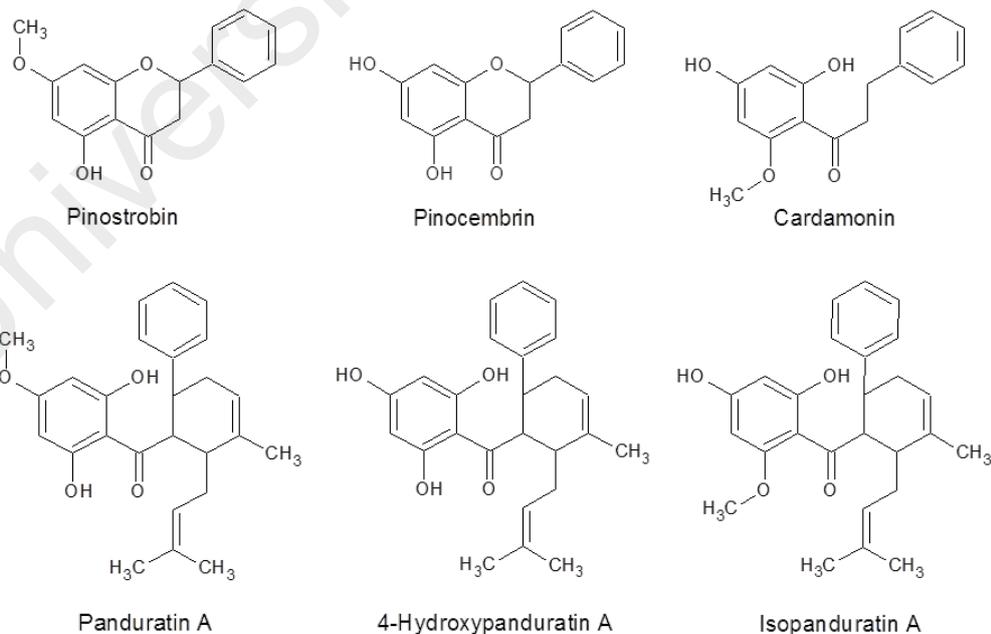


Figure 2.10: Bioactive compounds extracted from *B. rotunda* which include flavanones such as pinostrobin and pinocembrin; and chalcones such as cardamonin, panduratin A, 4-hydroxypanduratin A and isopanduratin A.

Table 2.4: The summary of medicinal properties in *B. rotunda* which include anti-cancer, anti-dengue, anti-inflammatory, anti-HIV, anti-aging, antioxidant, antibacterial, anti-tumor and antimutagenic.

Properties	Potential Use	Extract / Bioactive compounds	Reference
Anti-cancer	Colon cancer treatment	Panduratin A	(Yun et al., 2005)
Anti-cancer	Prostate cancer treatment	Panduratin A	(Yun et al., 2006)
Anti-cancer	Breast cancer treatment	Panduratin A	(Kirana et al., 2007)
Anti-cancer	Colon cancer treatment	Panduratin A	(Kirana et al., 2007)
Anti-cancer	Lung cancer treatment	Panduratin A	(Cheah et al., 2011)
Anti-cancer	Lung cancer treatment	Panduratin A	(Cheah et al., 2013)
Anti-cancer	Skin cancer treatment	Panduratin A	(Lai et al., 2015)
Anti-cancer	Pancreatic cancer treatment	Panduratin A Nicolaioidesin B	(Win et al., 2007)
Anti-cancer	Lung cancer treatment	Boesenbergin A	(Isa et al., 2012)
Anti-angiogenic	Cancer treatment	Panduratin A	(Lai et al., 2012)
Anti-dengue	Dengue treatment	Panduratin A 4-Hydroxypanduratin A	(Kiat et al., 2006)
Anti-HIV-1 protease	HIV	Chloroform and methanol extract	(Tewtrakul et al., 2003a)
Anti-HIV-1 protease	HIV	Cardamonin	(Tewtrakul et al., 2003b)
Anti-HIV-1 protease	HIV	4-Hydroxypanduratin A Panduratin A	(Cheenpracha et al., 2006)
Anti-inflammatory	Anti-inflammatory agent	Panduratin A	(Yun et al., 2003)
Anti-inflammatory	Prevent periodontal inflammation	Ethanollic extract	(Yanti et al., 2009d)
Anti-inflammatory	Prevent periodontal inflammation	Panduratin A	(Yanti et al., 2009b)
Anti-inflammatory	Prevent periodontal inflammation	Panduratin A	(Yanti et al., 2009a)
Anti-inflammatory	Periodontal disease treatment	Ethanollic extract	(Yanti & Hwang, 2010)

Table 2.4, continued

Anti-inflammatory	Anti-inflammatory agent	Panduratin A 4-Hydroxypanduratin A	(Tewtrakul et al., 2009)
Anti-inflammatory	Anti-inflammatory agent	Panduratin A 4-Hydroxypanduratin A	(Tuchinda et al., 2002)
Anti-inflammatory	Atopic dermatitis treatment	Ethanolic extract	(Kim et al., 2013)
Anti-inflammatory	Atopic dermatitis treatment	Panduratin A	(Kim et al., 2014)
Anti-inflammatory	Anti-inflammatory and hepatoprotective agent	Methanolic extract	(Morikawa et al., 2008)
Anti-aging	Prevention and treatment of skin aging	Panduratin A	(Shim et al., 2008a)
Anti-aging	Prevention and treatment of skin aging	Panduratin A	(Shim et al., 2008b)
Anti-aging	Prevention and treatment of skin aging	4-Hydroxypanduratin A	(Shim et al., 2009)
Anti-UV	Protective against UVB induced DNA damage	Ethanolic extract	(Listyawati et al., 2015)
Antioxidant	Protect against oxidative damage caused by toxic chemicals	Panduratin A	(Sohn et al., 2005)
Antioxidant	Treating hyperpigmentation as skin-whitening agents	Isopanduratin A 4-Hydroxypanduratin A	(Yoon et al., 2007)
Antioxidant	Treating hyperpigmentation as skin-whitening agents	Panduratin A	(Lee et al., 2010)
Antioxidant	Neuroprotection	Panduratin A 4-Hydroxypanduratin A	(Shindo et al., 2006)
Antioxidant	Hepatoprotectant/ liver cirrhosis treatment	Ethanolic extract	(Salama et al., 2012)
Antioxidant	Hepatoprotectant/ liver cirrhosis treatment	Ethanolic extract	(Salama et al., 2013a)
Antioxidant	Hepatoprotectant/ liver cirrhosis treatment	Panduratin A	(Salama et al., 2013b)
Antibacterial	Natural anticariogenic agent/ Prevent dental caries	Isopanduratin A	(Hwang et al., 2004)
Antibacterial	Natural anti-biofilm agent against multi-species oral biofilms	Panduratin A	(Yanti et al., 2009c)

Table 2.4, continued

Antibacterial	Gastric ulcer treatment	Pinostrobin	(Bhamarapavati et al., 2006)
Antibacterial	Treatment disease caused by staphylococci (MRSA, MSSA, MRCNS and MSCNS)	Panduratin A	(Rukayadi et al., 2009)
Antibacterial	Antibacterial against <i>S aureus</i>	Chloroform extract	(Voravuthikunchai et al., 2006)
Antibacterial	Treatment disease caused by enterococci	Panduratin A	(Rukayadi et al., 2010)
Antibacterial	Treatment of periodontitis and cariogenic causing bacterium	Panduratin A	(Park et al., 2005)
Antibacterial	Treatment of acne and skin disease causing microorganisms	Panduratin A Isopanduratin A	(Song et al., 2008)
Antibacterial	Diarhoea treatment	Essential oil	(Miksusanti et al., 2008)
Anti-tumour	Inhibit aminopeptidase activity	Methanolic extract	(Morikawa et al., 2008)
Anti-mutagenic	Strongly inhibit N-hydroxylation of Trp-P-2	Pinocembrin chalcone Cardamonin Pinocembrin Pinostrobin 4-hydroxypanduratin A Panduratin A	(Trakoontivakorn et al., 2001)
Anti-obesity	Prevention and treatment of obesity and associated fatty liver disease	Ethanolic extract	(Kim et al., 2012)
Anti-obesity	Treatment of metabolic disorder	Panduratin A	(Kim et al., 2011)
Anti-allergic	Agent against immediate-type hypersensitivity	Panduratin A	(Choi, Kim, & Hwang, 2012)
Anti-ulcerogenic	Gastric ulcer treatment	Methanolic extract and Pinostrobin	(Abdelwahab et al., 2011)
Wound healing	Wound healing	Ethanolic extract	(Mahmood et al., 2010)

2.6.2 Panduratin A and 4-hydroxypanduratin A

Panduratin A and 4-hydroxypanduratin A were reported to have stronger biological activities compared to other secondary metabolites in *B. rotunda* (Win et al., 2007). Both compounds were uniquely present in *B. rotunda*. Panduratin A can be used as an alternative medicine to replace chemotherapy to treat various types of cancer (Table 2.4). According to Yun et al. (2006), panduratin A inhibited the growth of cancerous cells by inducing apoptosis. Furthermore, panduratin A did not show toxic effects on normal healthy human prostate epithelial cells PrEC, thus making panduratin A a potential chemopreventing agent in cancer treatment (Yun et al., 2006).

Panduratin A also possess anti-viral activity which includes anti-dengue (Kiat et al., 2006) and anti-HIV-1 activities (Cheenpracha et al., 2006). The mode of action involves preventing virus replication by inhibiting the important enzymes that were essential for dengue and HIV virus propagation. Panduratin A prevents dengue virus replication by inhibiting DEN-2 virus NS3 protease (Kiat et al., 2006) and inhibits HIV virus activity by inhibiting HIV-1 protease activity (Cheenpracha et al., 2006). Recent reports indicate that the dengue outbreaks are increasing each year with 2.5 billion people at risk as reported by the World Health Organization (2012). Thus, developments of new drugs that possess anti-dengue properties are essential.

Panduratin A also has potential use in the dental care industry due to its anti-inflammatory activity preventing periodontal inflammation (Yanti et al., 2009d; Yanti et al., 2009b) and antibacterial activities thus preventing dental caries (Table 2.4) (Park et al., 2005). Panduratin A can also be used as a natural anti-biofilm agent against multispecies oral biofilm during early dental plaque formation (Yanti et al., 2009c). In addition, panduratin A and 4-hydroxypanduratin A also have potential use in the cosmeceutical industry. These compounds could inhibit melanin biosynthesis and can be

used as skin whitening agent (Lee et al., 2010; Yoon et al., 2007). In addition, the property of panduratin A as natural antibacterial agents could inhibit the growth of acnes and skin diseases caused by bacteria (Song et al., 2008).

Due to the increasing number of reports on the potential use of panduratin A as a new alternative medicine, dietary supplement and other health care products, demands on this compound in drug development and the health care industries is increasing. However, production of panduratin A in *B. rotunda* is limited and unable to cater the demands. According to Tewtrakul et al. (2009), 10 kilogram of dried *B. rotunda* rhizome yields only 715.2 mg of panduratin A. Additionally, it takes almost one year of planting, to harvest their mature rhizomes. Although panduratin A can be chemically synthesized, the production involves laborious processes and expensive starting materials (Chee et al., 2010). Besides, the manipulation of their production through metabolic engineering is hindered due to the lack of knowledge on the panduratin A biosynthetic pathway. Thus, elucidating the intermediate enzymes involved in the biosynthetic pathway is crucial to enable enhancement of panduratin A production. The intermediate enzymes include enzymes in the flavonoid pathway such as phenylalanine ammoniolyase (PAL), cinnamate-4-hydroxylase (C4H), 4-coumarate-CoA ligase (4CL), chalcone synthase (CHS) and prenyltransferase.

2.7 Prenyltransferase

2.7.1 Prenyltransferase classification

Prenyltransferase is a reaction process of transferring the prenyl moiety to a wide range of prenyl acceptors such as other prenyl moiety, proteins, phenolic or aromatic compounds (Brandt et al., 2009). Prenyltransferase that transfers the prenyl moiety on aromatic compounds are termed as aromatic prenyltransferases, which play a major role

in the diversification of most aromatic secondary metabolites (Yazaki et al., 2009). There are two major pathways involved in the biosynthesis of prenylated aromatic compounds; the shikimate pathway and isoprenoid pathway (Figure 2.11). The isoprenoid pathway derived from either mevalonate or methyl erythritol phosphate (MEP) pathways provide the prenyl chain moiety donor whereas, the shikimate or polyketide pathway provide an aromatic compound acceptor (Yazaki et al., 2009).

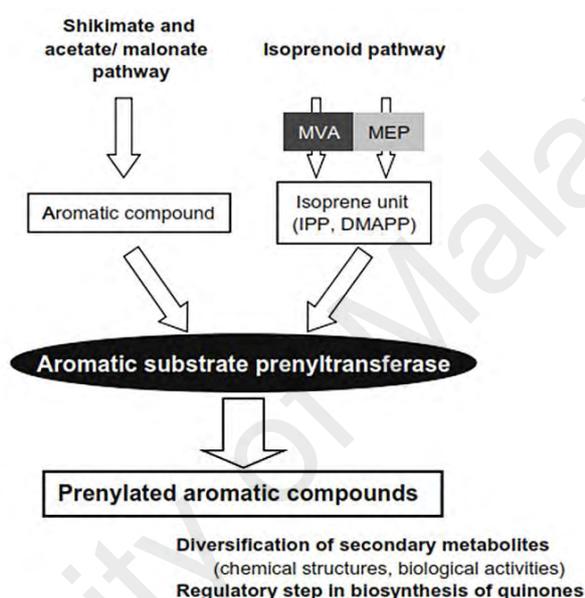


Figure 2.11: Coupling reaction of the shikimate/polyketide and isoprenoid pathways.

The isoprenoid pathway consists of two pathways, the mevalonate (MVA) pathway localized in the cytosol, and the methyl erythritol phosphate (MEP) pathway localized in the plastid. Most plant prenyltransferases responsible for prenylated aromatic compounds are membrane-bound proteins. IPP, isopentenyl diphosphate (Yazaki et al., 2009).

Aromatic prenyltransferases can be divided into two types; soluble-type prenyltransferase and membrane-bound prenyltransferase (Brandt et al., 2009). Membrane-bound prenyltransferase contain an aspartate rich motif in their sequence that conserved among the Mg-dependent prenyltransferase family (Brandt et al., 2009; Yazaki et al., 2009). Membrane-targeting signal can be detected at the N-terminus of the sequence (Yazaki et al., 2009). Research has been carried out and the enzyme has been

found in the mitochondria, endoplasmic reticulum and plastid. For instance, prenyltransferase is located in the inner membrane of mitochondria in *Oryza sativa* and *Arabidopsis thaliana* (Ohara et al., 2006; Okada et al., 2004), endoplasmic reticulum in *Lithospermum erythrorhizon* (Yazaki et al., 2002) and plastid in *Glycine max* and *Sophora flavescens* (Akashi et al., 2008; Sasaki et al., 2008).

Plant membrane-bound prenyltransferases can be further classified into two groups based on the aromatic compounds that they catalyze, namely; p-hydroxybenzoate (PHB) prenyltransferases and homogentisate (HG) prenyltransferases. Prenyltransferase that accept p-hydroxybenzoate as aromatic compound are involved in shikonin and ubiquinone biosynthesis. For instance, LePGT-1 and LePGT-2 are involved in the biosynthesis of shikonin using specific geranyl diphosphate as prenyl donor (Yazaki et al., 2002) whereas OsPPT1 from rice and AtPPT1 from *Arabidopsis* are involved in biosynthesis of ubiquinone by preference of longer prenyl moiety and solanesyl diphosphate as prenyl donor respectively (Ohara et al., 2006; Okada et al., 2004). The second prenyltransferase group; HG prenyltransferases are involved in biosynthesis of vitamin E and plastoquinone biosynthesis. For instance, AtVTE2-1 and AtVTE2-2 from *Arabidopsis* are involved in vitamin E, tocopherol biosynthesis plastoquinone biosynthesis respectively (Venkatesh et al., 2006).

Prenyl and aromatic compound substrate is recognized by amino acid conserved sequences in the membrane-bound prenyltransferase peptide sequence. Both PHB and HG prenyltransferases consist of similar aspartate-rich motif for prenyl moiety recognition; NDXXDXXXD and NQXXDXXXD respectively however did not show similarity in the aromatic recognition sequence (Akashi et al., 2008; Ohara et al., 2006; Okada et al., 2004; Sasaki et al., 2008; Yazaki et al., 2002). The prenyltransferase motif sequence that is responsible for recognizing PHB is GX(K/Y)STAL, while homogentisate

and flavonoids is KDXXDX(E/D)GD (Akashi et al., 2008; Ohara et al., 2006; Okada et al., 2004; Sasaki et al., 2008; Yazaki et al., 2002).

2.7.2 Plant flavonoid prenyltransferases

According to Yazaki et al. (2009), flavonoid prenyltransferases derived from HG prenyltransferases family as they share approximately 50% significant similarity and accept flavonoids as aromatic substrate. Several plant flavonoid prenyltransferases have been successfully isolated and characterized (Table 2.5). All of the identified flavonoid prenyltransferases were isolated from Leguminosae family except for CtIDT and MaIDT extracted from the Moraceae family. The first reported flavonoid-specific prenyltransferase was naringenin-8-prenyltransferase (SfN8D-1) isolated from *Sophora flavescens* (Sasaki et al., 2008). After that, SfN8D-2 and SfN8D-3 were cloned and molecularly characterized. Both SfN8D-2 and SfN8D-3 exhibited almost the same enzymatic properties as SfN8D-1 and suggested to be SfN8D-1 clone functional redundancy (Sasaki et al., 2008; Sasaki, Tsurumaru, Yamamoto, & Yazaki, 2011). In 2009, another flavonoid prenyltransferase was identified in soybean; pterocarpan 4-dimethylallyltransferase (GmG4DT) which catalyzes the biosynthesis of glyceollin (Akashi et al., 2009).

Besides, isoflavonoid-specific prenyltransferases, chalcone-specific prenyltransferases and flavone-specific prenyltransferase were also discovered. SfG6DT from *Sophora flavescens* and LaPT1 from *Lupinus albus* showed specific substrate specificity towards isoflavonoids and were responsible for genistein prenylation (Table 2.6). However, SfG6DT prenylates genistein at C-6 position on ring A, producing wighteone (Sasaki et al., 2011), while LaPT1 prenylates genistein at C-3' position on ring B, producing isowighteone (Shen et al., 2012). As for chalcone-specific

prenyltransferase, SfiLDT from *Sophora flavescens*, MaIDT from *Morus alba* and CtIDT from *Cudrania tricuspidata* strictly accepts isoliquiritigenin as the prenyl acceptor producing dimethylallyl isoliquiritigenin (Table 2.6). GuA6DT from *Glycyrrhiza uralensis* catalysed the prenylation of flavones such as apigenin at C-6 position (Li et al., 2014). Lastly, SffPT is the first flavonoid prenyltransferase that exhibit broad aromatic substrate specificity (Chen et al., 2013). Figure 2.12 shows some prenylation reaction of Sfn8DT-1, SffPT, Sfg6DT, SfiLDT, G4DT, and LaPT1.

Recently, other aromatic prenyltransferase such as coumarin prenyltransferase and phloroglucinol prenyltransferase were also discovered and molecularly characterized. PcPT from parsley and CIPT1 from *Citrus limon* were categorized as coumarin prenyltransferase as both prenylate umbiliferone (Table 2.6). However, PcPT only accepted DMAPP (Karamat et al., 2014), whereas CIPT1 only accepted GPP as prenyl donor (Munakata et al., 2014). HIPT-1, HIPT-1L and HIPT2 from hop were categorized as phloroglucinol prenyltransferase involved in β -bitter acid biosynthesis (Table 2.6). Although HIPT-1 was categorized as phloroglucinol prenyltransferase, HIPT-1 also has been found to accept naringenin chalcone as a substrate (Tsurumaru et al., 2012).

Aromatic prenyltransferase proteins are approximately 400 amino acids in length and is localized in the plastid (Table 2.6). The substrate specificity analysis showed that most of the flavonoid prenyltransferase isolated exhibited strict substrate specificity towards dimethylallyl diphosphate (DMAPP) moiety except for Sfg6DT, SffPT, GuA6DT and CIPT1. Sfg6DT also accepted geranyl diphosphate (GPP) and traces of farnesyl diphosphate (FPP) as prenyl donor besides DMAPP (Sasaki et al., 2011). SffPT and GuA6DT could accept both DMAPP and GPP as prenyl donors (Chen et al., 2013; Li et al., 2014). Among all, CIPT1 was the first aromatic prenyltransferase which specifically recognized GPP as prenyl donor (Munakata et al., 2014).

Table 2.5: Plant flavonoid prenyltransferases in NCBI database.

No	Plant Name	Family	Enzyme	Plant flavonoid prenyltransferase	Accession No.	Reference
1	<i>Sophora flavescens</i>	Leguminosae	SfN8DT-1	Naringenin 8-dimethylallyltransferase	BAG12671.1	(Sasaki et al., 2008)
2	<i>Sophora flavescens</i>	Leguminosae	SfN8DT-2	Naringenin 8-dimethylallyltransferase	BAG12673.1	(Sasaki et al., 2008)
3	<i>Sophora flavescens</i>	Leguminosae	SfN8DT-3	8-dimethylallyltransferase	BAK52289.1	(Sasaki et al., 2011)
4	<i>Sophora flavescens</i>	Leguminosae	SfG6DT	Genistein 6-dimethylallyltransferase	BAK52291.1	(Sasaki et al., 2011)
5	<i>Sophora flavescens</i>	Leguminosae	SfiLDT	Isoliquiritigenin dimethylallyltransferase	BAK52290.1	(Sasaki et al., 2011)
6	<i>Sophora flavescens</i>	Leguminosae	SfFPT	Flavonoid prenyltransferase	AHA36633.1	(Chen et al., 2013)
7	<i>Glycine max</i>	Leguminosae	GmG4DT	Pterocarpan 4-dimethylallyltransferase	NP_00123599 0.1	(Akashi et al., 2009)
8	<i>Lupinus albus</i>	Leguminosae	LaPT1	Genistein 3'-dimethylallyltransferase	AER35706.1	(Shen et al., 2012)
9	<i>Glycyrrhiza uralensis</i>	Leguminosae	GuA6DT	Flavone prenyltransferase	AIT11912.1	(Li et al., 2014)
10	<i>Cudrania tricuspidata</i>	Moraceae	CtiDT	Isoliquiritigenin 3'-dimethylallyltransferase	AJD80983.1	(Wang et al., 2014)
11	<i>Morus alba</i>	Moraceae	MaIDT	Isoliquiritigenin 3'-dimethylallyltransferase	AJD80982.1	(Wang et al., 2014)

Table 2.6: Summary of molecular characterization of plant aromatic prenyltransferases.

Category	Enzyme	Species	Family	Peptide length	Localization	Prenyl chain specificity	Aromatic compound specificity	References
Flavonoid prenyltransferase	SfN8DT-1	<i>Sophora flavescens</i>	Leguminosae	410	Plastid	DMAPP	Flavonone: Naringenin, Liquiritigenin, Hesperetin	(Sasaki et al., 2008)
Flavonoid prenyltransferase	SfG6DT	<i>Sophora flavescens</i>	Leguminosae	407	Plastid	DMAPP, GPP, FPP	Isoflavonoid: Genistein, Biochanin A	(Sasaki et al., 2011)
Flavonoid prenyltransferase	SfILDT	<i>Sophora flavescens</i>	Leguminosae	391	Plastid	DMAPP	Chalcone Isoliquiritigenin	(Sasaki et al., 2011)
Flavonoid prenyltransferase	SfFPT	<i>Sophora flavescens</i>	Leguminosae	407	Plastid	DMAPP & GPP	Flavonoids broad substrate specificity	(Chen et al., 2013)
Flavonoid prenyltransferase	GmG4DT	<i>Glycine max</i>	Leguminosae	409	Plastid	DMAPP	Pterocarpan (-) - Glycinol	(Akashi et al., 2009)
Flavonoid prenyltransferase	LaPT1	<i>Lupinus albus</i>	Leguminosae	408	Plastid	DMAPP	Isoflavonoid Genistein, 2'-hydroxygenistein	(Shen et al., 2012)
Flavonoid prenyltransferase	GuA6DT	<i>Glycyrrhiza uralensis</i>	Leguminosae	412	Plastid	DMAPP & GPP	Flavone Apigenin	(Li et al., 2014)
Flavonoid prenyltransferase	MaIDT	<i>Morus alba</i>	Moraceae	402	Plastid	DMAPP	Chalcone Isoliquiritigenin	(Wang et al., 2014)

Table 2.6, continued

Flavonoid prenyltransferase	CtIDT	<i>Cudrania tricuspidata</i>	Moraceae	398	Plastid	DMAPP	Chalcone Isoliquiritigenin	(Wang et al., 2014)
Phloroglucinol prenyltransferase	HIPT-1	<i>Humulus lupulus</i> L. hop 'Kirin II'	Cannabaceae	411	Plastid	DMAPP	Phloroglucinol derivatives PIVP, PIBP, PMBP Chalcone Naringenin chalcone	(Tsurumaru et al., 2010; Tsurumaru et al., 2012)
Phloroglucinol prenyltransferase	HIPT-1L and HIPT2	<i>Humulus lupulus</i> L. hop cultivar 'Nugget'	Cannabaceae	414 and 408	Plastid	DMAPP	Phloroglucinol derivatives PIVP, PIBP, PMBP Prenylated acylphloroglucinols	(Li et al., 2015)
Coumarin prenyltransferase	CIPT1	<i>Citrus limon</i>	Rutaceae	407	Plastid	GPP	Coumarin Umbelliferone	(Munakata et al., 2014)
Coumarin prenyltransferase	PcPT	<i>Petroselinum crispum</i> (Parsley)	Apiaceae	400	Plastid	DMAPP	Coumarin Umbelliferone	(Karamat et al., 2014)

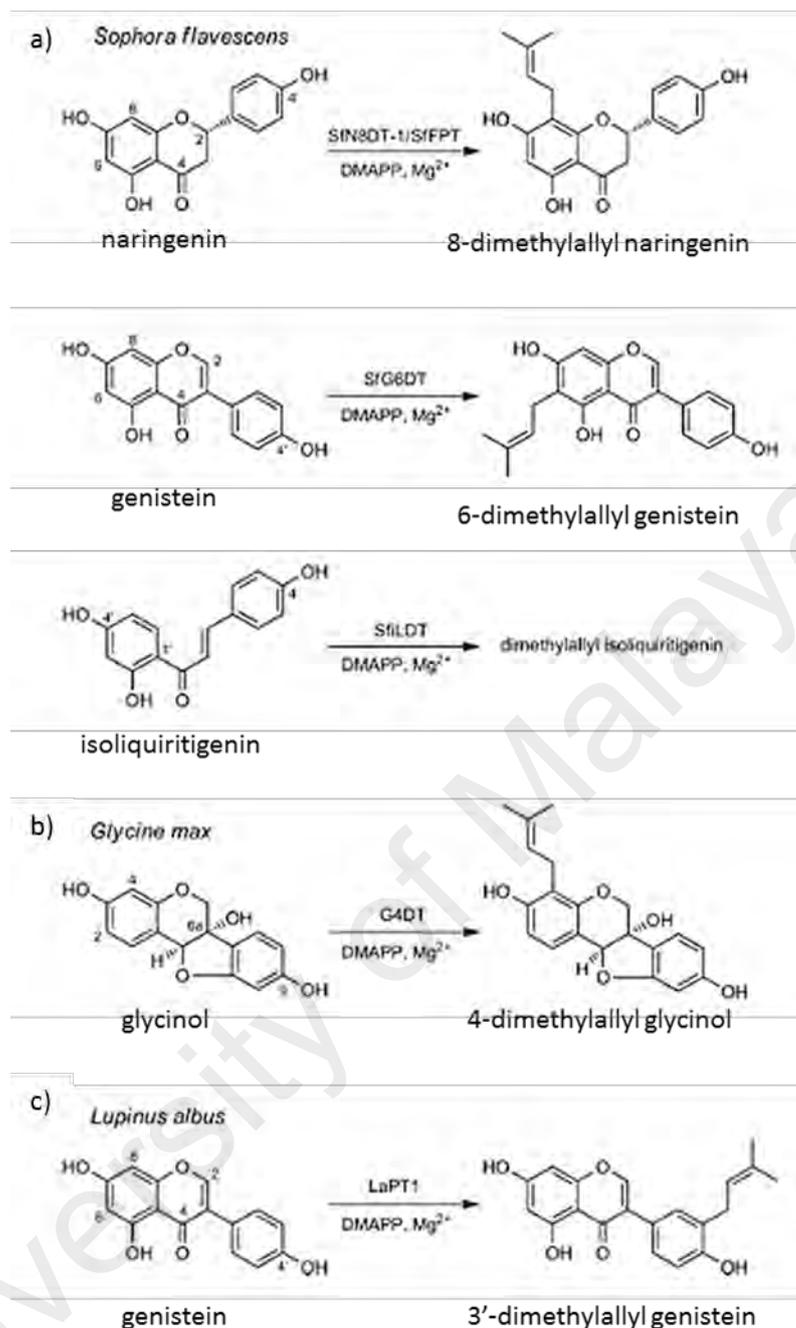


Figure 2.12: The flavonoid prenyltransferases from the Leguminosae (Li et al., 2014).

a) SfN8DT-1/SfFPT from *S. flavescens* catalyzes the 8-prenylation of naringenin, SfG6DT catalyzes the 6-prenylation of genistein, and SfILDt catalyzes the prenylation of isoliquiritigenin. b) G4DT from soybean (*G. max*) catalyzes the 4-prenylation of glycinol. c) LaPT1 from *L. albus* catalyzes the 3'-prenylation of genistein.

From the phylogenetic tree analysis constructed by Wang et al. (2014), flavonoid prenyltransferases evolved from HG prenyltransferases (Figure 2.13). Interestingly, leguminous flavonoid prenyltransferases and moraceous flavonoid prenyltransferases

branched into two different clusters. The leguminous flavonoid prenyltransferases evolved from the HG prenyltransferases that synthesize vitamin E, whereas moraceae flavonoid prenyltransferases evolved from the HG prenyltransferases that synthesize plastoquinone. The results implies that flavonoid prenyltransferases evolved independently in the two plant lineages (Wang et al., 2014).

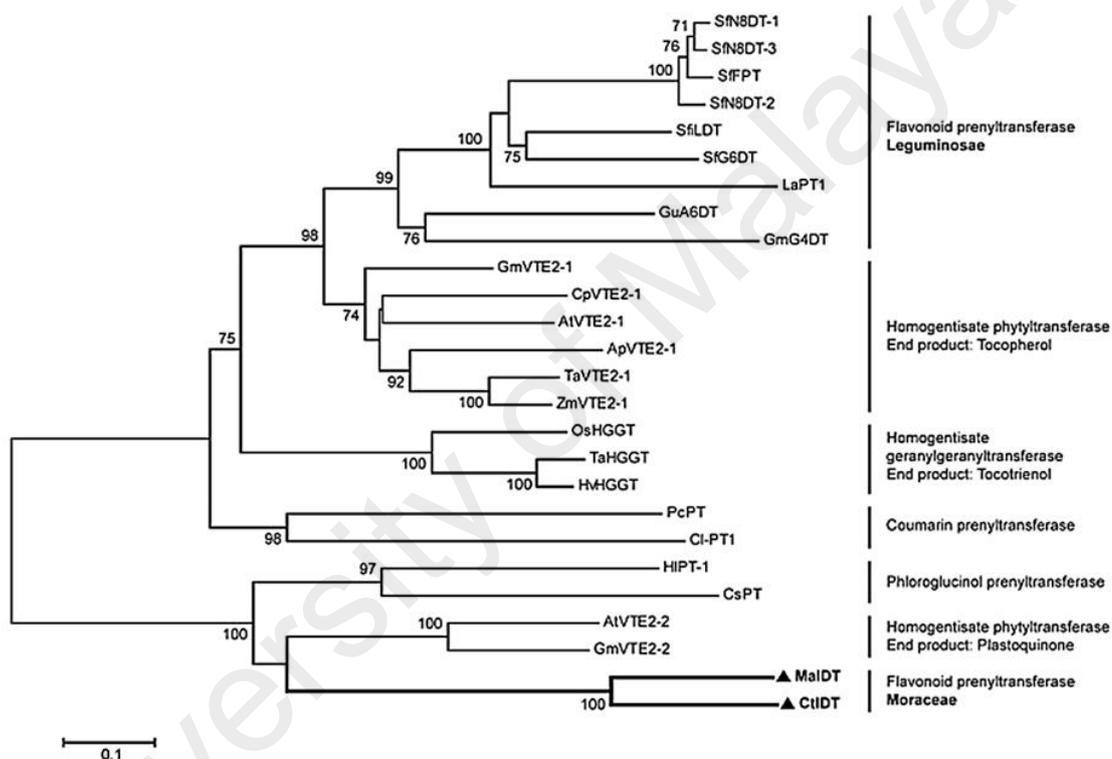


Figure 2.13: Phylogenetic tree relationship between plant aromatic prenyltransferases with homogentisate prenyltransferases generated by Wang et al. (2014).

2.7.3 Prenylated flavonoids in plants

Prenylated flavonoids are derived from flavonoids which attached a flavonoid skeleton with a lipophilic prenyl side chain. In plants, prenylated flavonoids are less abundant in nature as compared to flavonoids (Yang et al., 2015). Prenylation has been detected in different classes of flavonoids which include isoflavones, pterocarpan, flavanones,

flavones and chalcones (Barron & Ibrahim, 1996). Unlike flavonoids which are present in almost all plants, prenylated flavonoids are not widely distributed in plants. To date, approximately 1000 prenylated flavonoids have been identified from 37 plant genera (Yang et al., 2015). Most of prenylated flavonoids are found in plant families of Cannabaceae, Guttiferae, Leguminosae, Moraceae, Rutacea and Umbelliferae (Barron & Ibrahim, 1996; Yang et al., 2015).

Prenylated flavonoids have diverse structures depending on the flavonoid skeleton, length of prenyl side-chain, prenylation position on the aromatic rings and further modifications of the prenyl moiety (Yazaki et al., 2009). The length of prenyl side chain was based on the size of the carbon; C5 (isopentenyl), C10 (geranyl), C15 (farnesyl) and C20 (geranylgeranyl) (Figure 2.14). The modifications of prenyl side chain include oxidation, reduction, dehydration and cyclization (Barron & Ibrahim, 1996). In general, more C-prenylation on flavonoids occurs in plants compared to O-prenylation. C-prenylation take place more frequently on ring A at C-6 or C-8 and ring B at C-3' or C-5' (Barron & Ibrahim, 1996). Figure 2.15 shows the prenylation patterns on flavonoids and 3,3-dimethylallyl group is the most common pattern identified (Barron & Ibrahim, 1996; Yang et al., 2015).

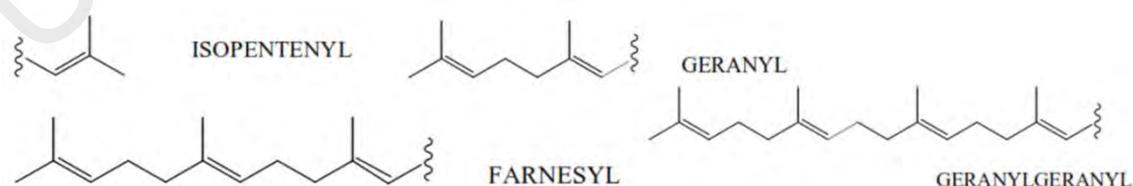


Figure 2.14: The prenyl side chain (Alhassan et al., 2014).

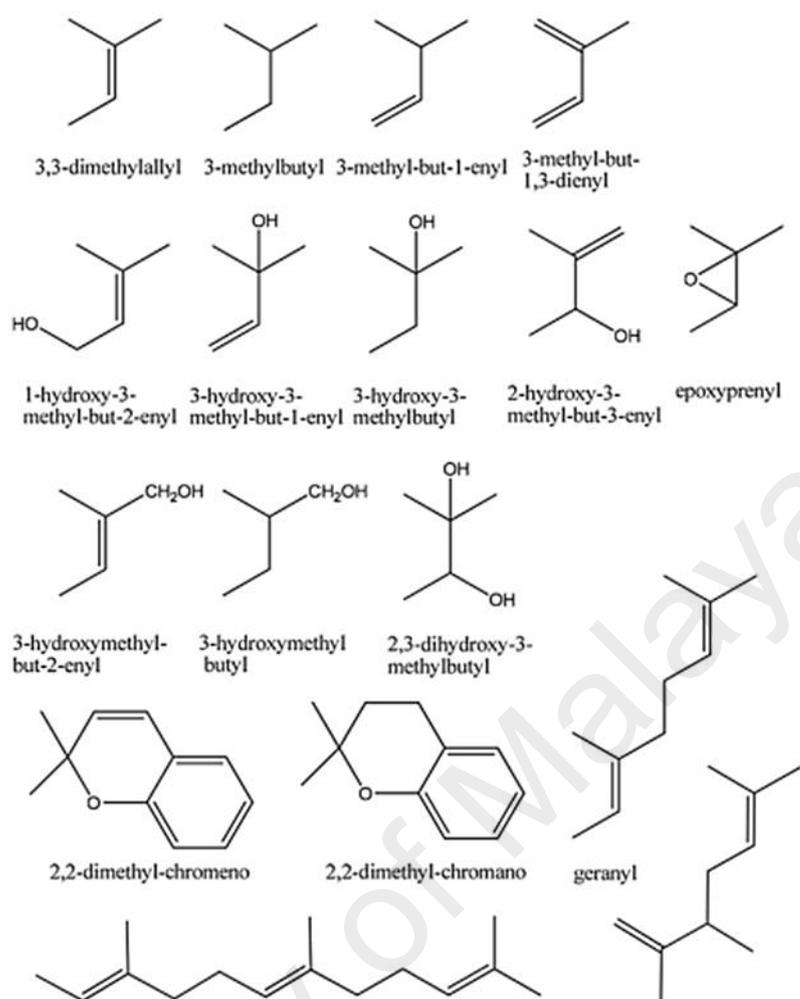


Figure 2.15: Prenylation patterns encountered on flavonoids (Yang et al., 2015).

Many prenylated flavonoids have been identified as bioactive compounds in medicinal plants as they possess medicinal properties such as anti-cancer, antioxidant, anti-allergic and anti-inflammatory. In general, prenylation improve the flavonoids bioactivities by increasing the lipophilicity of flavonoids, resulting a higher affinity to biological membranes and better interaction with target proteins (Yang et al., 2015). For instance, desmethylxanthohumol, a prenylated chalcone had shown high antioxidant activities as compared to the non-prenylated parent compound, chalconaringenin (Miranda et al., 2000). Moreover, prenylation at C-8 on flavonoids strongly enhanced the biological activity. For instance, licoflavone C and isobavachin strongly enhances the toxicity of

apigenin and liquiritigenin in rat H4IIE hepatoma and C6 glioma cells (Wätjen et al., 2007). Besides licoflavone C and 8-prenylnaringenin were also reported to be potent phytoestrogen (Garritano et al., 2005; Milligan et al., 2000). Xanthohumol, a C-8 prenylated chalcone in hop had shown anti-proliferative activity on human prostate cancer cell lines by inhibiting the growth of PC-3 and DU145 (Delmulle et al., 2006). It also inhibits the growth of human colon cancer (HT-29), breast cancer (MCF-7), and ovarian cancer (A2780) cell lines (Miranda et al., 1999).

2.7.4 Prenylated flavonoid in *B. rotunda*

There are more than 10 prenylated flavonoids including prenylated chalcones and flavanones identified in *B. rotunda* (Morikawa et al., 2008; Win et al., 2007). Figure 2.16 shows the structures of the prenylated flavonoids found in *B. rotunda*. As reported by Win et al. (2007), geranylated chalcone and geranylated flavanone showed higher cytotoxicity towards human pancreatic PANC-1 cancer cells compared to their parent compounds. For instance, chalcone, 3'-geranylcardamonin showed stronger cytotoxic activity compared to cardamonin (Win et al., 2007). Similarly, flavanone, 8-geranylpinostrobin and 6-geranylpinostrobin showed stronger cytotoxic activity compared to pinostrobin (Win et al., 2007). However, prenylated chalcone exhibit stronger cytotoxic activity compared to prenylated flavanone (Win et al., 2007).

In general, cyclohexenyl chalcone derivatives were more active than the other compounds, especially panduratin A. Although panduratin A has similar basic cyclohexenyl back bone with 4-hydroxypanduratin A and isopanduratin A2, the presence of a methoxyl group at C-4 and hydroxyls at C-2 and C-6 in cyclohexenyl chalcones seem to be important for activity (Win et al., 2007). Panduratin A was reported to exhibit more medicinal properties as compared to other cyclohexenyl chalcone. For instance,

panduratin A exhibited stronger *in vitro* preferential cytotoxic (PC₁₀₀) activity against human pancreatic cancer cells PANC-1 as compared to isopanduratin A2 and 4-hydroxypanduratin A (Win et al., 2007). Other biological activities of panduratin A are shown in Table 2.4.

Both panduratin A and 4-hydroxypanduratin A were also reported to exhibit stronger anti-inflammatory by inhibiting the nitric oxide (NO), prostaglandin E₂ (PGE₂) and tumour necrosis factor-alpha (TNF- α), as compared to panduratin C (Tewtrakul et al., 2009; Yun et al., 2003). Although panduratin C shares the same carbon skeleton as panduratin A, it has hydroxyl group at position C-4", which eventually reduces the compound's anti-inflammatory activity (Tewtrakul et al., 2009). Due to prenyl side chain on panduratin A structure, it exhibits stronger antioxidant activities compared to its parent compound, pinostrobin chalcone by inhibiting the L-glutamate toxicity in N18-RE-105 cells (Shindo et al., 2006). Beside panduratin A, 4-hydroxypanduratin A was also reported to exert this neuroprotective effects (Shindo et al., 2006). These compounds also exhibit anti-aging properties by inhibiting the UV-induced matrix metalloproteinase-1 (MMP1) expression (Shim et al., 2009; Shim et al., 2008a).

Even though panduratin A exhibits numerous medicinal properties, this compound or other prenylated cyclohexanyl chalcones are not easily isolated due to their low abundance in nature and their biosynthetic pathway is also not yet understood. Thus, establishing the biosynthetic pathway through elucidating the enzymatic genes involved will open up the possibilities to increase the yield of panduratin A via approaches such as metabolic engineering. Other than enzymes, regulating proteins are the key players in regulating the pathway by controlling enzyme production. Among the most important regulating proteins are the transcription factors.

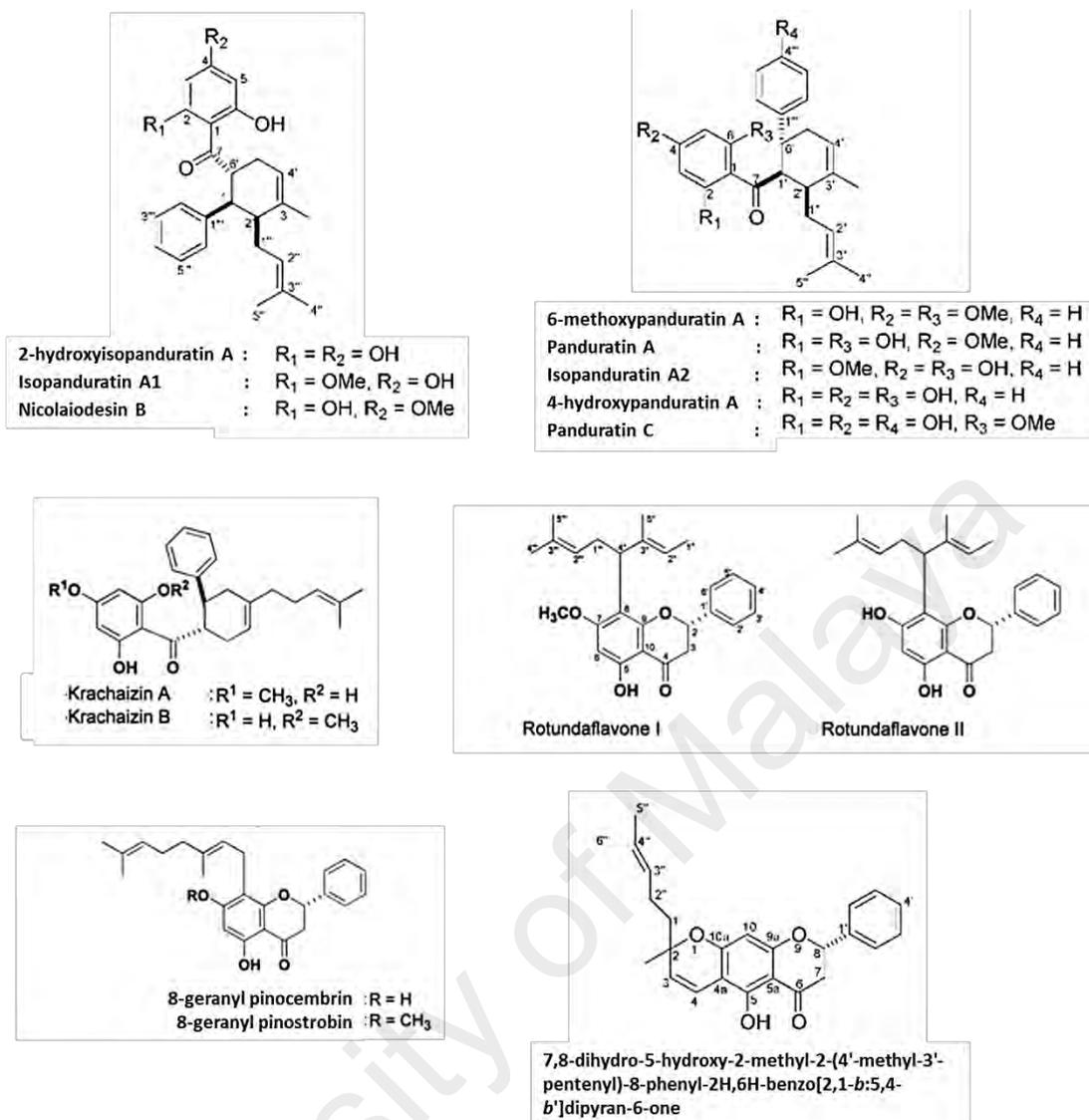


Figure 2.16: The structure of prenylated flavonoids found in *B. rotunda* (Morikawa et al., 2008; Win et al., 2007).

2.8 Plant MYB transcription factors

2.8.1 MYB transcription factors classification in plants

Transcription factors (TFs) are regulatory proteins that control the complex network of metabolic pathways. TFs are sequence specific DNA-binding proteins that recognize specific cis-regulatory sequences in the promoter regions of target genes and modulate their expression either through activation or repression depending on tissue types, developmental stages and/or environmental cues (Du et al., 2009; Vom Endt, Kijne, &

Memelink, 2002). However, some of the TFs do not directly bind to the DNA sequences, but form complexes with other co-factors or proteins to regulate the expression of the target genes (Yang et al., 2012).

Several families of TFs have been identified to participate in controlling the secondary metabolic pathways through regulating the expression of enzyme-coding genes in the pathways (Yang et al., 2012). Transcription factors have been classified into different families based on the conserved DNA-binding domains (Pabo & Sauer, 1992). Helix-turn-helix, zinc finger, leucine zipper, scissors and MADS cassette are some examples of structural motifs that are capable of binding target DNA sequences (Du et al., 2009). Several transcription factor families have been described to regulate the secondary metabolism such as MYB, bHLH, AP2/ERF, WRKY, Zinc finger, DOF, SPL, NAC and bZIP (Yang et al., 2012). Among all, MYB transcription factors comprise one of the largest transcription factor families in plants.

The first MYB gene identified was the 'oncogene' v-Myb gene from avian myeloblastosis virus (AMV) (Klempnauer et al., 1982). Based on sequence comparison analysis, the gene appears to have originated from a vertebrate gene and has three members, namely c-Myb, A-Myb and B-Myb (Weston, 1998). The first plant MYB gene identified was C1 from *Zea mays*, which regulates anthocyanin synthesis (Paz-Ares et al., 1987). The common feature of MYB transcription factor proteins is the presence of highly conserved MYB DNA-binding domain (Ambawat et al., 2013).

The MYB DNA-binding domain generally consists of one to four imperfect repeats (R1, R2 and R3); with 50 – 53 amino acids where each repeat forms three α -helices (Figure 2.17). The second and third helices of each repeat form a helix-turn-helix (HTH) structure. The third helix in the HTH structure is called the recognition helix, which is essential to recognize and bind to specific DNA sequences by forming hydrogen bonds

with bases located in the major groove of the DNA double helix. The second helix stabilizes the overall configuration through hydrogen interactions with the recognition helix (Becker et al., 2003; Ogata et al., 1992). Three regularly spaced tryptophan residues typically located in the MYB repeats, which form a tryptophan cluster in three-dimensional HTH structure (Ogata et al., 1992). In plants, the first tryptophan of R3 is replaced by phenylalanine or isoleucine (Ambawat et al., 2013).

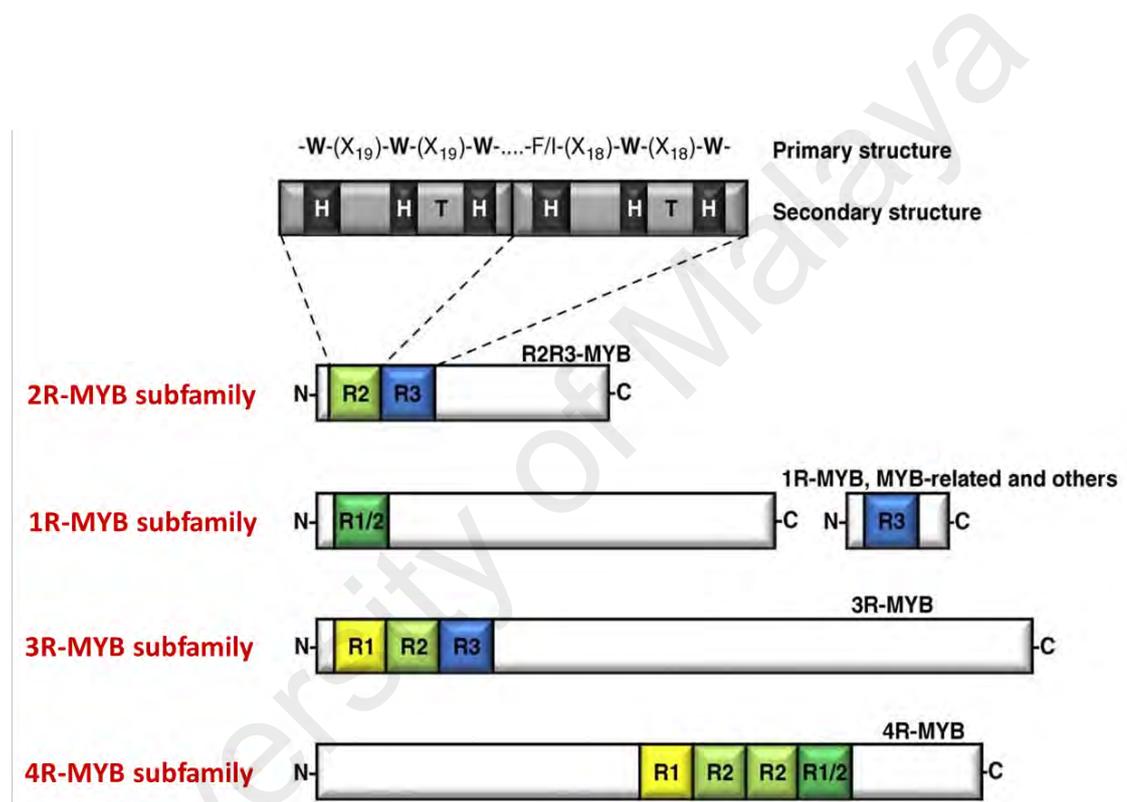


Figure 2.17: Schematic diagram of plant MYB transcription factor classes.

Illustration showing different plant MYB protein classes depending on the number of adjacent MYB repeats (R). The primary and secondary structures of a typical R2R3-MYB are indicated. H, helix; T, turn; W, tryptophan; X, any amino acid (Dubos et al., 2010).

The plant MYB repeats - R1, R2 or R3 are named according to their similarity with the three repeats (R1, R2 and R3) in protolytic c-Myb protein. The MYB transcription factor family can be classified into four subfamilies based on the number of adjacent repeats in the binding domains, namely 1R-, 2R-, 3R- and 4R-MYB subfamilies (Dubos

et al., 2010). The first MYB subfamily, the heterogenous 1R-MYB subfamily, also called MYB-related group comprises of an intact single or partial MYB repeat (R1/2, R3-MYB). 1R-MYB subfamily is widely distributed in plants and is the second largest group of the MYB family. There are 68 members of 1R-MYB in Arabidopsis, 47 members in grapevine and 64 members in rice (Table 2.7). 1R-MYB transcription factors are involved in morphogenesis (Simon et al., 2007), secondary metabolism (Dubos et al., 2008; Matsui et al., 2008), circadian clock control (Lu et al., 2009), response to phosphate starvation (Bustos et al., 2010) and also flower and fruit development (Barg et al., 2005).

Next, the second MYB subfamily is 2R-MYB subfamily also known as R2R3-MYB which is the largest group of plant MYB family consisting of R2 and R3 repeats. There are 138 R2R3-MYB members in Arabidopsis, 126 in rice and 244 in soybean (Table 2.7). R2R3-MYB transcription factors are further classified into 28 subgroups based on the conserved motifs at the C-terminus (Dubos et al., 2010; Liu et al., 2015; Stracke et al., 2001). Therefore, the regulatory activity of the proteins is dependent on the C-terminal region. These R2R3-MYB transcription factors are involved in primary and secondary metabolism, plant development, determination of cell fate and identity, hormone signal transduction, and response to abiotic and biotic stresses (Ambawat et al., 2013).

Subsequently, the third MYB subfamily, the 3R-MYB contains R1, R2 and R3 repeats (R1R2R3-MYB), which have been identified in most eukaryotes. In contrast to their predominant roles in cell cycle control, relatively few 3R-MYB proteins are found in plants (Ito, 2005). For instance, there are five 3R-MYB proteins in Arabidopsis, grapevine and rice (Table 2.7). Lastly, the smallest MYB subfamily is the 4R-MYB which contains four R1/R2-like repeats. Only a single or two of 4R-MYB proteins have been found in plant species (Table 2.7). Little is known about the functions of the 4R-MYB subfamily in plants.

Table 2.7: Distribution MYB transcription factor subfamilies within selected plant species.

Plant species	Number				References
	1R-MYB	2R-MYB	3R-MYB	4R-MYB	
<i>Arabidopsis thaliana</i>	68	138	5	2	(Du et al., 2013; Katiyar et al., 2012)
Grapevine (<i>Vitis vinifera</i>)	47	118	5	1	(Du et al., 2013; Matus et al., 2008; Wilkins et al., 2009)
Orange (<i>Citrus sinensis</i>)	90	87	1	1	(Hou et al., 2014)
Rice (<i>Oryza sativa</i>)	64	126	5	1	(Chen et al., 2006)
Soybean (<i>Glycine max</i>)	127	244	6	2	(Du et al., 2013; Du et al., 2012)

2.8.2 MYB transcription factor functions in plants

In plants, the MYB transcription factor family plays a wide variety of important roles in regulating the plant specific processes. The MYB proteins might act either as transcription activator or transcription repressor or as both transcription activator and repressor. The functions of MYB proteins have been identified through genetic and molecular analysis in numerous plant species including *Arabidopsis*, rice, maize (*Zea mays*), petunia (*Petunia hybrida*), soybean (*Glycine max*), snapdragon (*Antirrhinum majus*), grapevine (*Vitis vinifera*) and apple (*Malus domestica*).

The MYB family has been found to be involved in controlling cell morphogenesis and differentiation, response to environmental stresses, phytohormones and regulating phenylpropanoid metabolism. Some of the MYB proteins are expressed in most tissues, while the others are only expressed in specific tissues at certain developmental stages. For instance, AtMYB21 is only expressed in flower buds (Shin et al., 2002), AtMYB26 in anthers (Steiner-Lange et al., 2003) and cotton GhMYB109 is specifically expressed in cotton fiber initials and elongation (Suo et al., 2003). Accordingly, GhMYB7 and GhMYB9 are developmentally regulated in cotton fibers (Hsu et al., 2005).

Cell fate determination and morphogenesis in petal, root hair, trichome and seed coat are critical steps in plant development. For instance, the size and shape of the flower petals are important to attract the pollinators from a distance. Pollinators showed preference for flowers with conical shaped epidermal cells. The shape increase the proportion of incident light that enters the epidermal cells and enhancing light absorption by the floral pigments resulting a brighter petal colour (Noda et al., 1994). MYB transcription factor proteins such as PhMYB1 from *Petunia hybrida*, AmMYBL2 from snapdragon, *Antirrhinum majus* and AtMYB16 from *Arabidopsis thaliana* share the same function involves in petal morphogenesis by controlling the shape of petal epidermal cells (Baumann et al., 2007). In addition, besides controlling the conical cell development in the petal hinge epidermis, AmMYBJ1 is also reported to be involved in controlling trichome and mesophyll cell morphogenesis in *Antirrhinum majus* (Perez-Rodriguez et al., 2005).

In trichome and root hair development, the MYB proteins form complexes with other transcription factors such as bHLH and WD40-repeat (WDR) proteins for transcriptional regulation. This type of regulation also termed as combinatorial control. In trichome development, activator complex consist of an R2R3 MYB-type protein, GLABRA1 (GL1) (Oppenheimer et al., 1991), a WD40-repeat protein, TRANSPARENT TESTA GLABRA1 (TTG1) (Galway et al., 1994) and bHLH transcription factor, GLABRA (GL3) or ENHANCER OF GLABRA3 (EGL3) (Payne et al., 2000; Zhang et al., 2003). The TTG1-GL3/EGL3-GL1 activator complex induces the expression of GLABRA2 (GL2), a homeodomain protein (Rerie et al., 1994) which is required for trichome initiation.

In root hair development, GL1 is replaced by another R2R3 MYB-type protein, WAREWOLF (WER) in the activator complex (Lee & Schiefelbein, 1999). The TTG1-

GL3/EGL3-WER activator complex also induces the expression of a homeodomain protein, GL2 and R3-MYB protein. In contrast to trichome initiation system, the R3 MYB functions as positive regulator in root hair development by suppressing the GL2 expression, resulting in root hair cell differentiation (Tominaga-Wada & Wada, 2014). GL1 and WER are categorized under subgroup 15 of R2R3 MYB proteins. Similar to R3 MYB, these proteins also contain the bHLH interaction conserved motif [D/E]Lx2[R/K]x3Lx6Lx3R in their R3 repeats (Zimmermann et al., 2004).

Plant development, growth and productivity are often affected by various environmental stresses. Regulation of MYB family has been shown to be one of essential strategies for plant to tolerate the stresses. Several MYB proteins have been found regulated in plants in response to drought such as AtMYB96, AtMYB15 and AtMYB2 in Arabidopsis (Ding et al., 2009; Seo et al., 2009; Urao et al., 1996), MdoMYB121 in apple (Cao et al., 2013), VvMYB60 in grapevine (Galbiati et al., 2011) and StMYB1R-1 in potato (Shin et al., 2011).

High salinity is also another major abiotic stress which has caused ion imbalance and water deficiency in plants. Several MYB proteins are regulated in plants to tolerate salt stress. For instance, AtMYB20 enhances salt resistance by repression of abscisic acid (ABA) signalling negative regulators, the type 2C serine/threonine protein phosphatases (PP2Cs) expression (Cui et al., 2013). AtMYB73 acts as a negative regulator of salt responses by suppression of *SOS1* and *SOS3* genes, the salt overly sensitive (SOS) transcripts (Kim et al., 2013).

In addition to abiotic stress, plants also regulate MYB proteins to confer biotic stresses for plant survival. In Arabidopsis, AtMYB102 plays a role in defense against the insect herbivores *Pieris rapae* (De Vos et al., 2006); whereas AtMYB44 increased resistance towards the bacterial pathogen *Pseudomonas syringae* by enhancing the expression of

pathogenesis-related (PR) genes through the salicylic acid signalling pathway (Zou et al., 2013). In wheat, TaPIMP1 increased resistance against *Bipolaris sorokiniana* by regulating the defense-related genes expression (Zhang et al., 2012). Besides plant development and environmental stress, MYB transcription factors are also involved in controlling the secondary metabolite production in plants.

2.8.3 MYB transcription factors in secondary metabolic pathway in plant

Several MYB transcription factors are involved in regulation of flavonoid biosynthetic pathway such as proanthocyanidins, anthocyanins flavonols, and isoflavonoids. Genes encoding enzymes that regulate the flavonoid biosynthetic pathway are divided into two groups; namely early biosynthetic genes (EBGs) which include *CHS*, *CHI*, *F3H*, *F3'H* and *FLSI*) catalysing the flavonol biosynthesis, whereas late biosynthetic genes (LBGs) consist of *DFR*, *LDOX* and *ANR*) regulating proanthocyanidin and anthocyanin biosynthesis (Figure 2.18).

In plants, R2R3 MYB transcription factors in subgroups 4, 5, 6, 7 and 27 are found to be the major subgroup in controlling flavonoid biosynthetic pathway. Subsequently, single repeat R3 MYB proteins are also involved in the flavonoid pathway regulation. However, some of the MYB regulators do not fit this classification perfectly. Table 2.8 shows the MYB transcription factors that regulate phenylpropanoid pathway in plants. Depending on species, the transcription regulation might involve the MYB transcription factor alone or formation of MYB transcription factor with other transcription factors such as a MYB-bHLH-WD40 complex or a MYB-bHLH dimer (Hichri et al., 2011).

As mentioned previously, both structural and regulating genes are important in regulating the biosynthesis of bioactive compounds. Hence, elucidating structural and

regulating genes that are directly or indirectly involved in the flavonoid biosynthetic pathway are important initial steps for designing strategies to increase panduratin A production in *B. rotunda*. The identification can be done through sequencing of the total RNA transcript using cutting edge next generation sequencing technology.

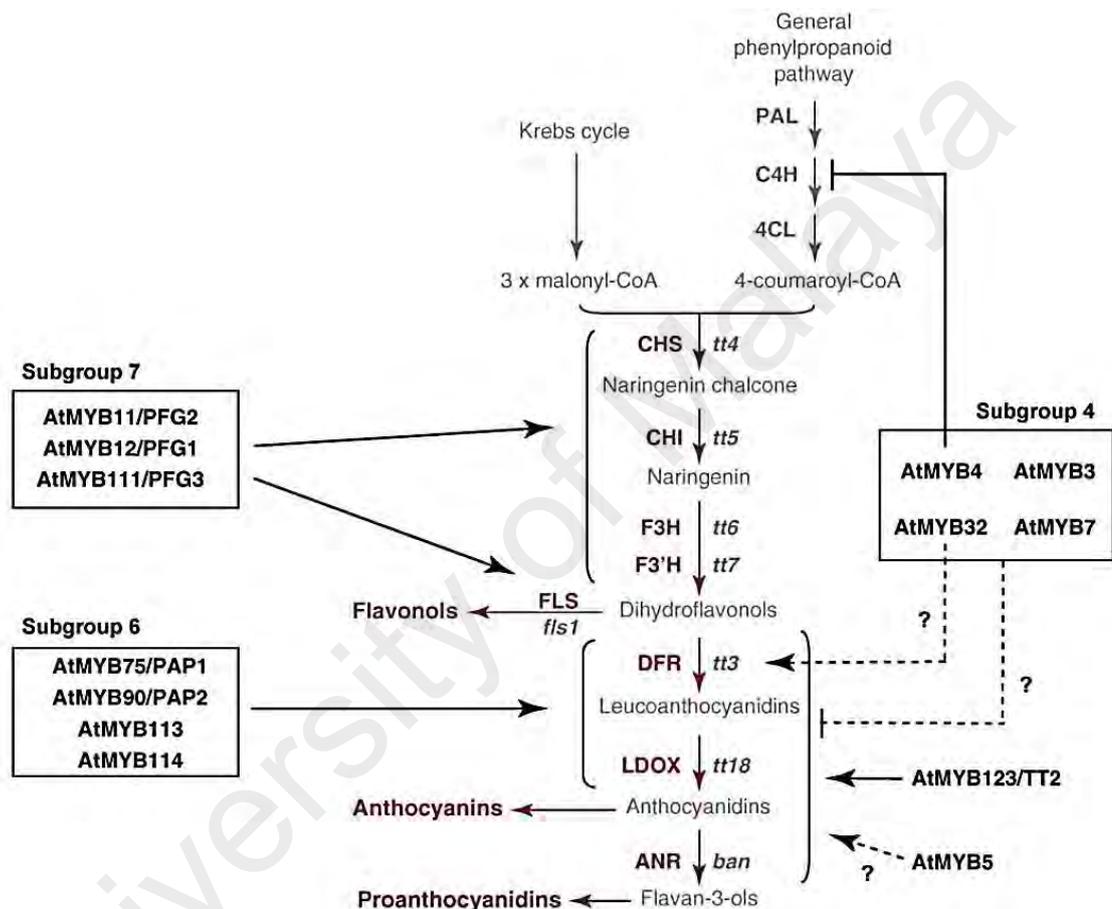


Figure 2.18: Regulation of flavonoid biosynthetic pathway by R2R3 MYB transcription factors subgroup 4, 6 and 7 in Arabidopsis (Dubos et al., 2010).

Early biosynthetic genes (EBGs); *CHS*, *CHI*, *F3H*, *F3'H* and *FLS1* indicated by purple font, whereas late biosynthetic genes (LBGs); *DFR*, *LDOX* and *ANR* indicated by red and maroon font. Abbreviations: PAL, phenylalanine ammonia lyase; C4H, cinnamate-4-hydroxylase; 4CL, 4-coumaroyl-CoA synthase; CHI, chalcone isomerase; F3H, flavonol 3-hydroxylase; F3'H, flavonol 3'-hydroxylase; FLS, flavonol synthase; DFR, dihydroflavonol-4-reductase; LDOX, leucoanthocyanidin dioxygenase; ANR, anthocyanin reductase. Enzymes are indicated in upper case letters and genetic loci in lower case italics. Dotted lines represent putative regulations.

Table 2.8: MYB transcription factors that are involved in the regulation of the phenylpropanoid pathway. This table was modified from Liu et al., 2015.

Plant species	Gene	Type	Subgroup	Compound	Function	Reference
<i>Arabidopsis thaliana</i>	AtMYB60	R2R3	1	Anthocyanin	Repressor	(Park et al., 2008)
<i>Arabidopsis thaliana</i>	AtMYB14	R2R3	2	Stilbene	Activator	(Höll et al., 2013)
<i>Arabidopsis thaliana</i>	AtMYB15	R2R3	2	Stilbene	Activator	(Höll et al., 2013)
Lotus (<i>Lotus japonicus</i>)	LjMYB14	R2R3	2	Isoflavonoid	Activator	(Shelton et al., 2012)
<i>Arabidopsis thaliana</i>	AtMYB58	R2R3	3	Monolignol	Activator	(Zhou et al., 2009)
<i>Arabidopsis thaliana</i>	AtMYB63	R2R3	3	Monolignol	Activator	(Zhou et al., 2009)
Apple (<i>Malus domestica</i>)	MdMYB3	R2R3	4	Anthocyanin/ flavonol	Activator	(Vimolmangkang et al., 2013)
<i>Arabidopsis thaliana</i>	AtMYB4	R2R3	4	Sinapate ester	Repressor	(Jin et al., 2000)
<i>Arabidopsis thaliana</i>	AtMYB32	R2R3	4	Monolignol	Repressor	(Preston et al., 2004)
Cider gum (<i>Eucalyptus gunnii</i>)	EgMYB1	R2R3	4	Monolignol	Repressor	(Legay et al., 2007; Legay et al., 2010)
Danshen (<i>Salvia miltiorrhiza</i>)	SmMYB39	R2R3	4	Phenolic acid	Repressor	(Zhang et al., 2013)
Garden chrysanthemum (<i>Chrysanthemum morifolium</i>)	CmMYB1	R2R3	4	Monolignol	Repressor	(Zhu et al., 2013)
Grapevine (<i>Vitis spp.</i>)	VvMYBC2-L1	R2R3	4	Proanthocyanidin	Repressor	(Huang et al., 2014)
Leucaena (<i>Leucaena leucocephala</i>)	LlMYB1	R2R3	4	Monolignol	Repressor	(Omer et al., 2013)
Maize (<i>Zea Mays</i>)	ZmMYB31	R2R3	4	Monolignol	Repressor	(Fornalé et al., 2010)
Maize (<i>Zea Mays</i>)	ZmMYB42	R2R3	4	Monolignol	Repressor	(Sonbol et al., 2009)
Petunia (<i>Petunia hybrida</i>)	PhMYB27	R2R3	4	Anthocyanin	Repressor	(Albert et al., 2014)
Rabbit-foot clover (<i>Trifolium arvense</i>)	TaMYB14	R2R3	4	Proanthocyanidin	Activator	(Hancock et al., 2012)
Snapdragon (<i>Antirrhinum majus</i>)	AmMYB308	R2R3	4	Monolignol	Repressor	(Tamagnone et al., 1998)

Table 2.8, continued

Snapdragon (<i>Antirrhinum majus</i>)	AmMYB330	R2R3	4	Monolignol	Repressor	(Tamagnone et al., 1998)
Switchgrass (<i>Panicum virgatum</i>)	PvMYB4a	R2R3	4	Monolignol	Repressor	(Shen et al., 2012)
<i>Arabidopsis thaliana</i>	AtMYB123/TT2	R2R3	5	Proanthocyanidin	Activator	(Nesi et al., 2001)
Barley (<i>Hordeum vulgare</i>)	HvMYB10	R2R3	5	Proanthocyanidin	Activator	(Himi et al., 2012)
Grapevine (<i>Vitis spp.</i>)	VvMYBPA1	R2R3	5	Proanthocyanidin	Activator	(Bogs et al., 2007)
Grapevine (<i>Vitis spp.</i>)	VvMYBPA2	R2R3	5	Proanthocyanidin	Activator	(Terrier et al., 2009)
Kiwifruit (<i>Actinidia spp.</i>)	DkMYB2	R2R3	5	Proanthocyanidin	Activator	(Akagi et al., 2010)
Kiwifruit (<i>Actinidia spp.</i>)	DkMYB4	R2R3	5	Proanthocyanidin	Activator	(Akagi et al., 2009)
Lotus (<i>Lotus japonicus</i>)	LjTT2a	R2R3	5	Proanthocyanidin	Activator	(Yoshida et al., 2008)
Lotus (<i>Lotus japonicus</i>)	LjTT2b	R2R3	5	Proanthocyanidin	Activator	(Yoshida et al., 2008)
Lotus (<i>Lotus japonicus</i>)	LjTT2c	R2R3	5	Proanthocyanidin	Activator	(Yoshida et al., 2008)
Maize (<i>Zea Mays</i>)	C1	R2R3	5	Anthocyanin	Activator	(Cone et al., 1994; Paz-Ares et al., 1987)
Maize (<i>Zea Mays</i>)	PL	R2R3	5	Anthocyanin	Activator	(Cone et al., 1993)
Maize (<i>Zea Mays</i>)	PL-BLOTCHED1 (PL-BH)	R2R3	5	Anthocyanin	Activator	(Cocciolone & Cone, 1993)
Nectarine (<i>Prunus persica</i>)	PpMYBPA1	R2R3	5	Proanthocyanidin	Activator	(Ravaglia et al., 2013)
Orchid (<i>Oncidium spp.</i>)	OgMYB1	R2R3	5	Anthocyanin	Activator	(Chiou & Yeh, 2008)
Poplar (<i>Populus spp.</i>)	PtMYB134	R2R3	5	Proanthocyanidin	Activator	(Mellway et al., 2009)
Strawberry (<i>Fragaria ananassa</i>)	FaMYB9	R2R3	5	Proanthocyanidin	Activator	(Schaart et al., 2013)
Strawberry (<i>Fragaria ananassa</i>)	FaMYB11	R2R3	5	Proanthocyanidin	Activator	(Schaart et al., 2013)
Apple (<i>Malus domestica</i>)	MdMYB1	R2R3	6	Anthocyanin	Activator	(Tako et al., 2006)
Apple (<i>Malus domestica</i>)	MdMYB10	R2R3	6	Anthocyanin	Activator	(Espley et al., 2007)

Table 2.8, continued

Apple (<i>Malus domestica</i>)	MdMYB110a	R2R3	6	Anthocyanin	Activator	(Chagné et al., 2013)
Apple (<i>Malus domestica</i>)	MdMYBA	R2R3	6	Anthocyanin	Activator	(Ban et al., 2007)
<i>Arabidopsis thaliana</i>	AtMYB75/PAP1	R2R3	6	Phenylpropanoid-derived compound	Activator	(Zuluaga et al., 2008; Zvi et al., 2012)
<i>Arabidopsis thaliana</i>	AtMYB90/PAP2	R2R3	6	Anthocyanin	Activator	(Gonzalez et al., 2008)
<i>Arabidopsis thaliana</i>	AtMYB113	R2R3	6	Anthocyanin	Activator	(Gonzalez et al., 2008)
<i>Arabidopsis thaliana</i>	AtMYB114	R2R3	6	Anthocyanin	Activator	(Gonzalez et al., 2008)
Cabbage (<i>Brassica oleracea</i> var. <i>capitata</i>)	BoMYB2	R2R3	6	Anthocyanin	Activator	(Yuan et al., 2009)
Cabbage (<i>Brassica oleracea</i> var. <i>botrytis</i>)	PURPLE (PR)	R2R3	6	Anthocyanin	Activator	(Chiu et al., 2010)
Chinese bayberry (<i>Myrica rubra</i>)	MrMYB1	R2R3	6	Anthocyanin	Activator	(Huang et al., 2013; Niu et al., 2010)
Gerbera (<i>Gerbera hybrida</i>)	GMYP10	R2R3	6	Anthocyanin	Activator	(Elomaa et al., 2003)
Grapevine (<i>Vitis labruscana</i> .)	VIMYBA1-1	R2R3	6	Anthocyanin	Activator	(Kobayashi et al., 2002)
Grapevine (<i>Vitis labruscana</i>)	VIMYBA1-2	R2R3	6	Anthocyanin	Activator	(Kobayashi et al., 2002)
Grapevine (<i>Vitis labruscana</i>)	VIMYBA1-3	R2R3	6	Anthocyanin	Activator	(Azuma et al., 2008)
Grapevine (<i>Vitis labruscana</i>)	VIMYB2	R2R3	6	Anthocyanin	Activator	(Geekiyana et al., 2002)
Grapevine (<i>Vitis vinifera</i>)	VvMYBA1	R2R3	6	Anthocyanin	Activator	(Kobayashi et al., 2005)
Herba epimedii (<i>Epimedium sagittatum</i>)	EsMYBA1	R2R3	6	Anthocyanin	Activator	(Huang et al., 2013)
Mangoesteen (<i>Garcinia mangostana</i>)	GmMYB10	R2R3	6	Anthocyanin	Activator	(Palapol et al., 2009)
Nectarine (<i>Prunus persica</i>)	PpMYB10	R2R3	6	Anthocyanin	Activator	(Ravaglia et al., 2013)

Table 2.8, continued

Petunia (<i>Petunia hybrida</i>)	AN2	R2R3	6	Anthocyanin	Activator	(Quattrocchio et al., 1999; Quattrocchi, et al., 1998)
Petunia (<i>Petunia hybrida</i>)	DPL	R2R3	6	Anthocyanin	Activator	(Albert et al., 2011)
Petunia (<i>Petunia hybrida</i>)	PHZ	R2R3	6	Anthocyanin	Activator	(Albert et al., 2011)
Snapdragon (<i>Antirrhinum majus</i>)	ROSEA1	R2R3	6	Anthocyanin/phenolic acid	Activator	(Schwinn et al., 2006)
Snapdragon (<i>Antirrhinum majus</i>)	ROSEA2	R2R3	6	Anthocyanin	Activator	(Schwinn et al., 2006)
Snapdragon (<i>Antirrhinum majus</i>)	VENOSA	R2R3	6	Anthocyanin	Activator	(Schwinn et al., 2006)
Strawberry (<i>Fragaria ananassa</i>)	FaMYB10	R2R3	6	Anthocyanin	Activator	(Laitinen et al., 2008)
Sweet pepper (<i>Capsicum annuum</i>)	CaA	R2R3	6	Anthocyanin	Activator	(Borovsky et al., 2004)
Tobacco (<i>Nicotiana tabacum</i>)	NtAN2	R2R3	6	Anthocyanin	Activator	(Pattanaik et al., 2010)
Tomato (<i>Lycopersicon esculentum</i>)	LeANT1	R2R3	6	Anthocyanin	Activator	(Mathews et al., 2003)
<i>Arabidopsis thaliana</i>	AtMYB11	R2R3	7	Flavonol	Activator	(Stracke et al., 2007)
<i>Arabidopsis thaliana</i>	AtMYB12	R2R3	7	Flavonol/chlorogenic acid	Activator	(Mehrtens et al., 2005; Stracke et al., 2007)
<i>Arabidopsis thaliana</i>	AtMYB111	R2R3	7	Flavonol	Activator	(Stracke et al., 2007)
Grapevine (<i>Vitis spp.</i>)	VvMYBF1	R2R3	7	Flavonol	Activator	(Czemmel et al., 2009)
Japanese gentian (<i>Gentiana triflora</i>)	GtMYBP3	R2R3	7	Flavonol	Activator	(Nakatsuka et al., 2012)
Japanese gentian (<i>Gentiana triflora</i>)	GtMYBP4	R2R3	7	Flavonol	Activator	(Nakatsuka et al., 2012)
Maize (<i>Zea Mays</i>)	P1	R2R3	7	Phobaphene/Flavonol	Activator	(Grotewold et al., 1994)
<i>Arabidopsis thaliana</i>	AtMYB46	R2R3	13	Monolignol	Activator	(McCarthy et al., 2009)
<i>Arabidopsis thaliana</i>	AtMYB61	R2R3	13	Monolignol	Activator	(Newman et al., 2004)

Table 2.8, continued

<i>Arabidopsis thaliana</i>	AtMYB83	R2R3	13	Monolignol	Activator	(McCarthy et al., 2009)
Poplar (<i>Populus spp.</i>)	PtoMYB216	R2R3	13	Monolignol	Activator	(Tian et al., 2013)
Soybean (<i>Glycine max</i>)	GmMYB12B2	R2R3	20	Isoflavonoid	Activator	(Li et al., 2013)
Apple (<i>Malus domestica</i>)	MdMYB6	R2R3	22	Anthocyanin	Repressor	(Gao et al., 2011)
Grapevine (<i>Vitis spp.</i>)	VvMYB5a	R2R3	27	Flavonoid	Activator	(Deluc et al., 2006)
Grapevine (<i>Vitis spp.</i>)	VvMYB5b	R2R3	27	Proanthocyanidin and anthocyanin	Activator	(Deluc et al., 2008)
<i>Epimedium sagittatum</i>	EsMYB9	R2R3	27	Anthocyanin	Activator	(Huang et al., 2013)
<i>Arabidopsis thaliana</i>	AtCPC	R3		Anthocyanin	Repressor	(Zhu et al., 2009)
<i>Arabidopsis thaliana</i>	AtMYBL2	R3		Flavonoid	Repressor	(Dubos et al., 2008; Matsui et al., 2008)
Petunia (<i>Petunia hybrida</i>)	PhMYB27	R2R3		Anthocyanin	Repressor	(Albert et al., 2014; Albert et al., 2011)
Petunia (<i>Petunia hybrida</i>)	PhMYBx	R3		Anthocyanin	Repressor	(Albert et al., 2014; Albert et al., 2011)

2.9 Transcriptome and RNA-Seq

Transcriptome is defined as the complete set of transcripts in a cell, and their quantity, for a specific developmental stage or physiological condition (Wang et al., 2009). Whole transcripts are divided to two groups; coding and non-coding ribonucleic acid (RNA). Coding RNA consists of messenger RNA (mRNA) which carries genetic information that directs the synthesis of proteins; while non-coding RNA consists of transfer RNA (tRNA) and ribosomal RNA (rRNA) both of which are involved in protein synthesis. Therefore, RNA acts as a bridge connecting genotype and phenotype. According to Wang et al. (2009), understanding transcriptome is essential for interpreting the functional elements of the genome; revealing the molecular constituents of cells and tissue, and also understanding plant developmental and phatogenesis (Wang et al., 2009).

There are various technologies developed to deduce and quantify the transcriptome including microarray, serial analysis of gene expression (SAGE) and massively parallel signature sequencing (MPSS). Microarray is based on hybridization approach in which cDNA is hybridized to an array of complementary oligonucleotide probes corresponding to gene of interest and the abundance of a particular mRNA species is estimated from its hybridization intensity to the relevant probe. Whereas, both SAGE and MPSS are based on a sequencing approach that involves sequencing of short cDNA fragments followed by counting the number of times a particular fragment has been observed.

Recently, RNA-Seq has been developed as a new approach that offers several advantages over existing technologies. It provides high single-base resolution for annotation and quantitative digital gene expression levels at much lower cost compared to large scale Sanger sequencing. Additionally, as no prior knowledge of gene sequence required; this method is attractive to non-model organisms that do not have the genomic sequences yet. Consequently, RNA-Seq can detect and accurately quantitate a wide range

of expressed genes either at low or very high levels. Furthermore, RNA-Seq requires less RNA sample because no laborious cloning steps are needed.

RNA-Seq uses next-generation sequencing (NGS) technologies for sequencing the whole transcriptome. In general, total mRNA or fragmented mRNA is converted to cDNA followed by a fragmentation step to generate a library of cDNA fragments. Next, adaptors are ligated at both ends of fragmented cDNAs. Subsequently, each molecule is sequenced using next-generation sequencer technology to obtain short sequences from one end (single-end sequencing) or both ends (paired-end sequencing). Paired-end sequencing yields much more information than single-end sequencing. The resulting sequencing reads are either mapped to reference genome or assembled *de novo* to produce a genome scale transcription map that consists of both the transcriptional structure and/or expression level for each gene (Wang et al., 2009).

2.10 Public protein databases

Annotation of the assembled *de novo* transcripts generated from RNA-Seq is the first important step for further understanding the experimental subjects. There are several public protein databases that are normally used for transcript annotation such as the National Center for Biotechnology Information (NCBI) reference sequence database, Swiss-Prot database, Clusters of Orthologous Groups (COGs) database, Gene Ontology (GO) database and Kyoto Encyclopedia of Genes and Genomes (KEGG) database. Different database has different style of protein annotation and classification.

RefSeq or NCBI's reference sequence database is a curated non-redundant collection of sequences (<http://www.ncbi.nlm.nih.gov/RefSeq/>) that compared of complete or incomplete genomes, transcripts and proteins that are represented in living organisms

including prokaryotes, eukaryotes and viruses (Pruitt et al., 2006). RefSeq database allows the user to retrieve the background information of recorded nucleotide and protein sequences such as annotation, names, origin, coding regions and conserved domains (Pruitt et al., 2006).

Swiss-Prot database is a protein sequence and knowledge database (<http://www.expasy.org/sprot/> and <http://www.ebi.ac.uk/swissprot/>) that provide a high level of annotation, minimal level of redundancy and high integration with other databases (Bairoch & Apweiler, 2000). Compared to NCBI's RefSeq, the protein annotation in Swiss-Prot database is more detailed. Swiss-Prot consists of two main components of data: the core data and the annotation. The core data consists of the amino acid sequence, the protein name, taxonomic data and citation information. Some of the entries have further information on specific protein. Swiss-Prot display detailed annotation such as protein function, enzyme-specific information, biologically relevant domains and sites, post-translational modification, molecular weight, subcellular localization, splice isoform and polymorphism (Boeckmann et al., 2003).

COGs protein database is designed to classify proteins from completely sequenced genomes based on orthology concept (Tatusov et al., 2000). As orthologous proteins have the same domain architecture and the same function, the function of unknown proteins can be predicted based on the the relationship between other orthologous proteins that has been experimentally characterized using COGNITOR program (Tatusov et al., 2000). Subsequently, the COGs classify proteins by functional categories such as translation, replication, transcription and signal transduction.

GO database provide ontologies, a set of controlled, structured vocabularies to describe attributes of gene products in three non-overlapping domains of molecular biology which include Molecular Function (MF), Biological Process (BP) and Cellular

Component (CC). Based on Gene Ontology Consortium (2004), Molecular Function describes activities, while Biological Process describes biological goals accomplished by one or more ordered assemblies of molecular functions. Cellular Component describes locations, at the molecular levels of subcellular structures and macromolecular complexes.

KEGG database provide understanding of biological functions from a genomic perspective by linking genomic information with the network of interacting molecules (Kanehisa & Goto, 2000). KEGG database consist of three databases: GENES, LIGAND and PATHWAY which represent genomic, chemical and pathway network information, respectively (Kanehisa & Goto, 2000).

2.11 Applications of transcriptome profiling

Recently, the RNA-Seq technology has become the main approach for high-throughput gene discovery and gene expression analysis in model and non-model organisms. The technologies give advantages to plant genomics in sequencing difficult plant species especially the non-model plant that mostly contain large and complex genome, high levels of ploidy and large proportions of repeats. RNA-Seq provides a better alternative for rapid and efficient access to genetic information compared to whole genome sequencing. Consequently, the plant genomic data has grown rapidly over the past few years. Among NGS technologies, the Illumina/Solexa platform has been widely used as it offers fast, inexpensive, accurate, efficient and reliable tool for transcriptome characterization and gene discovery in many plants. Knowledge on, gene regulation and complex metabolic networks in plants have also improved through RNA-Seq analysis.

RNA-Seq has been used to discover the important genes for economic traits improvement and nutrient accumulation. For instance, fruit color which is mainly attributed to anthocyanin, is one of the most important economic trait. Therefore, in order to understand the molecular mechanism of anthocyanin biosynthesis and its regulation, the anthocyanin pathway in Korean black raspberry (*Rubus coreanus* Miquel) and also red and yellow sweet cherry (*Prunus avium* L.) fruits were elucidated through transcriptome analysis (Hyun et al., 2014; Wei et al., 2015). Through transcriptome profiling of tomato (*Solanum lycopersicum*) fruit development, and transcription factors associated with ascorbic acid, carotenoid and flavonoid biosynthesis that determine nutrient levels in tomato were elucidated (Ye et al., 2015).

RNA-Seq has also been applied for plant species that are difficult to grow *in vitro*. It was reported that transcriptome profiling at different plant embryo developmental stages such as embryogenic callus, yellow embryogenic callus and global embryo was done in order to understand the molecular mechanisms regulating early somatic embryogenesis in endangered medicinal *Eleutherococcus senticosus* Maxim (Tao et al., 2016). Moreover, besides providing molecular basis on fiber and secondary metabolite biosynthesis in Sodom apple (*Calotropis gigantea*), it was also reported that molecular markers to facilitate breeding and genetic improvement of varieties can also be developed using RNA-Seq approach (Muriira et al., 2015).

Furthermore, RNA-Seq helps to provide information to improve plant tolerance towards diseases or harsh environment by allowing discovery of genes that are involved in biotic and abiotic stress. For instance, the identification of important defense-related genes was achieved by *de novo* characterization of the banana root transcriptome and gene expression analysis under *Fusarium oxysporum* f. sp. *Cubense* tropical race 4 infection (Wang et al., 2012). Genes involved in systemic symptom development caused

by *Cucumber mosaic virus* in infected *Nicotiana tabacum* were also determined (Lu et al., 2012). In addition to biotic stresses, RNA-Seq combined with digital gene expression (DGE) analysis allows plant scientists to identify genes that are involved in abiotic stresses in plants such as drought, cold and high salinity. For instance, RNA-Seq aids in identification of dehydration-related genes in eucalyptus, chrysanthemum and cotton (Bowman et al., 2013; Villar et al., 2011; Xu et al., 2013).

Among transcriptome profile applications, identifying the genes related to secondary metabolite biosynthetic pathways especially in medicinal plants have been among the most popular objectives. Most biosynthetic pathways directing secondary metabolites production still remains unclear due to limited knowledge in the plant database. Besides, some bioactive compounds are uniquely present in specific medicinal plants and are derived from unestablished biosynthetic pathways. Therefore RNA-Seq has been applied in understanding plant biosynthetic pathways at the genomic level. This approach is essential in order to elucidate genes that are involved directly or indirectly in secondary metabolite pathways in medicinal plants by analyzing gene regulations of transcriptome profiles.

Due to the lack of information in the plant genome database, *de novo* transcriptome sequencing of medicinal plants can be considered as an alternative for plant database enrichment. Recently, there has been dramatic growth on transcriptome profiling studies of medicinal plants to elucidate secondary metabolite biosynthetic pathway using RNA-Seq approach which include reports for *Gingko biloba* L., *Gentiana macrophylla*, *Nervilia fordii*, *Chlorophytum borivilianum*, *Polygonum minus*, *Hypericum perforatum*, *Lophophora williamsii*, safflower, *Polygala tenuifolia*, *Uncaria rhynchophylla*, *Sophora flavescens*, *Rhodiola algida* and *Gentiana straminea* (Guo et al., 2014; Han et al., 2015; Han et al., 2015; He et al., 2012; Hua et al., 2014; Huang et al., 2015; Ibarra-Laclette et

al., 2015; Kumar et al., 2016; Li et al., 2012; Loke et al., 2016; Tian et al., 2015; Zhang et al., 2014; Zhou et al., 2016).

Elucidation of genes involved in the downstream terpenoid biosynthetic pathway are commonly studied. Although the main terpenoid biosynthetic pathway is established, further downstream of the biosynthetic pathway of complex terpenoid derivatives remains unclear. Medicinal bioactive compounds that are classified under terpenoids are mostly complex terpenoids and some are species-specific compounds. For instance, *de novo* transcriptome sequencing was performed to resolve triterpenoid biosynthetic pathway such as mogrosides biosynthesis in *Siraitia grosvenorii* (Tang et al., 2011), ginsenosides biosynthesis in *Panax ginseng*, *Panax quiquefolius* and *Panax notoginseng* (Cao et al., 2015; Luo et al., 2011; Sun et al., 2010) and gypenosides biosynthesis in *Gynostemma pentophyllum* (Subramaniyam et al., 2011). In addition, tanshinone derived from diterpene together with rosminic acid biosynthetic pathway were also enriched in *Salvia miltiorrhiza* through *de novo* transcriptome approach (Wenping et al., 2011). It was reported that RNA-Seq of *Curcuma longa* L. rhizome revealed novel transcripts related to anticancer and antimalarial terpenoids (Annadurai et al., 2013). RNA-Seq also has been applied to enrich flavonoid biosynthetic pathway. For instance, in *Pueraria lobata* putative genes involved in isoflavones biosynthesis was identified (Wang et al., 2015).

Digital gene expression (DGE) analysis for elucidating differentially expressed genes (DEGs) is an approach that can be used to further understand the nature of a plant's response towards various stimuli or stresses (Molina et al., 2011). RNA-Seq and DEGs analysis has been used to identify genes that are directly or indirectly involved in the biosynthetic pathways of target bioactive compounds. For instance, seven cytochrome P450 (CYP450) and five UDP-glucosyltransferase (UDPG) were identified as candidate genes that direct mogrosides, one of complex terpenoid derivative biosynthesis in *Siraitia*

grosvenorii. This was carried out by comparing its transcriptome profile in different fruiting stages (Tang et al., 2011).

Another approach, through the combination of RNA-Seq and methyl jasmonate induction experiments, genes encoding one CYP450 and four UDP-glycosyltransferases were proposed to be key enzymes in the ginsenoside biosynthesis in *Panax quinquefolius* (Sun et al., 2010). In addition, by combining RNA-Seq and phylogenetic tree analysis based on the previously identified CYP450 and UDP-glycosyltransferase in *Panax quinquefolius*, two CYP450 and one UDP-glycosyltransferase were also elucidated as candidates for ginsenoside biosynthesis in *Panax notoginseng* (Luo et al., 2011; Sun et al., 2010).

University of Malaya

CHAPTER 3: METHODOLOGY

3.1 Plant materials

3.1.1 *B. rotunda* plant

The yellow rhizome variety of *B. rotunda* was purchased from KIZA herb field, Kuala Krau, Pahang. Some of the rhizomes were planted and grown under natural field conditions and maintained under the Plant Biotechnology Research Laboratory, University of Malaya.

3.1.2 Establishment of *B. rotunda* cell suspension cultures

Callus induction was initiated with surface sterilization of shoot buds of *B. rotunda* (4 - 5 cm). The shoot buds were rinsed thoroughly under running tap water. Then, eight shoot buds per treatment were soaked in 20 % (v/v) commercial bleach, chlorox (containing 5 – 10 % (v/v) of sodium hypochloride) and a few drops of Tween 20 for 10 to 15 minutes. The container containing shoot buds were shaken occasionally and the shoot buds were then rinsed in sterile distilled water 3 times. Subsequently, shoot buds were rinsed in 70 % (v/v) ethanol for few seconds and finally with sterile distilled water. The shoot buds were blotted dry and cut horizontally into discs of 1 mm thick. The discs were then transferred into callus induction medium (CIM), which contained Murashige and Skoog (MS) basal salt (Murashige & Skoog, 1962) supplemented with 1 mg/ml indole-3-acetic acid (IAA), 1 mg/ml naphthalene acetic acid (NAA) and 2 mg/ml 2,4-dichlorophenoxyacetic acid (2,4-D). The explants were placed in the dark under controlled temperature (25°C).

Callus initiated from CIM was identified and transferred into propagation media containing MS media supplemented with 3 mg/ml 2,4-D. Mixed callus was used to establish cell suspension cultures propagated in MS liquid media (LM) (Murashige &

Skoog, 1962), supplemented with 1 mg/l benzylaminopurine (BAP), 1 mg/l NAA, 1 mg/l biotin, 2 mg/l 2,4-D and 99.42 mg/l L-glutamine. The pH medium was adjusted to pH 5.7. The cell suspension cultures were placed on orbital shaker (70 - 80 rotations per minute, rpm) under a 16 h photoperiod with a light intensity of 31.4 $\mu\text{mol}/\text{m}^2/\text{s}$ provided by cool fluorescent lamp in the growth room. The cell suspension culture was subcultured every 10 – 12 days by decanting off 40 ml of old medium and replaced with fresh medium in 250 ml conical flask.

3.1.3 Phenylalanine treatment

Equal amounts of the cell suspension (5 ml of settled cell volume) were used in all experiments. Settled cell volume is defined as sedimentation of cell suspensions in graduated tubes (50 ml falcon conical centrifuge tubes). For the control, no phenylalanine was used, while for phenylalanine-treated samples, 40 mg/l of phenylalanine was added into LM immediately after subculture.

3.2 General molecular technique

3.2.1 DNA extraction using CTAB method

Genomic DNA from *B. rotunda* was isolated using modified cetylmethylammonium bromide (CTAB) method (Doyle, 1991). Initially, 2 gram of *B. rotunda* leaf was ground to powder in liquid nitrogen. Subsequently the ground sample was added into 15 ml falcon tube containing 10 ml of pre-heated CTAB extraction buffer with 60 μl β -mercaptoethanol. CTAB extraction buffer consist of 2 % (w/v) CTAB, 2 % (w/v) of polyvinylpyrrolidone (PVP), 1.4 M of sodium chloride (NaCl), 0.5 M of ethylenediaminetetraacetic acid (EDTA) at pH 8.0 and 1 M of Tris-HCL at pH 8.0. Then,

the tube was heated at 65°C for 10 minutes. Next, an equal volume of chloroform:isoamyl alcohol (24:1; v/v) was added to the mixture and the mixture was mixed for a few seconds. The sample was centrifuged in an Eppendorf 5417R centrifuge (Eppendorf, Hamburg, Germany) at 10,621 X g (10,000 rpm) for 15 minutes to remove protein impurities. The supernatant was recovered and transferred into a new tube and the steps repeated once. The repeated steps involved the addition of equal volume of chloroform:isoamyl alcohol (24:1; v/v), followed by centrifugation and supernatant recovery. Next, 2/3 volumes of isopropanol was added to the supernatant. The mixture was kept at -80°C for overnight to precipitate nucleic acid and then centrifuged at 10,621 X g (10,000 rpm) for 15 minutes. The supernatant was discarded and the remaining pellet was washed with 1 ml of cold 70% (v/v) ethanol and the mixture was transferred into a new microcentrifuge tube. The sample was again centrifuged at 10,621 X g (10,000 rpm) for 10 minutes. The pellet was air-dried and dissolved in 500 µl TE buffer and 2 µl of 20 mg/ml RNase A. The mixture was incubated for 30 minutes at 37°C to remove the RNA. Finally, the extracted DNA was stored at -20°C freezer.

The concentration and purity of total DNA samples were determined using an Eppendorf Biophotometer. In general, the acceptable 260/280 ratio of pure DNA is between the range of 1.8 – 2.0. In addition, the acceptable range of 260/230 ratio is 2.0 – 2.2. For measuring the DNA concentration, 2 µl of samples were mixed with 498 µl sterile distilled water. The diluted samples were measured at wavelength of 260 nm and 280 nm. The readings were taken against a blank of 500 µl sterile distilled water.

Agarose gel electrophoresis was carried out to verify the success of DNA extraction using CTAB method. About 2 µl (1 µg) of Perfect 1 kb DNA ladder (EURx, Poland) and 2 µl (~ 1.4 µg) of DNA were loaded into 1% (w/v) agarose gel containing ethidium bromide (see Appendix A.11 for agarose gel preparation). Prior to loading the sample

into the agarose gel, the DNA sample was pre-mixed with 10 μ l of 6X DNA loading dye (Fermentas, USA). The 1% (w/v) agarose gel was electrophoresed at 120 volts for 25 min. The migrated bands were visualized under UV light using AlphaImagerTM 2200 (Alpha Innotech, USA).

3.2.2 RNA extraction from cell suspension cultures using CTAB method

A modified CTAB method was employed to extract total RNA from both control and phenylalanine-treated cell suspension cultures (Al-Obaidi et al., 2010). Initially, 300 to 500 milligram of cell suspension culture was ground in liquid nitrogen. Subsequently the ground sample was added into a 2 ml microcentrifuge tube containing 1 ml of pre-heated CTAB extraction buffer with 20 μ l β -mercaptoethanol. Then, the tube was heated at 65°C for 10 minutes. The mixture was mixed for few seconds. An equal volume of chloroform:isoamyl alcohol (24:1; v/v) was then added to the mixture and the mixture mixed for few seconds.

The sample was centrifuged in an Eppendorf 5417R centrifuge (Eppendorf, Hamburg, Germany) at 10,621 X g (10,000 rpm) for 15 minutes to remove protein impurities. The supernatant was recovered and transferred into a new microcentrifuge tube and the steps repeated 2-3 times. The repeated steps involved the addition of equal volume of chloroform:isoamyl alcohol (24:1; v/v), followed by centrifugation and supernatant recovery. Next, 0.1 volume of 3 M sodium acetate together with 3 volumes of pre-cooled absolute ethanol were added to the supernatant. The mixture was kept at -80°C for 2–3 days to precipitate RNA and then centrifuged at 10,621 X g (10,000 rpm) for 30 minutes at 4°C. The supernatant was discarded and the remaining pellet was washed with 1 ml of cold 70% (v/v) ethanol. The sample was again centrifuged at 10,621 X g (10,000 rpm)

for 5 minutes at 4°C, removed and pellet was air-dried and dissolved in 20 µl diethyl pyrocarbonate (DEPC)-treated water.

The concentration and purity of total RNA samples were determined using an Eppendorf Biophotometer. The reading at 260 nm indicates the concentration of nucleic acid; while reading at 280 nm indicates the amount of protein in the sample. The reading at 230 nm indicates the presence of organic compounds in the sample. The ratio of absorbance at 260 nm and 280 nm was used to assess the purity of DNA and RNA. In general, the acceptable 260/280 ratio of pure RNA is between the range of 1.8 – 2.0. Whereas, the acceptable range of 260/230 ratio is 2.0 – 2.2. For measuring the RNA concentration, 2 µl of samples were mixed with 498 µl sterile distilled water. The diluted samples were measured at wavelengths of 260 nm and 280 nm. The readings were taken against a blank of 500 µl sterile distilled water.

3.2.3 RNA extraction from cell suspension cultures using Easy Spin RNA extraction kit method

RNA was extracted using Easy Spin RNA extraction kit (iNtRON Biotechnology, Kyungki-Do, Korea) according to the manufacturer's instructions. About 100 mg of cell suspension culture was ground in liquid nitrogen and transferred into 1.5 ml microcentrifuge tube. About 1 ml of Lysis buffer was added into the tube and the mixture was homogenized by vortex at room temperature for 10 seconds. Then, 200 µl of chloroform was added and the mixture was mixed. After that, the mixture was centrifuged at 17,949.49 X g (13,000 rpm) using Eppendorf 5417R centrifuge (Eppendorf, Hamburg, Germany) at 4°C for 10 minutes. Then 400 µl of the upper layer was transferred to a new microcentrifuge tube. About 400 µl of Binding buffer was added and mixed by pipetting or gently inverted for 2 -3 times. The mixture was incubated at room temperature for 1

minute. The upper layer containing the Binding buffer was then loaded to the column and centrifuged at 17,949.49 X g (13,000 rpm) for 30 seconds. Then, 700 µl of Washing buffer A was added to the column and centrifuged at 17,949.49 X g (13,000 rpm) for 30 seconds. Next, 700 µl of Washing buffer B was added to the column and centrifuged at 17,949.49 X g (13,000 rpm) for 30 seconds. The tube was centrifuged for another 1 – 2 minutes to dry the column membrane. In order to elute the RNA from the column membrane, the column was placed in a clean 1.5 ml tube and 50 µl of Elution buffer was added directly on the membrane. The column was incubated at room temperature for 1 minute and centrifuged at 17,949.49 X g (13,000 rpm) for 1 minute.

The concentration and purity of the extracted RNA was determined using NanoDrop® 2000 UV-Vis Spectrophotometer (Thermo Scientific, USA) and agarose gel electrophoresis. The ratio of absorbance at 260 nm and 280 nm was used to assess the purity of RNA. In general, the acceptable 260/280 ratio of pure RNA is between the range of 1.8 – 2.0 (Sambrook et al., 1989).

3.2.4 Gel electrophoresis

In order to determine the RNA integrity, the extracted RNA was electrophoresed on 1% (w/v) agarose gel containing ethidium bromide at 120 volts for 25 minutes. 1 X Tris-Borate-EDTA (TBE) buffer was used as running buffer as well as gel preparation buffer. About 2 µl (~ 1 - 8 µg) of extracted total RNA samples were mixed with 2 µl 2 X RNA loading dye and loaded into 1% (w/v) agarose gel. About 2 µl of RiboRuler High Range RNA ladder (Thermo Scientific, USA) was used as the molecular weight marker. The migrated bands were visualized and photographed under UV light using AlphaImager 2200 (Alpha Innotech, Miami, USA). The presence of two ribosomal RNA bands (28S and 18S) indicates a sample with good RNA integrity.

3.2.5 Preparation of cDNA

Reverse transcription reactions were performed using *TransScript*®II Reverse Transcriptase (TransgenBiotech, Beijing, China) with approximately 2 µg total RNA (unless stated) following the manufacturer's instructions. RNA template for cDNA synthesis was RNA samples from control and phenylalanine treated cell suspension culture. Initially, DNase treated RNA together with 1 µl random primer was incubated at 65°C for 5 minutes to remove any RNA secondary structure. Then, the mixture was chilled on ice for 2 minutes. Next, reverse transcription mixture was added into RNA template and primer. A total of 20 µl reverse transcription reaction mixture consist of 10 µl of TS reaction mixture, 1 µl of transcript RT enzyme and 1 µl of gDNA remover. For reverse transcription, the reaction mixture was incubated at 25°C for 10 minutes followed by 42°C for 30 minutes. The mixture was heated at 85°C for 5 minutes to inactivate the reverse transcription reaction. Finally, the cDNA template can be used immediately for amplification or stored at -20°C.

3.2.6 Polymerase chain reaction (PCR)

3.2.6.1 PCR reaction mixture

PCR amplifications were performed using HotStarTaq DNA polymerase (Qiagen). A total PCR reaction mixture of 20 µl consist of 1 X CoralLoad PCR buffer, 0.2 mM dNTP mixture, 0.5 µM forward primer, 0.5 µM reverse primer, 0.025 U/ µl HotStart Taq DNA polymerase and sample. For positive control, DNA or RNA was used as template while for negative control, sterile distilled water was used.

3.2.6.2 PCR parameters

The amplification was performed on T100 Thermal Cycler (Bio-Rad, California, USA). PCR was conducted to ensure the cDNA prepared was free from genomic DNA contaminations. The PCR cycling conditions consisted of initial denaturation at 95°C for 5 minutes, denaturation at 95°C for 1 minute, annealing at 56°C for 1 minute, extension at 72°C for 1 minute and final extension at 72°C for 10 minutes. The denaturation, annealing and extension step was repeated for 35 times.

However, for amplification of target genes (putative *BrMYB1*, *BrMYB2* and *BrMYB3*), the PCR cycling conditions consisted of an initial denaturation at 95°C for 5 minutes, denaturation at 95°C for 10 seconds, annealing for 10 seconds, extension at 72°C for 45 seconds and final extension at 72°C for 10 minutes. The denaturation, annealing and extension steps were repeated for 35 times. The annealing temperature between 45 – 55°C was tested. The optimal annealing temperature for putative *BrMYB1*, *BrMYB2* and *BrMYB3* genes were 52.9°C, 51.2°C and 52.9°C, respectively.

3.2.6.3 Gel electrophoresis

About 2 µl of 1 kb Plus DNA ladder (Thermoscientific, USA) and 10 µl of PCR product were loaded into 1% (w/v) agarose gel. The agarose gel was electrophoresed at 120 volts for 25 min. The migrated bands were visualized using ChemiDoc MP System (Bio-Rad, California, USA).

3.2.7 Real-time or quantitative polymerase chain reaction (qPCR)

qPCR was performed on QuantStudio 12 K Flex realtime PCR platform (Applied Biosystem, Carlsbad, CA, USA) using Power SYBR® Green Master Mix (Applied

Biosystem, Carlsbad, CA, USA) to detect transcript abundance. The 20 μl of SYBR Green reaction consisted of 1 X SYBR Green master mix, 0.5 μM forward primer, 0.5 μM reverse primer, 40 ng cDNA template and sterile distilled water. The amplification was achieved by the following PCR protocol: first denaturation at 95°C for 10 minutes, then 40 cycles of denaturation at 95°C for 15 s, annealing and extension at 60°C for 1 minute. The dissociation curve was established at the end of PCR cycle at 95°C for 15 s, 60°C for 1 minute followed by 95°C for 15 s.

3.3 Transcriptome profiling sequencing

3.3.1 Library preparation and sequencing

Control and phenylalanine-treated cell cultures were harvested after 14 days of propagation. The RNA extraction and gel electrophoresis was done as described in subsections 3.2.2 and 3.2.4, respectively. The quality and quantity of RNA samples were analysed using an Agilent 2100 Bioanalyzer (Agilent, Waldbronn, Germany) to ensure RNA concentrations of more than 400 ng/ μl and to obtain RNA quality with an OD 260/280 of between 1.8 – 2.2, 28S/18S > 1.8 and an RNA integrity number (RIN) = 8. Whole transcriptome sequencing was performed by Beijing Genome Institute (BGI), Shenzhen, China using HiSeq 2000 Sequencing System (Illumina, Inc., USA).

3.3.2 Transcript assembly and annotation

Paired-end reads sequence data with a length of 75 bp that generated from the HiSeq 2000 Sequencing System (Illumina, Inc., USA) was transformed by base calling into sequence data, called raw data or raw reads. The raw reads that only have adaptor fragments were filtered out by the Illumina pipeline, leaving the remaining reads which

later denoted as clean reads or clean data. The RNA-seq data were deposited at the NCBI under BioProject with accession number PRJNA256116 with SRA Study accession number SRR1524841 for control (untreated) and SRR1524842 for phenylalanine treated samples. To ensure that clean reads of RNA-Seq data are of high quality and suitable for subsequent analyses, the quality control of RNA-Seq clean data were determined by using FastQC software (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>).

Next, the clean data were assembled into transcript contigs by short reads assembling program SOAPdenovo (Short Oligonucleotide Analysis Package) software (Li et al., 2010). This software adopts the de Bruijn graph data sequence data structure to construct contigs. The reads were mapped to the contigs using the paired-end relationship between reads, contigs from the same transcript can be detected. Next, scaffolds were made by connecting the contigs using SOAPdenovo software, in which N represents the unknown sequence between each two contigs. Paired-end reads were used again to fill the intra-scaffold gaps to form unigenes. As two samples, which were treated and control from the sample species were sequenced, unigenes from each assembly was further assembled to acquire longer non-redundant unigenes using TGI clustering tools (Pertea et al., 2003).

Longer unigenes that were generated by combining both transcripts from control and treated samples were annotated against protein databases such as NR, Swiss-Prot, KEGG and COG by Blastx (e-value cutoff of < 0.00001) alignment. The best aligned results were used to determine the sequence direction of the unigenes. Next, for other unaligned unigenes, sequence orientation as well as its coding regions was predicted by using ESTscan software (Iseli et al., 1999).

3.4 Transcriptome profiling analysis

3.4.1 Functional annotation

Functional annotation was done to reveal the protein functional information of unigenes, which include protein orthologous groups and pathway annotation. All unigenes with sequence orientation were subjected to functional annotation. Homology search was done by Blastx alignment of unigenes against public protein databases such as non-redundant (NR), Swiss-Prot, KEGG and COG with e -value < 0.00001 . Next, in order to classify unigenes in GO functional annotation, unigenes with nr annotation information was mapped to their respective ontologies using Blast2GO program (Conesa et al., 2005) and further gene classification was done using WEGO software (Ye et al., 2006). Unigenes were classified under three GO-terms namely molecular function, cellular component and biological process.

3.4.2 CDS prediction

For codon sequence (CDS) prediction, all unigenes were firstly aligned by Blastx (e -value cut off of < 0.0001) to protein databases in the priority order of nr, Swiss-Prof, KEGG and COG. Codon region sequences of unigenes with the highest Blastx ranks was selected and translated into amino acid sequences using standard codon table. Unaligned unigenes were scanned using ESTscan (Iseli et al., 1999) for codon sequence and amino acid sequence of the coding region.

3.4.3 Digital Gene Expression analysis for elucidating differentially expressed genes (DEGs)

DEGs analysis was carried out based on BGI's analysis pipeline. A rigorous algorithm that developed based on method described by Audic & Claveris (1997) was employed to identify DEGs from two samples of RNA-Seq data. Because the expression of each gene occupies only a small part of the library, the number of unambiguous clean tag from gene A was denoted as x and the $p(x)$ is in the Poisson distribution (Figure 3.1). The total clean tag number of sample 1 is N_1 , and the total clean tag number of sample 2 is N_2 ; gene A holds x tags in sample 1 and y tags in sample 2. Hence, the probability of gene A expressed equally between control and phenylalanine treated samples can be calculated as depicted in the formula in Figure 3.2. P value corresponds to differential gene expression test.

$$p(x) = \frac{e^{-\lambda} \lambda^x}{x!}$$

Figure 3.1: Poisson distribution calculation to identify differentially expressed genes between control and phenylalanine treated samples.
 λ is the real transcripts of the gene;
 x is tags from gene A in sample 1

$$2 \sum_{i=0}^{i=y} p(i|x)$$

$$\text{or } 2 \times \left(1 - \sum_{i=0}^{i=y} p(i|x) \right) \quad \left(\text{if } \sum_{i=0}^{i=y} p(i|x) > 0.5 \right)$$

$$p(y|x) = \left(\frac{N_2}{N_1} \right)^y \frac{(x+y)!}{x! y! \left(1 + \frac{N_2}{N_1} \right)^{(x+y+1)}}$$

Figure 3.2: The probability formula to calculate the expressed gene equally between two samples. x is tags from gene A in sample 1; y is tags from gene A in sample 2; N_1 is total clean tag number of sample 1; N_2 is total clean tag number of sample 2.

False Discovery Rate (FDR) is a method to determine the threshold of P value in multiple test and analysis through manipulating the FDR value. In this study, the level of transcripts or unigenes was determined using Reads per kb per Million reads (RPKM) method (Mortazavi et al., 2008). Figure 3.3 show the RPKM equation to quantify the transcripts level.

$$RPKM = \frac{10^6 C}{NL/10^3}$$

Figure 3.3: Reads per kb per Million reads (RPKM) equation to calculate the transcripts level. C is the number of reads that uniquely mapped to unigene A; N is the number of reads that uniquely mapped to all unigenes in the sample experiment; L is the length of unigene A in basepair.

Subsequently, $FDR \leq 0.001$ and the absolute value of $\log_2 \text{Ratio} \geq 1$ as the thresholds were used to judge the significance of gene expression difference. Ratio was defined as the ratio of RPKM of phenylalanine treated over RPKM of control. The up- and down-regulation of transcripts were indicated by the positive and negative value of $\log_2 \text{Ratio}$,

respectively. Besides, the unigene expression level fold changes in phenylalanine treated sample compared to control sample was determined by the calculation of $\log_2\text{Ratio}$.

3.4.4 KEGG pathway enrichment analysis

Pathway-based analysis helps to further understand biological functions. KEGG is the major public pathway-related database. Some unigenes were annotated in more than one pathways and similarly, one enzyme in a specific pathway was annotated by multiple unigenes. For pathway enrichment analysis, hypergeometric test was used to identify significantly enriched metabolic pathways or signal transduction pathways in DEGs comparing to the genome background. Figure 3.4 shows the calculation formula, where N is the number of genes with KEGG annotation, n is the number of DEGs in N ; M is the number of genes related to specific pathways and m is the number of DEGs in M . Pathways with Q value ≤ 0.05 were considered significantly enriched in DEGs. The Q value was used to determine the P value threshold. Q values are the name given to the adjusted P values found using an optimized FDR approach. The FDR approach is optimized by using characteristics of the P value distribution to produce a list of Q value. Q value (FDR adjusted P value) of 0.05 implies that 5% of significant tests will result in false positive.

$$P = 1 - \sum_{i=0}^{m-1} \frac{\binom{M}{i} \binom{N-M}{n-i}}{\binom{N}{i}}$$

Figure 3.4: The probability formula to identify significantly enriched metabolic pathways or signal transduction pathways in DEGs comparing to the genome background.

N is the number of genes with KEGG annotation;

M is the number of genes related to specific pathways;

n is the number of DEGs in N ;

m is the number of DEGs in M .

3.4.5 Identification of genes in phenylpropanoid and flavonoid pathway

The genes that mapped to phenylpropanoid and flavonoid pathways in KEGG pathway were identified. Additionally, the expression fold changes either up-regulated or down-regulated were also determined.

3.4.6 Analysis on genes involved in panduratin A proposed pathway

3.4.6.1 Gene identification and analysis

Flavonoid O-methyltransferase and prenyltransferase were two enzymes that were proposed to be directly involved in panduratin A production. Initially, to elucidate flavonoid O-methyltransferase and prenyltransferase from *B. rotunda* transcriptome data, the unigene sequences that were annotated as flavonoid O-methyltransferase and flavonoid prenyltransferase were extracted from the data. Every unigene was checked manually for the nucleotide sequences, domain region, open reading frame (ORF) and gene fold changes. First, the nucleotide sequences that contain gaps were eliminated. Next, a protein translates software - ExPasy (<http://web.expasy.org/translate/>) was employed to predict the ORF of the unigene sequences. Complete ORF consists of both start and stop codons. Then, using the predicted ORF sequence, protein similarity search against the GenBank database was performed using Protein blast to check the conserved domain region and protein annotation.

3.4.6.2 Multiple sequence alignments

The multiple sequence alignment of flavonoid O-methyltransferase and prenyltransferase proteins were done using Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>) by employing default parameters. Flavonoid

O-methyltransferase and prenyltransferase protein sequences from other plants used in the multiple sequence alignment were downloaded from GenBank (Appendix J).

3.4.6.3 Phylogenetic tree analysis

A phylogenetic tree of flavonoid O-methyltransferase and prenyltransferase from *B. rotunda* were constructed using neighbour-joining method (MEGA 5.10 software). The phylogenetic tree was constructed using the following parameters: Poisson correction, pairwise deletion and bootstrap analysis with 1000 replicates. The flavonoid O-methyltransferase and prenyltransferase sequences were used for phylogenetic tree constructions were downloaded from GenBank (Appendix J). The selection of the flavonoid O-methyltransferase and prenyltransferase protein sequences from other plants for phylogenetic tree analysis were based on the protein similarity to the corresponding unigenes annotated as flavonoid O-methyltransferase and prenyltransferase from blastx results, respectively. Besides, well characterized flavonoid O-methyltransferase and prenyltransferase proteins from other plants were selected.

3.4.7 Identification and classification of transcription factors and transcription regulators

Transcription factors (TFs) and transcription regulators (TRs) in *B. rotunda* were identified and classified using iTAK software (<http://bioinfo.bti.cornell.edu/cgi-bin/itak/index.cgi>). iTAK is a program to identify and classify plant transcription factors (TFs) and transcription regulators (TRs) from protein or nucleotide sequences based on the rules (required and forbidden protein domains of each gene family) described in (Perez-Rodriguez et al., 2009). Protein sequences that were translated from nucleotide

sequences generated from Illumina sequencer were used to find both transcription factors and transcription regulators. iTAK searches both TFs and TRs based on homology search using TFs and TRs from rice database. Subsequently, the differentially regulated TFs and TRs in response to phenylalanine were also identified.

3.4.8 Analysis on R2R3 MYB transcription factors

3.4.8.1 Identification and analysis of R2R3 MYB transcription factors

All unigenes that were annotated as R2R3 MYB transcription factors were identified from *B. rotunda* transcriptome data for further analysis. Each unigene was checked manually for the nucleotide sequences, domain region, ORF and gene regulation. First, the nucleotide sequences that contain gaps were eliminated. Then, using the remaining unigene sequences, protein similarity search against the GenBank database was performed using Blastx to check the availability of the domain region. Next, a protein translates software; ExPasy (<http://web.expasy.org/translate/>) was employed to predict the ORF of the unigene sequences. Complete ORF consists of both start and stop codons. Lastly, the unigene regulation profile was determined by comparing the fold changes of R2R3 MYB transcription factors in control and phenylalanine treated samples.

3.4.8.2 Multiple sequence alignments

The sequences of predicted open reading frame of R2R3 MYB transcription factors in *B. rotunda* were manually aligned with other plants R2R3 MYB proteins using MEGA 5.2 software based on the location of the corresponding amino acids in the R2 and R3 repeat domains. Multiple sequence alignment was done using Clustal Omega using default parameters (<http://www.ebi.ac.uk/Tools/msa/clustalo/>).

3.4.8.3 Protein motif identification

Protein motif identification was done using the MEME version 4.10.1 tool (<http://meme-suite.org/tools/meme>) to identify the conserved motif shared among *B. rotunda* R2R3 MYB proteins.

3.4.8.4 Phylogenetic tree analysis

A phylogenetic tree of R2R3 MYB transcription factors from *B. rotunda* were constructed using neighbour-joining method (MEGA 5.10 software). The phylogenetic tree was constructed using the following parameters: Poisson correction, pairwise deletion and bootstrap analysis with 1000 replicates. The plant R2R3 MYB proteins used for phylogenetic tree constructions were downloaded from the GenBank (Appendix J). In this study, phylogenetic tree analysis was done to predict the function of R2R3 MYB transcription factors in *B. rotunda*.

3.5 Validation and expression pattern analysis

Unigene expression data in *B. rotunda* transcriptome was validated by quantitative real-time PCR (qRT-PCR) analysis. The RNA from control and 14 days post-phenylalanine treated cell suspension cultures sample was used for the validation. The RNA extraction was as described in 3.2.2. Subsequently, the cDNA was prepared as mentioned in 3.2.4.

3.5.1 Primer design

A total of 9 unigenes were selected for qPCR analysis to experimentally validate the transcriptional abundance results from the sequencing and computational analysis. The unigenes include 4 random up-regulated unigenes (Unigene58054_All, Unigene57613_All, Unigene555838_All and Unigene54651_All) and 5 unigenes that are annotated in the flavonoid pathway (Unigene10327_All; PAL, Unigene 67845_All; C4H, Unigene41852_All; 4CL, Unigene1735_All; CHS and Unigene49558_All; F3H). Elongation factor 1 alpha (*Ef1 α*) was used as the reference gene. Primers that were used for the experimental validation are shown in Table 3.1. Primers for qPCR were designed using Primer 3 software (<http://bioinfo.ut.ee/primer3-0.4.0/>) using the following parameters: primer length of 20 – 23 bp, product size of 120 – 150 bp, 45 – 60% of GC content and default primer annealing temperatures. The example of primer 3 output was presented in Appendix A.10. Next, the primer properties and PCR suitability test of the selected primers were analysed using Sequence Manipulation Suite: PCR Primer Stats software (http://www.bioinformatics.org/sms2/pcr_primer_stats.html). Appendix A.11 shows the example of PCR primer stats results. Primers that pass the PCR suitability tests with no GC clamp, no self-annealing and no hairpin formation were selected. All primers were synthesized by AITbiotech Pte Ltd, Singapore.

Table 3.1: Forward and reverse primer sequences used to validate transcriptome data which consist of 4 random up-regulated unigenes, 5 flavonoid unigenes and one reference unigene (*Ef1 α*).

Gene	Primers	Nucleotide sequence
Random up-regulated unigenes	Unigene58054_ All Forward	5'-ACAACGCCTTCAACAACCTC-3'
	Unigene58054_ All Reverse	5'-CTGGTGGTCTGGAATGTGGT-3'
	Unigene57613_ All Forward	5'-CGAGCTGTATGGGGAAGAAG-3'
	Unigene57613_ All Reverse	5'-GCCGCGTAGGAGATGAAGTA-3'
	Unigene55838_ All Forward	5'-GGTGGTGTACCAGGTTTCAGG-3'
	Unigene55838_ All Reverse	5'-GACGCCAGAGATCGTGACAG-3'
	Unigene54651_ All Forward	5'-ATAAACAGCCAGGGCAACAG-3'
	Unigene54651_ All Reverse	5'-TTCAGCATCAGATCCCCTTC-3'
PAL	Unigene10327_ All Forward	5'-GCTCTGCAAATGGACCTCA-3'
	Unigene10327_ All Reverse	5'-ATCGAGCAGTTGGTGTCTTC-3'
C4H	Unigene67845_ All Forward	5'-ATTCCTGTTTTGGGGCTTTC-3'
	Unigene67845_ All Reverse	5'-GGCATTGTAGCGTTGCTTT-3'
4CL	Unigene41852_ All Forward	5'-CCTAAATCTGCCTCAGGAAA-3'
	Unigene41852_ All Reverse	5'-GAGTTTACTGAAGGCTTTTGC-3'
CHS	Unigene1735_ All Forward	5'-GCTACAAGGCCAAGAAGACG-3'
	Unigene1735_ All Reverse	5'-CTCAGGGAGGTACGTGTCGT-3'
F3H	Unigene49558_ All Forward	5'-ATAAGACGCTCTCCCGTTGT-3'
	Unigene49558_ All Reverse	5'-ATGGGTTTCAGGCTTGACTCG-3'
<i>Ef1α</i>	<i>Ef1α_rt_F1</i>	5'-GAGGCTGGTATCTCAAAGG -3'
	<i>Ef1α_rt_R1</i>	5'-GTGCCTTTGAGTACTTTGG -3'

3.5.2 qPCR analysis

qPCR analysis was carried out to verify the transcriptome data. The method used for qPCR was described in 3.2.6. Primers used for target genes were shown in Table 3.1. Elongation factor 1-alpha was used as the reference genes. For non-template control, sterile distilled water was used as template. The relative expression levels of the selected unigenes normalized to elongation factor was calculated using $2^{-\Delta\Delta C_t}$ method. All reactions were performed with four experimental replicates and data were analyzed using QuantStudio 12 K Flex software. For transcriptome data validation, the fold changes between control and post phenylalanine treatment from sequencing was compared to the fold changes generated from the qPCR analysis.

3.6 Sequencing of MYB transcription factor in *B. rotunda*

For sequencing the MYB transcription factor from *B. rotunda* cell suspension cultures, DNA and RNA from 14 days post-phenylalanine treated cell suspension cultures was extracted using methods as described in subsections 3.2.1 and 3.2.3, respectively. Then, using DNA-free RNA, the cDNA was prepared using methods as described in subsection 3.2.4. The total RNA used to prepare the cDNA was 1.5 µg.

Subsequently, PCR verification by amplifying MYB1_Partial gene was done to ensure no traces of DNA presence in the cDNA sample. Therefore, DNase-treated RNA sample was used as the PCR template. The PCR reaction mixture and parameters was done as described in 3.2.5. The sequence of forward MYB1_Partial primer is 5'-TTCCCTCCCTGGTCGTACTIONGA-3', while reverse MYB1_Partial primer is 5'-CGGATATGGACAAGGCAAAA-3'. Forward and reverse primers were designed based on unigene14544_All sequence. For positive control, genomic DNA was used as template while for negative control; sterile distilled water was used as the template. Non reverse transcriptase cDNA also used as a negative control template.

3.6.1 Primer design

Forward and reverse primers to amplify the putative *BrMYB1*, *BrMYB2* and *BrMYB3* ORF were designed manually using MEGA 5.10 software based on Unigene14544_All, Unigene51701_All and Unigene19149_All sequences; respectively. Primers were designed with the following criteria: primer length of 20 bp, 40 – 60% of GC content and default primer annealing temperatures. Then, the primer location within the sequences, primer properties and PCR suitability test of the selected primers were analysed using Primer3 software (<http://bioinfo.ut.ee/primer3-0.4.0/>) and Sequence Manipulation Suite: PCR Primer Stats software (http://www.bioinformatics.org/sms2/pcr_primer_stats.html).

Primers that pass the PCR suitability tests with no GC clamp, no self-annealing and no hairpin formation were selected. All primers were synthesized by AITbiotech Pte Ltd, Singapore. Figure 3.5 show the illustration of forward and reverse primer position of putative *BrMYB1*, *BrMYB2* and *BrMYB3*. The forward primers were located at the upstream of start codon and the reverse primers were located at downstream of stop codon. Table 3.2 shows all of the forward and primer sequences to amplify the target MYB genes.

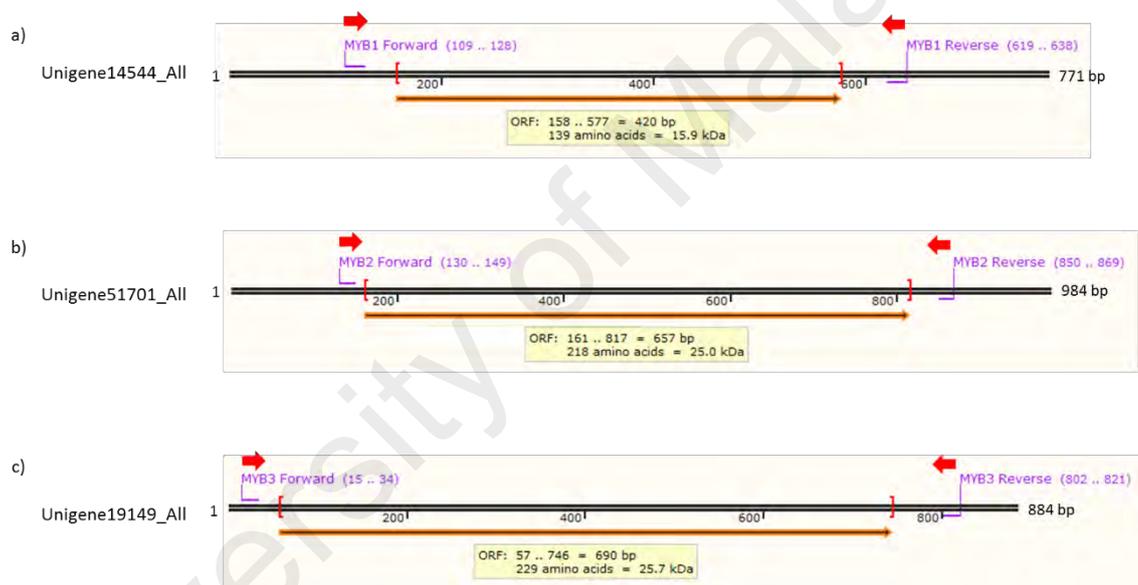


Figure 3.5: The position of a) putative *BrMYB1*, b) putative *BrMYB2* and c) putative *BrMYB3* primers relative to the Unigene14544_All, Unigene51701_All and Unigene19149_All, respectively.

The bold red arrows represent the primers. The orange arrow represents the ORF region. The red open and close brackets indicate start and stop codon; respectively.

Table 3.2: Forward and reverse primer sequences to amplify the putative *BrMYB1*, *BrMYB2* and *BrMYB3*.

Gene	Primers	Nucleotide sequence
<i>BrMYB1</i>	MYB1 Forward	5'-GACTTGGGGATGTTGAGGTG-3'
	MYB1 Reverse	5'-ACAGAGGCAATCAGCTGAGC-3'
<i>BrMYB2</i>	MYB2 Forward	5'-TAGAGAGATTGTGTATGCAC-3'
	MYB2 Reverse	5'-GAGCAGAACAGTACAGGAAG-3'
<i>BrMYB3</i>	MYB3 Forward	5'-ATGCACACCTAGCTATCCTC-3'
	MYB3 Reverse	5'-CAGCATCTTGTTTCAGATGGC-3'

3.6.2 Putative *BrMYB1*, *BrMYB2* and *BrMYB3* genes amplifications

Putative *BrMYB1*, *BrMYB2* and *BrMYB3* gene amplifications from both DNA and cDNA was done using the PCR reaction mixture and parameters as described in subsection 3.2.5. The cDNA template used was free from genomic DNA contaminations as verified previously in section 3.6. For negative control; sterile distilled water was used as the replacement. In order to select the best annealing temperature to amplify putative *BrMYB1*, *BrMYB2* and *BrMYB3* genes, gradient PCR was carried out. Finally, to sequence the genes, PCR was done using the selected annealing temperature using HotStartTaq DNA polymerase.

3.6.3 Gene cloning

3.6.3.1 Purification of PCR fragments

Gel purification of target gene fragments were performed according to the manufacturer's instructions (Qiagen). A total of 100 µl of PCR products was electrophoresed on 1% (w/v) agarose gel electrophoresis. The specific bands corresponding to the amplicon size was excised using a clean, sharp blade. The gel was transferred into a colorless tube. Then, the weight of the gel was measured. After that, 3 volume of Buffer QC was added into the tube. To dissolve the gel, the tube was incubated

at 50°C for 10 minutes and mixed 2 -3 times. Then, 1 volume of isopropanol was added to the sample and the mixture was mixed well. QIAquick spin column was placed in a provided 2 ml collection tube. Next, the column was centrifuged at 16,532.425 X g (13,000 rpm) for 1 minute (K2015R centrifuge, Centurion Scientific, Laguna Niguel, USA). The flow-through was discarded. Then, 500 µl of Buffer QC was added followed by centrifugation at 16 532,425 X g (13,000 rpm) for 1 minute. About 750 µl of PE was added to wash the column. Next, the column was centrifuged at 16,532.425 X g (13,000 rpm) for 1 minute and the flow through was discarded. The centrifugation step was repeated to completely remove the residual washing buffer. Finally, for elution of purified product, QIAquick column was placed in a 1.5 ml microcentrifuge tube and 30 µl of sterile distilled water was added to the center of the QIAquick membrane. After 1 minute the column stand at room temperature, the column was centrifuged at 16,532.425 X g (13,000 rpm) for 1 minute.

Agarose gel electrophoresis was carried out to confirm the size of gel purified PCR product. About 2 µl (1 µg) of 1kb Plus DNA ladder (Thermoscientific, USA) and 5 µl of gel purified PCR product were loaded into 1 % (w/v) agarose gel. Prior to loading the sample, the gel purified PCR product was mixed with 1 µl of 6 X loading dye (Fermentas, USA). The agarose gel was electrophoresed at 120 volts for 25 min. The migrated bands were visualized using ChemiDoc MP System (Biorad, California, USA).

3.6.3.2 Ligation into pCR4-TOPO vector

The purified PCR products were ligated and cloned into pCR4-TOPO vector (TOPO TA Cloning Kit, Invitrogen, Carlsbad, USA) by following the manufacturer's instructions. Ligation reactions were carried out in mixture of 4 µl of purified PCR product, 1 µl of salt solution (1.2 M NaCl and 0.06 M MgCl) and 1 µl (10 ng) of TOPO

vector. The reaction mixture was mixed gently and incubated at room temperature for 5 minutes. Then, the mixture was placed on ice and ready for transformation.

3.6.3.3 Transformation One Shot TOP10 competent cells

TOPO TA vector harbouring the PCR product was introduced into TOP10 competent cells according to the manual's instruction (TOPO TA Cloning Kit, Invitrogen, Carlsbad, USA). First, competent cells were thawed on ice. Next, 2 μ l of ligation mixture was added into the competent cells and mixed gently. Then, the mixture was incubated on ice for 20 minutes. The cells were heat-shocked for 30 seconds at 42°C followed by immediate transfer on ice. Next, 250 μ l of SOC medium (2% Tryptone, 0.5% Yeast Extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl, 10 mM MgSO₄ and 20 mM glucose) at room temperature was added to the tube. The culture was propagated at 37°C at 220 rpm for 1 hour. After that, 50 μ l of transformed culture with addition of 20 μ l of SOC media was spread on LB agar plate containing 50 μ g/ml kanamycin and incubated overnight (16 hours) at 37°C. After incubation, only positive recombinants were obtained on the selective plate.

3.6.3.4 Bacteria colony screening

To analyze the positive clones, PCR library colony screening was carried out. Single colony was transferred to a labeled grid selection media (LB agar containing 50 μ g/ml of kanamycin) using sterile tooth pick. The plates were incubated for 8 hours at 37°C. To prepare the template for PCR, the same tooth pick was immersed into 30 μ l of sterile distilled water and boiled at 98°C for 10 minutes. The PCR reaction mixture and parameters are as described in 3.2.5.

3.6.3.5 Propagating the positive transformants

The positive transformants from labeled grid media were transferred and cultured on 10 ml of LB broth (containing 50 µg/ml of kanamycin). The broth was propagated overnight at 37°C at 220 rpm (LM-575 D incubator shaker, YIHDER, Beijing, China).

3.6.3.6 Plasmid Purification

Plasmids containing the target sequences were extracted using Wizard Plus SV Minipreps DNA Purification System (Promega, Madison, USA) according to the manufacturer's instructions. First, overnight culture was centrifuged at 4,829.76 X g (6,000 rpm) for 15 minutes (K2015 centrifuge, Centurion Scientific, Laguna Niguel, USA). Then, the entire medium was decanted. About 250 µl of Cell Resuspension Solution was added to resuspend the cell pellet. The resuspended cells were transferred into a new 1.5 ml microcentrifuge tube. To lyse the cells, 250 µl of Cell Lysis Solution was added to the sample. The mixture was inverted few times to mix. Then, 10 µl of Alkaline Protease Solution was added to the sample. The mixture was inverted few times and incubated at room temperature for 5 minutes. Next, 350 µl of Neutralization Solution was added to the sample and inverted few times to mix. The sample was centrifuged at 16,532.425 X g (13,000 rpm) for 10 minutes at room temperature (K2015R centrifuge, Centurion Scientific, Laguna Niguel, USA).

The cleared lysate was transferred into the spin column and centrifuged at 16,532.425 X g (13,000 rpm) for 1 minute at room temperature to bind the plasmid DNA. The flow through was discarded. About 750 µl of Wash Solution was added into the spin column and the spin column was centrifuged at 16,532.425 X g (13,000 rpm) for 1 minute to wash the plasmid. The washing step was repeated with 250 µl of Wash Solution. In order to completely remove the Wash Solution in the spin column, additional centrifugation at

16,532.425 X g (13,000 rpm) for 2 minutes was done. The Spin Column was transferred into 1.5 ml microcentrifuge tube. About 30 µl of Nuclease-Free water was added into the spin column followed by centrifugation at 16,532.425 X g (13,000 rpm) for 1 minute at room temperature to elute the plasmid DNA. The plasmid was stored at -20°C.

3.6.3.7 Preparation of plasmid DNA for sequencing

For sequencing, 15 µl of prepared plasmids harbouring putative *BrMYB1*, *BrMYB2* and *BrMYB3* fragment sequences were transferred into 1.5 ml microcentrifuge tubes. Representative clones were sent to AITbiotech Pte Ltd, Singapore and sequenced using automated DNA sequencer. M13 forward and M13 reverse primers were employed to sequence the target fragments. Both M13 forward and reverse sequences were located in the pCR4-TOPO vector.

3.6.4 Sequence analysis

The putative sequence of *BrMYB1*, *BrMYB2* and *BrMYB3* were checked manually by looking at the chromatogram peaks. A good sequencing results were indicated by a good chromatogram peaks. Then, the vector sequences were removed from the sequence by using MEGA 5.10 software. In order to check the reliability of the reference transcriptomic data sequence, multiple sequence alignment of nucleotide sequences of transcriptomic unigenes (Unigene14544_All, Unigene51701_All or Unigene19149_All), MYB from DNA and cDNA template (putative *BrMYB1*, *BrMYB2* or *BrMYB3*) was conducted using Clustal Omega. In order to confirm the annotation of Unigene14544_All, Unigene51701_All, or Unigene19149_All encoded as MYB protein, Blastx homology search using putative *BrMYB1*, *BrMYB2* and *BrMYB3* generated from cDNA template

and Blastp using predicted putative BrMYB1, BrMYB2 and BrMYB3 ORF were performed.

Subsequently, a phylogenetic tree was constructed using neighbour-joining method (MEGA 5.10 software). The phylogenetic tree was constructed using the following parameters: Poisson correction, pairwise deletion and bootstrap analysis with 1000 replicates. In this study, phylogenetic tree analysis was done to further classify putative BrMYB1, BrMYB2 and BrMYB3; and predict their functions through motif homologies. The selection of the MYB protein sequences from other plants for phylogenetic tree analysis were based on the protein similarity to the corresponding putative BrMYB encoded as MYB proteins from blastx results, respectively. Besides, well characterized MYB proteins from other plants were also selected based on the MYB repeat domain and R2R2 MYB subgroups. For putative BrMYB1, R3-type MYB proteins involved in flavonoid pathway were chosen. In addition, for putative BrMYB2, subgroup 4 R2R3-type MYB protein with the repressor motifs and putative BrMYB3, R2R3-MYB proteins from subgroup 4, 5, 6, 15 and 27 that contain bHLH interaction motif were chosen.

For molecular characterization of putative BrMYB1, BrMYB2 and BrMYB3, multiple sequence alignment with selected MYB proteins from other plants was conducted using Clustal Omega software <http://www.ebi.ac.uk/Tools/msa/clustalo/>. The other plant MYB proteins used for phylogenetic tree and multiple sequence alignment analysis were downloaded from the GenBank (Appendix J).

3.7 Putative *BrMYB2* transcription factor in response to phenylalanine treatment

In order to study putative *BrMYB2* gene regulation pattern in response to phenylalanine treatment, the cell suspension cultures of control and phenylalanine treated were harvested after 1 day, 7 days and 14 days, respectively. Total RNA extraction followed by cDNA preparation was done as described in 3.2.3 and 3.2.4, respectively.

3.7.1 Primer design of reference genes for qPCR analysis

Elongation factor 1-alpha and β -tubulin were two reference genes that were chosen from *B. rotunda* transcriptome data. Forward and reverse primers to amplify the elongation factor 1-alpha and β -tubulin were designed using Primer3 software (<http://bioinfo.ut.ee/primer3-0.4.0/>) based on Unigene12153_All and Unigene35585_All sequences. Primers were designed using the following parameters: primer length of 20 bp, product size of 120 – 150 bp, 45 – 60% of GC content and default primer annealing temperatures. Next, the primer properties and PCR suitability test of the selected primers were analysed using Sequence Manipulation Suite: PCR Primer Stats software (http://www.bioinformatics.org/sms2/pcr_primer_stats.html). Primers that pass the PCR suitability tests, no self-annealing and no hairpin formation were selected. All primers were synthesized by AITbiotech Pte Ltd, Singapore. Table 3.3 shows all of the forward and primer sequences to amplify the reference genes in qPCR analysis.

Table 3.3: Forward and reverse primer sequences for reference genes in qPCR analysis

Gene	Primers	Nucleotide sequence
Elongation factor 1-alpha	EF1 α Forward	5'-GAGTTGAGACGGGTGTCCTC-3'
	EF1 α Reverse	5'-TTCCTGGAGAGCTTCATGGT-3'
β -tubulin	β -tub Forward	5'-GCCGAGAAGATGAGGGAGAT-3'
	β -tub Reverse	5'-GGCGTATTTTCCTGTGCTGT-3'

3.7.2 Primer design of putative *BrMYB2* gene primer design for qPCR analysis

Forward and reverse primers to amplify the putative *BrMYB2* gene was designed manually using MEGA 5.10 software based on the *BrMYB2* sequencing results. Primers were designed with the following criteria: primer length of 20 bp, 40 – 60% of GC content and default primer annealing temperatures. Then, the primer location within the sequences, primer properties and PCR suitability test of the selected primers were analysed using Primer3 software (<http://bioinfo.ut.ee/primer3-0.4.0/>) and Sequence Manipulation Suite: PCR Primer Stats software (http://www.bioinformatics.org/sms2/pcr_primer_stats.html). Primers that pass the PCR suitability tests with no GC clamp, no self-annealing and no hairpin formation were selected. All primers were synthesized by AITbiotech Pte Ltd, Singapore. Figure 3.6 shows the illustration of the forward and reverse primer positions of *BrMYB2* gene. The forward primer is located at the R2 domain region while the reverse primer is located at the R3 domain region; between two exon-exon junctions. Table 3.4 shows all of the forward and primer sequences to amplify the putative *BrMYB2* gene in qPCR analysis.

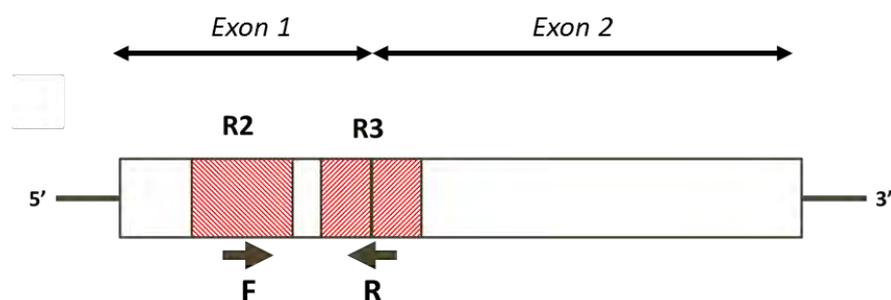


Figure 3.6: The position putative *BrMYB2* primers for qPCR analysis.

There were two exons indicated by the block and red striped box indicate the domain regions. The bold brown arrows represent the forward (F) and reverse (R) primers.

Table 3.4: Forward and reverse primer sequences for *BrMYB2* gene in qPCR analysis

Gene	Primers	Nucleotide sequence
<i>BrMYB2</i>	MYB2 Forward	5'-TTCTCCGATGCGGCAAGAGC-3'
	MYB2 Reverse	5'-ATAAGAGACCATTTGTTGCC-3'

3.7.3 qPCR analysis

qPCR analysis was carried out to study the putative *BrMYB2* gene regulation patterns in response to phenylalanine treatment during 1, 7 and 14 days post-treatment. The qPCR was done as described in subsection 3.2.6. Forward and reverse putative *BrMYB2* primers used in qPCR analysis as showed in Table 3.4. Elongation factor 1-alpha and β -tubulin genes were used as the reference genes. For non-template control, sterile distilled water was used as template whereas for non-reverse transcriptase control, the cDNA was prepared without the reserve transcriptase to verify that the cDNA used for the qPCR analysis were free from any DNA contaminations. The relative expression levels of the putative *BrMYB2* gene was normalized to that of the elongation factor 1-alpha and β -tubulin; and calculated using $2^{-\Delta\Delta C_t}$ method. All reactions were performed with four experimental replicates and the data generated were analyzed using QuantStudio 12 K Flex software.

CHAPTER 4: RESULTS

4.1 Establishment of *B. rotunda* cell suspension cultures

Callus is a mass of undifferentiated cells that derived from plant tissue or explants. In this study, callus was induced from meristemic cells of *B. rotunda* shoot bud (Figure 4.1 a). There were three types of callus induced from meristemic cells of *B. rotunda* that include friable embryogenic callus, compact callus and mixed callus. Mixed calli consisted of both friable and compact callus. According to Tan (2005), panduratin A production was increased in the phenylalanine-treated cell suspension cultures that derived from mixed calli. As shown in Figure 4.1 b, mixed callus was chosen for the entire experiment in this thesis and was propagated in liquid media which was later referred to as cell suspension cultures (Figure 4.1 c and d).

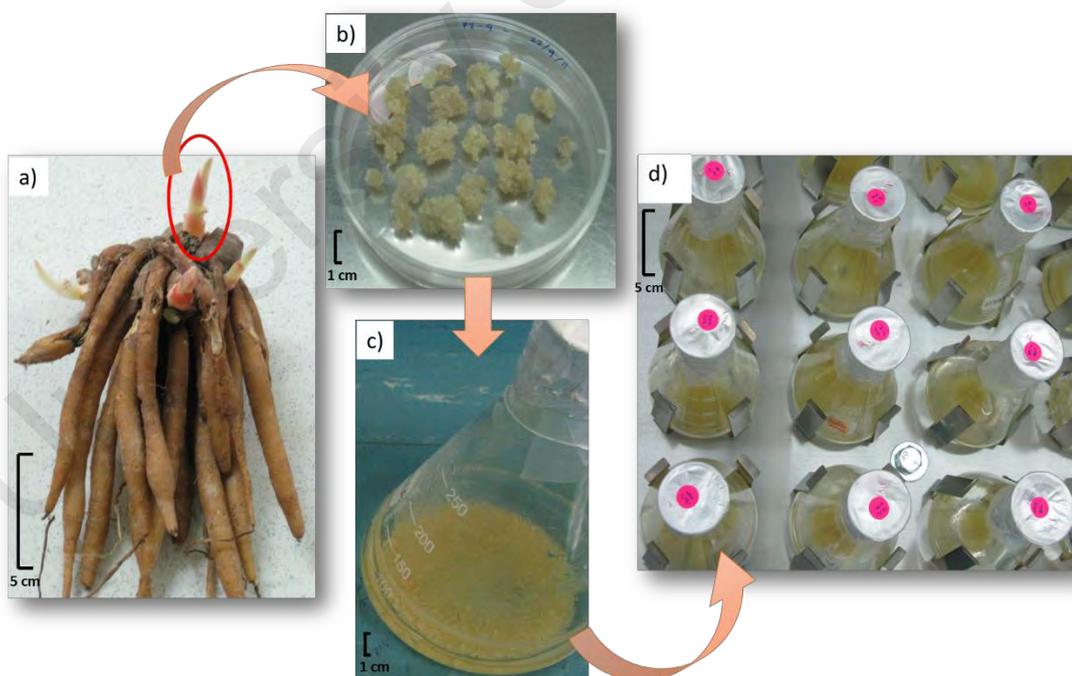


Figure 4.1: Establishment of *B. rotunda* cell suspension cultures.

a) *B. rotunda* rhizome with shoot bud indicated by red circle b) Propagation of mixed callus initiated from shoot bud c) Cell suspension cultures established from mixed callus and d) Mass propagation of cell suspension cultures in liquid media.

4.2 Total RNA assessment

Total RNA samples were extracted from control and phenylalanine treated *B. rotunda* callus using a modified CTAB method (Figure 4.2). Table 4.1 shows the concentration and purity assessment results for both RNA samples. The RNA integrity was assessed using an Agilent 2100 Bioanalyzer (Agilent, Waldbronn, Germany) (Figure A.2 in Appendix C) prior sending the samples for sequencing. The RNA samples were reassessed at the BGI, Shenzhen, China to ensure no degradation of the RNA during shipment (Figure A.3 and Table A.12 in Appendix C). From the results obtained, both RNA samples were of suitable quality for the transcriptome sequencing using Next-generation Illumina-Solexa platform, HiSeq 2000 Sequencing System (Illumina, Inc., USA). DNase treatment was carried out by the BGI upon receiving the RNA samples.

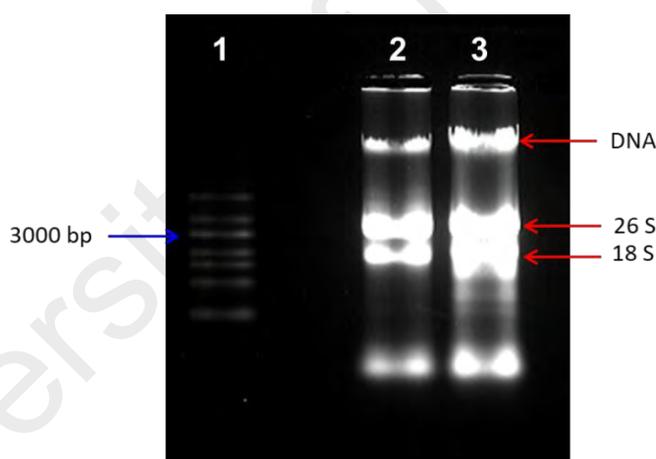


Figure 4.2: Total RNA extraction of *B. rotunda* callus using modified CTAB method. Lane 1: RiboRuler High Range RNA ladder (Thermo Scientific, USA); lane 2: total RNA extracted from control callus and lane 3: total RNA extracted from phenylalanine treated callus.

Table 4.1: Concentration and purity of total RNA from control and phenylalanine treated samples.

Samples	Concentration of RNA ($\mu\text{g}/\mu\text{l}$)	Ratio value of absorbance	
		A260/A280	A260/A230
Control	3.106	1.87	1.92
Treated	4.432	1.89	1.72

4.3 Transcriptome profile analysis

4.3.1 RNA-Seq quality control

Quality control is a crucial step to ensure that clean reads of RNA-Seq data are of high quality and suitable for subsequent analyses. Figure 4.3 showed an overview of the range of quality values across all bases at each position in the clean paired-end reads FastQ file of control and phenylalanine treated samples, while Figure 4.4 showed the quality score distribution over all sequences. From the FastQC reports, the quality scores of all paired-end clean reads from control and phenylalanine-treated samples were more than 30, which indicate a very good quality reads (Figure 4.3 and Figure 4.4).

Next, Figure 4.5 showed that paired-end clean reads from control and phenylalanine-treated samples produce biased sequence content at the start of the reads. However, this technical bias does not affect the downstream analysis. There were no base N content (Figure 4.6) and adaptor contaminations (Figure 4.7) in control and phenylalanine treated RNA-Seq data. Other FastQC reports such as GC distribution over all sequences (Figure A.4), distribution of sequence length over all sequences (Figure A.5), percent of sequences remaining (Figure A.6), Kmer content (Figure A.7), Kmer sequences (Table A.13) and overpresented sequences (Table A.14) were presented in Appendix D.1.

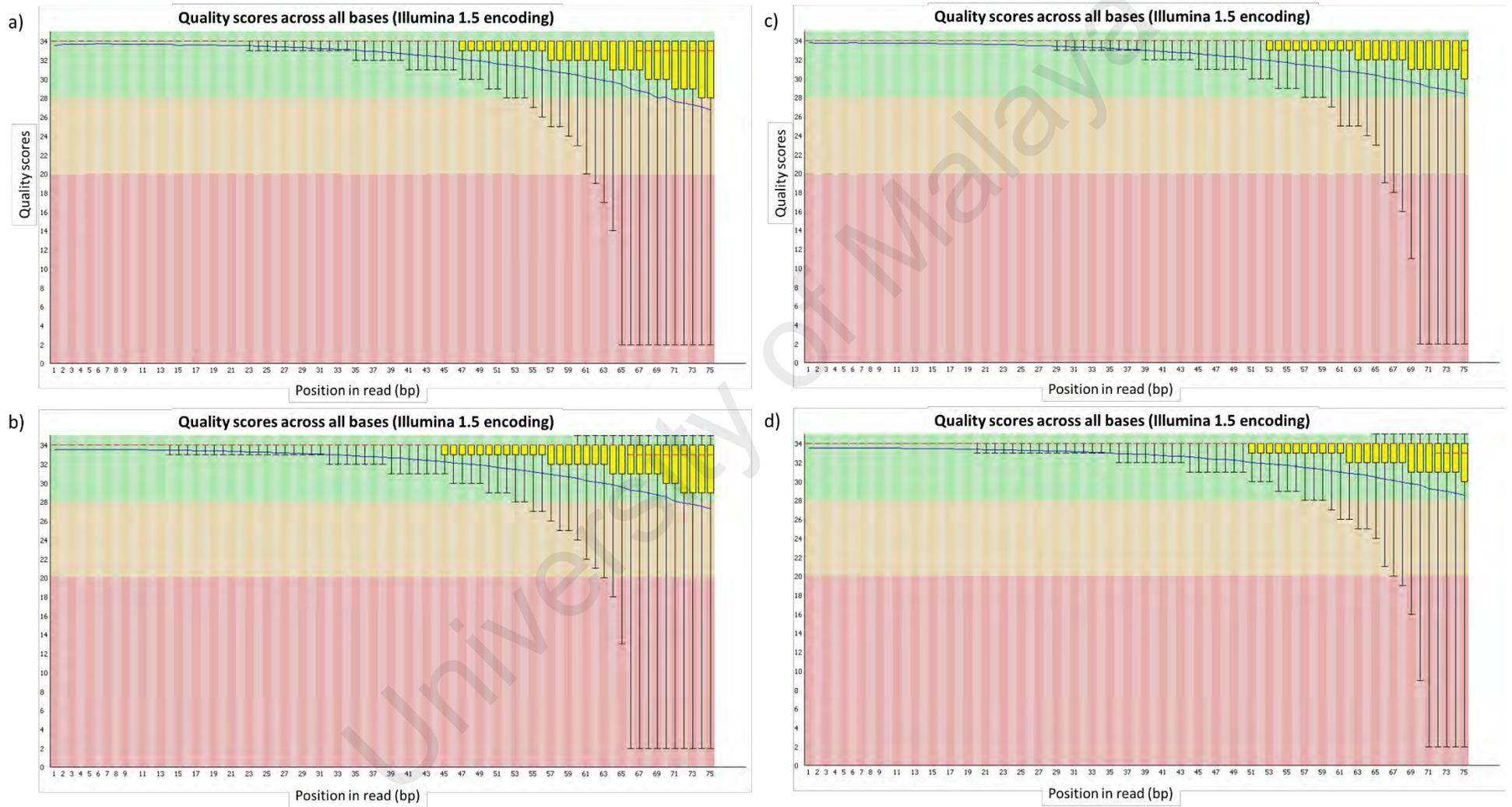


Figure 4.3: Quality scores across all bases of RNA-Seq data from paired-end reads of control (a and b) and phenylalanine-treated (c and d) samples..

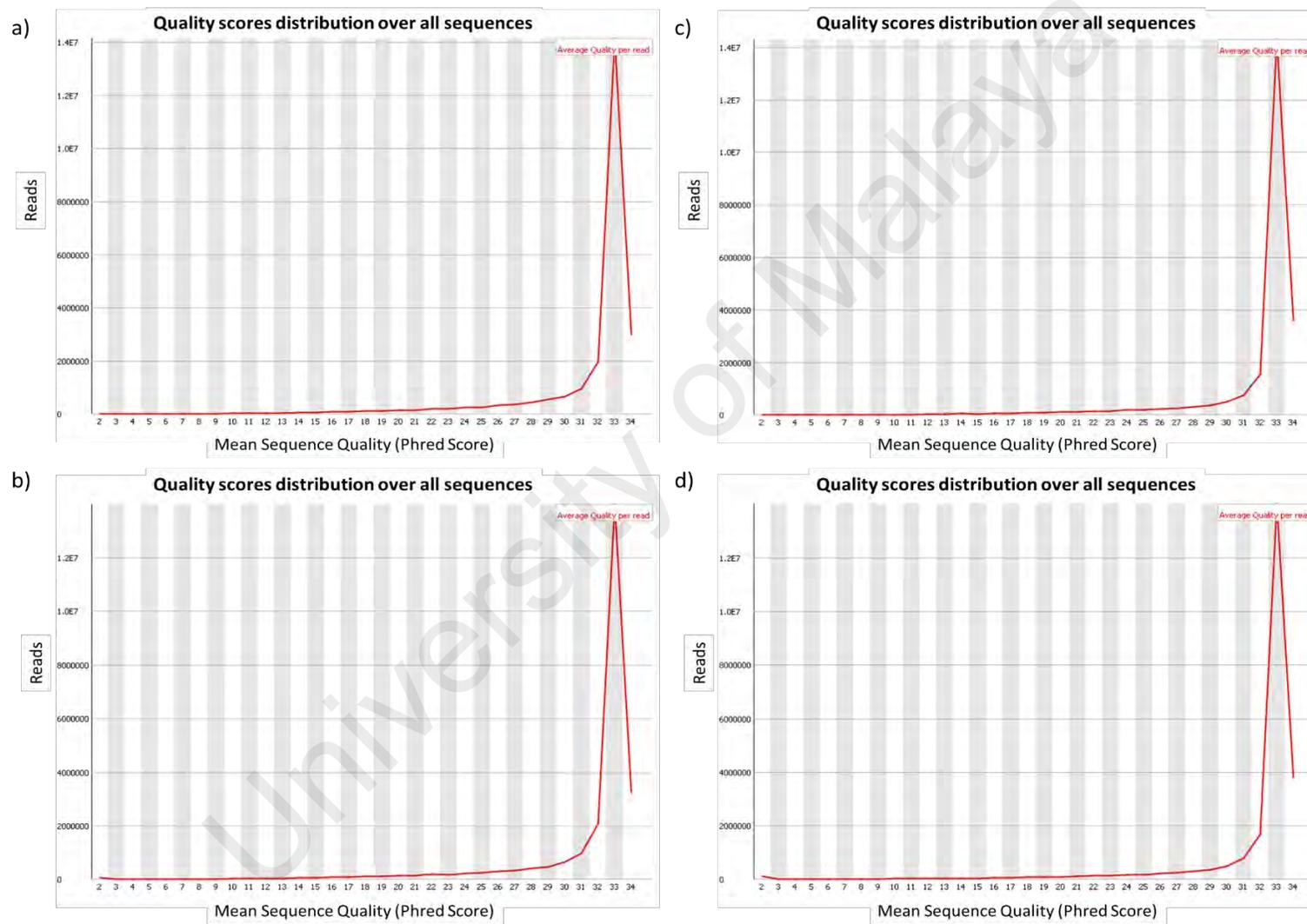


Figure 4.4: Quality scores distribution over all sequences of RNA-Seq data from paired-end reads of control (a and b) and phenylalanine-treated (c and d) samples.

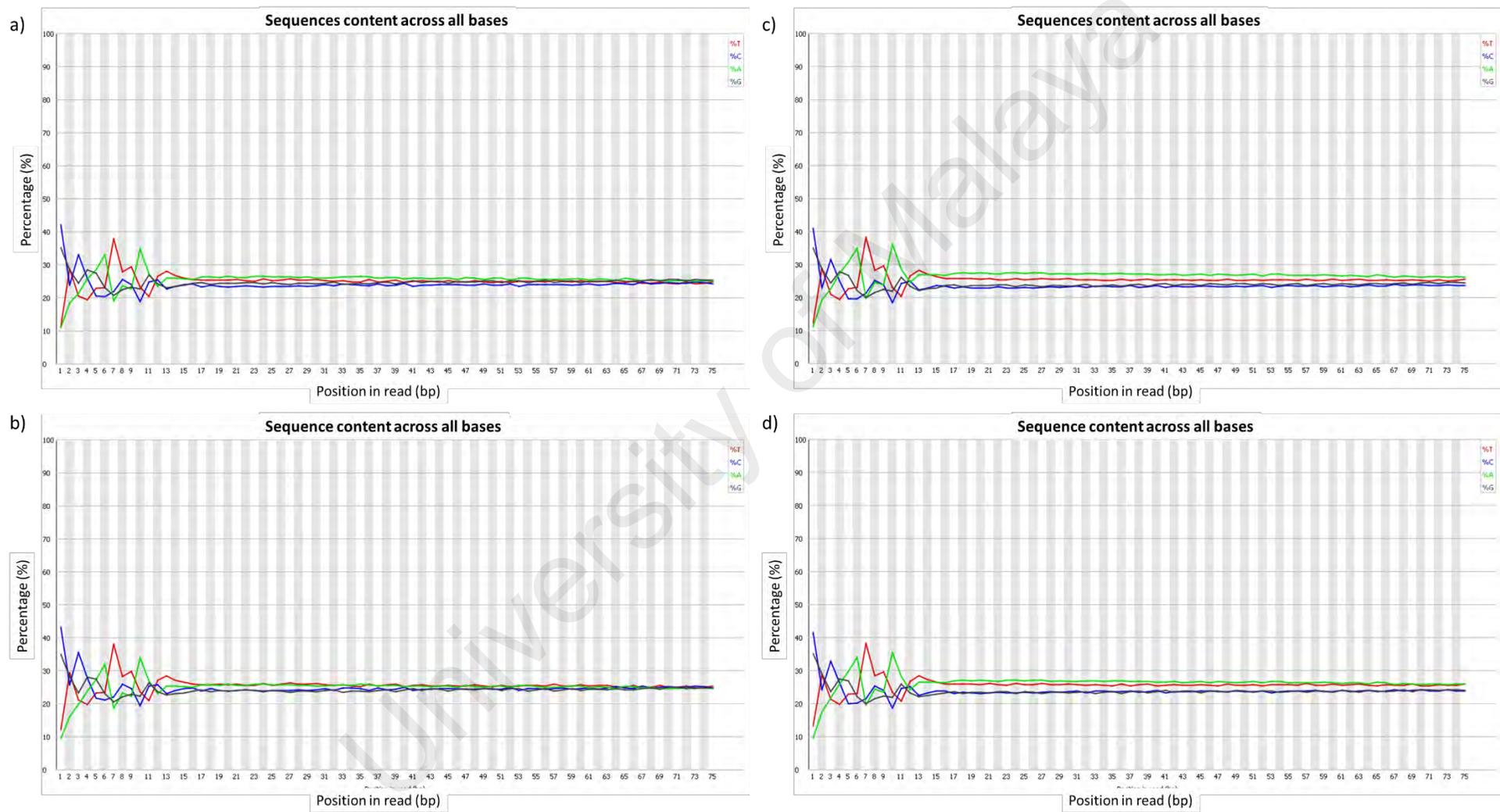


Figure 4.5: Sequence content of all bases in RNA-Seq data from paired-end reads of control (a and b) and phenylalanine-treated (c and d) samples.

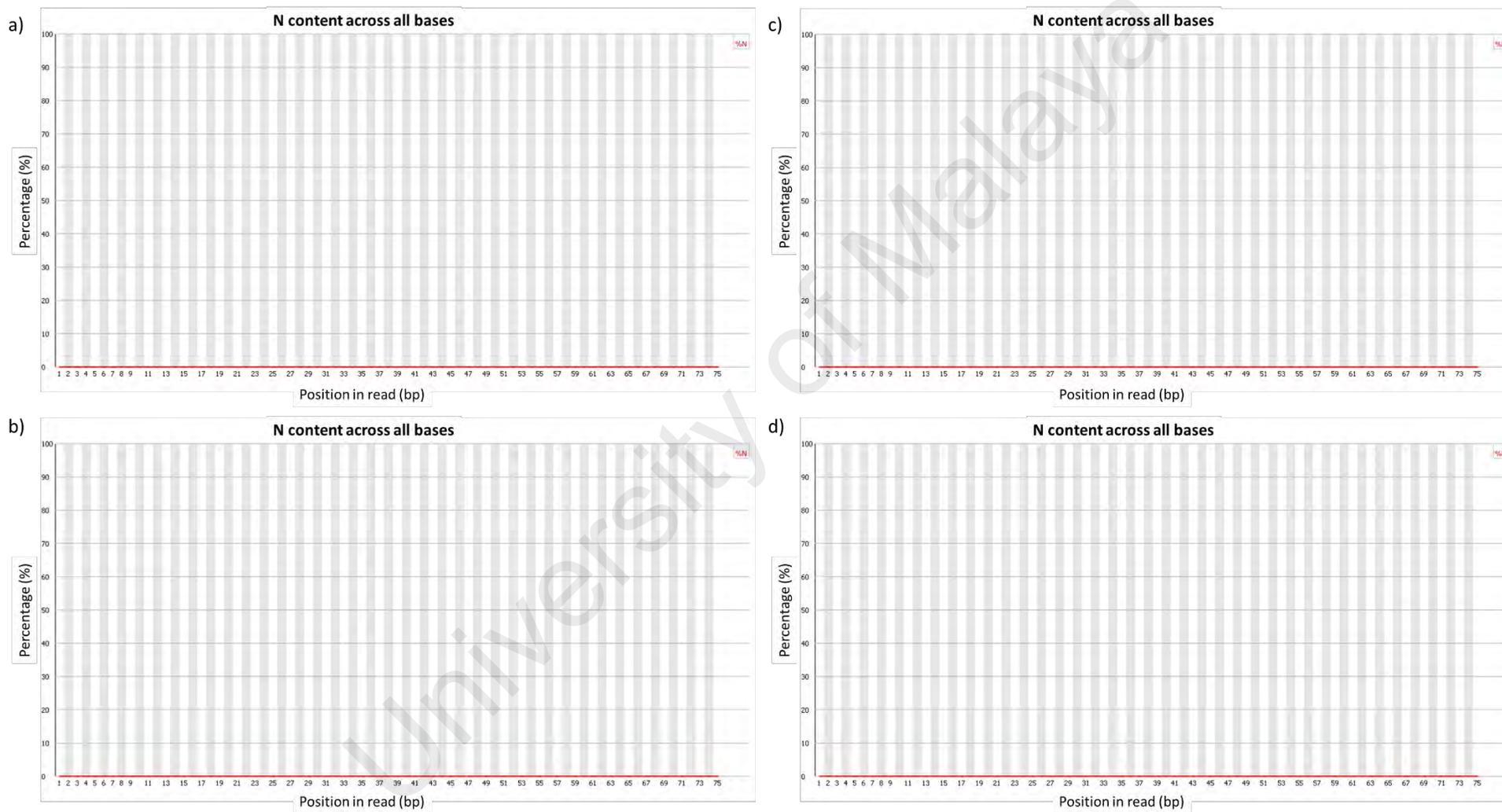


Figure 4.6: N content across all bases in RNA-Seq data from paired-end reads of control (a and b) and phenylalanine-treated (c and d) samples.

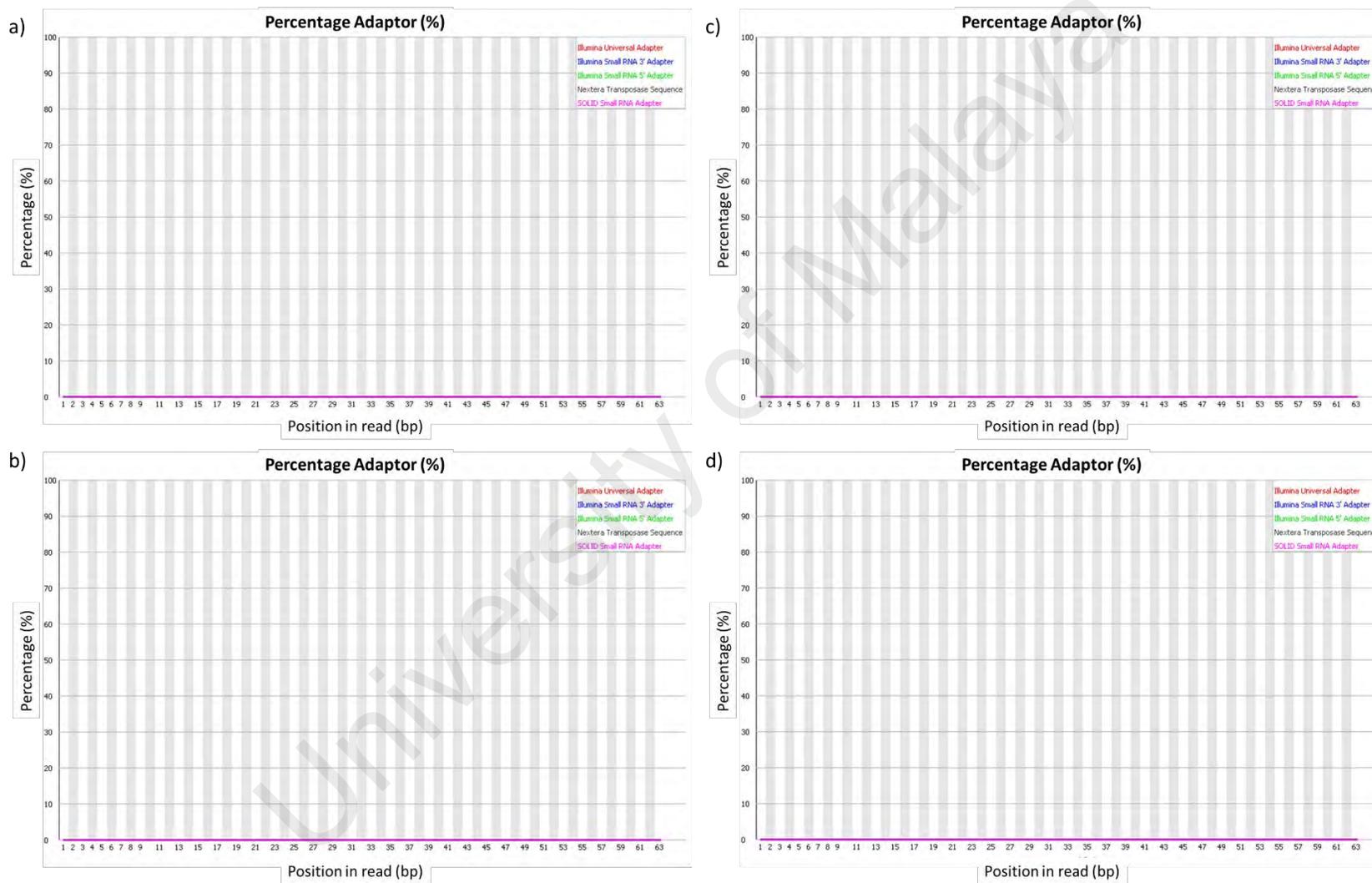


Figure 4.7: Adaptor content in RNA-Seq data from paired-end reads of control (a and b) and phenylalanine-treated (c and d) samples.

4.3.2 Short-read *de novo* sequencing and assembly

Illumina-Solexa RNA sequencing technology was used to sequence the whole transcriptome of *B. rotunda*. After stringent data filtering and quality checks, approximately 50 million high-quality clean reads were obtained from both samples with 95.13% and 96.06% Q20 bases (the number of bases with scores of greater than or equal to 20) for control and treated sample, respectively. In total, there were 24,473,594 and 23,470,648 clean paired-end reads generated with a total of 3,671,039,100 and 3,520,597,200 nucleotides from control *B. rotunda* callus and phenylalanine treated callus respectively (Table 4.2).

Table 4.2: Summary of reads assembly generated by SOAPdenovo from control and phenylalanine treated *B. rotunda* callus.

N50 size of contigs, scaffolds or unigene were calculated by ordering all sequences then adding the lengths from longest to shortest until the summed length exceeded 50% of the total length of all sequences.

	Control	Phenylalanine treated
Total number of reads	24,473,594	23,470,648
Total nucleotides (nt)	3,671,039,100	3,520,597,200
GC%	49.31%	47.89%
Q20%	95.13%	96.06%
Step-wise assembly		
Total number of contig	287,451	273,979
Average sequence size of contigs	199	191
N50 length of contig	236	221
Total number of scaffolds	149,648	147,381
Average sequence size of scaffolds	359	330
N50 length of scaffold	535	465
Total number of unigenes	78,998	77,541
Total nucleotides (nt) in unigenes	44,279,890	39,284,596
Average sequence size of unigenes	561	507
N50 length of unigene	703	610
Combined control and phenylalanine treated unigenes		
Total number of all unigenes	101,043	
Average sequence size of all unigenes	599	
N50 length of all unigene	804	
Unigenes with orientation	54,284	
Unigenes without orientation	46,759	

Clean reads that were generated from the Illumina Genome analyzer were assembled into contigs, scaffolds and unigenes using open source SOAPdenovo assembler program (Li et al., 2009). A total of 287,451 and 273,979 contigs with lengths ranging between 75 to 5,680 bp and 75 to 3,739 bp with N50 lengths of 236 and 221, were generated for control and phenylalanine treated samples, respectively. Contigs were then overlapped using paired-end read information to assemble into scaffolds. There were 149,648 and 147,381 scaffolds assembled from the control and treated samples with average scaffold sizes of 359 (control) and 330 (treated sample). Scaffolds from both samples ranged from 100 to 12,211 bp for the control and from 100 to 5,943 bp for the treated sample.

Subsequently, scaffolds were overlapped and paired-end reads were used to fill the scaffold gaps to obtain unigenes. For the control sample, there were 78,998 unigenes assembled with lengths ranging from 200 to 12,209 bp and a N50 length of 703 bp; while for the treated sample, there were 77,541 unigenes assembled with lengths ranging from 200 to 5,944 bp with a N50 length of 610 bp. Finally, longer sequences denoted as All Unigenes, were assembled by overlapping both unigenes of the control and phenylalanine treated samples followed by removing redundant sequences using TGICL software. There were in total approximately 101,043 All Unigenes assembled with lengths ranging from 200 to 12,209 bp. The N50 lengths of All unigenes were 804 bp. Appendix D.1 shows the length distribution. Appendix D.2 shows gap distribution of unigene of the control and phenylalanine treated samples and All Unigenes.

In order to determine the unigenes' sequence orientation, all unigenes were aligned using BlastX alignment (e-value < 1.00E -05) against four protein databases with the priority order of GenBank NR, Swiss-Prot, KEGG and COG. Remaining unaligned unigenes were analyzed using ESTscan software (Iseli et al., 1999) to predict the coding

regions and to decide on sequence direction. The best-aligned results showed that 54,284 unigenes are oriented while 46,759 are non-oriented unigenes (Table 4.2).

4.3.3 Functional annotation and gene ontology classification

Functional annotation gave information on protein function annotation, pathway annotation, COG annotation and GO annotation. Unigenes with sequence orientation were aligned against public protein databases such as NR, Swiss-Prot, KEGG and COG using BlastX homology search (e-value < 1.00E -05), which is based on sequence similarities to the published protein databases. In total, 50,932 (50.41%) unigenes were successfully annotated (Table 4.3). Most of the unigenes were annotated using the NR database (49.93%) followed by Swiss-Prot (34.63%), KEGG (24.07%) and COG (16.26%). The remainder had no matches.

Table 4.3: Function annotation of *B. rotunda* transcriptome data in four public protein databases.

The databases include NR, Swiss-Prot, KEGG and COG.

Public protein database	No. of unigene hits	Percentage
NR	50,447	49.93
Swiss-Prot	34,989	34.63
KEGG	24,316	24.07
COG	16,426	16.26
Total	50,932	50.41

Clusters Orthologous Groups of proteins (COG) database contain orthologous proteins that were classified under several categories. Unigenes were aligned to COG database to predict and classify their possible function. Figure 4.8 shows the distributions of 16,526 unigenes assigned into 25 orthologous clusters in COG. Some unigenes may be assigned into several clusters in COG categories; while some unigenes were assigned to the same

cluster but with different protein orthologous similarity. In total, there were 34,434 unigenes that were assigned to COG database (Appendix D.3). The majority of the unigenes were distributed in general function prediction (4,851) followed by transcription (3,691); and replication, recombination and repair (3,053). A total of 1,863 functionally unknown unigenes were identified. Whereas, 753 unigenes were assigned to secondary metabolites biosynthesis, transport and catabolism and 395 as defense mechanism unigenes.

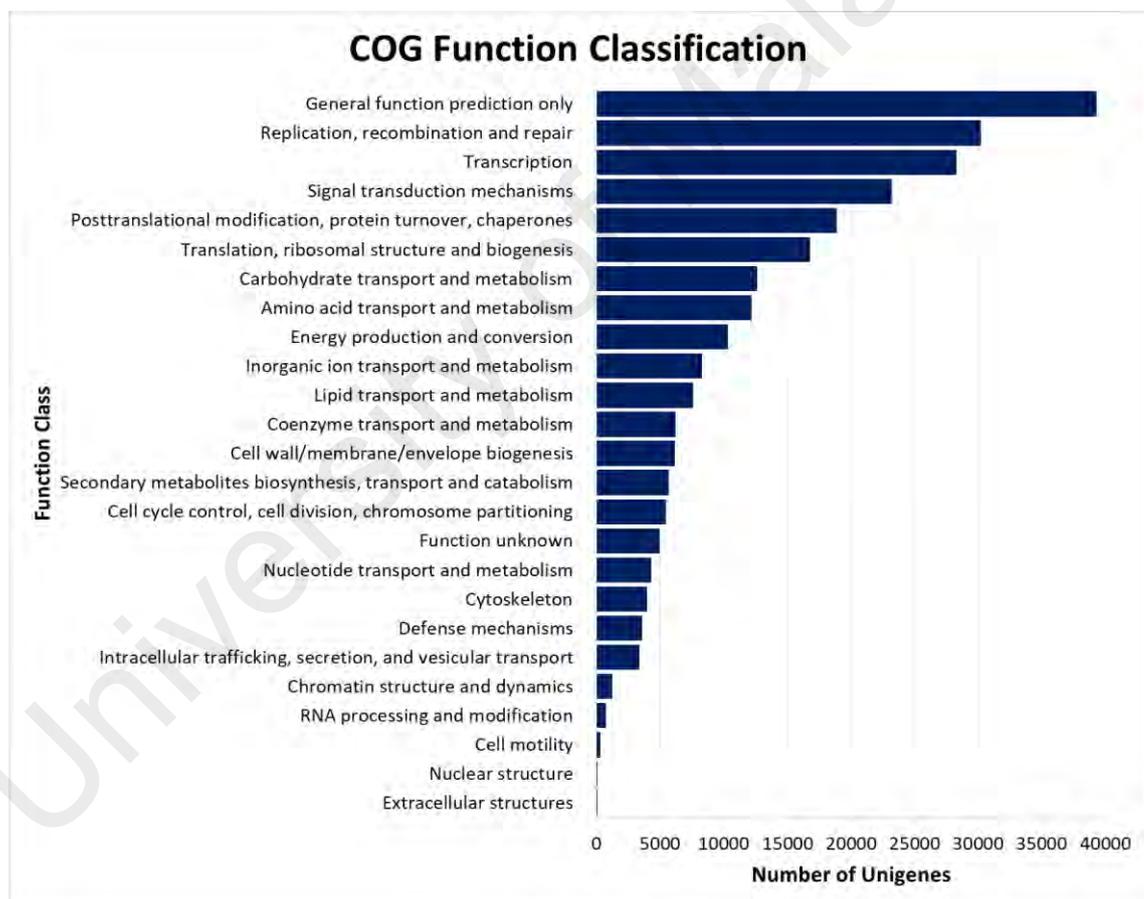


Figure 4.8: COG functional annotation of *B. rotunda* transcripts.

Unigene with NR annotation was further annotated and classified under GO. GO is an international standardized gene functional classification system. It has three ontologies which include molecular function, cellular component and biological properties. The basic unit of GO is GO-term and every GO-term belongs to a type of ontology. Figure 4.9 shows the distribution of unigenes assigned in GO. In total, 33,984 unigenes were mapped to GO with 7,451 unigenes assigned to molecular function, 16,493 unigenes assigned to cellular components and 10,040 unigenes assigned to biological process (Appendix D.4). One unigene may be assigned into several different GO-terms.

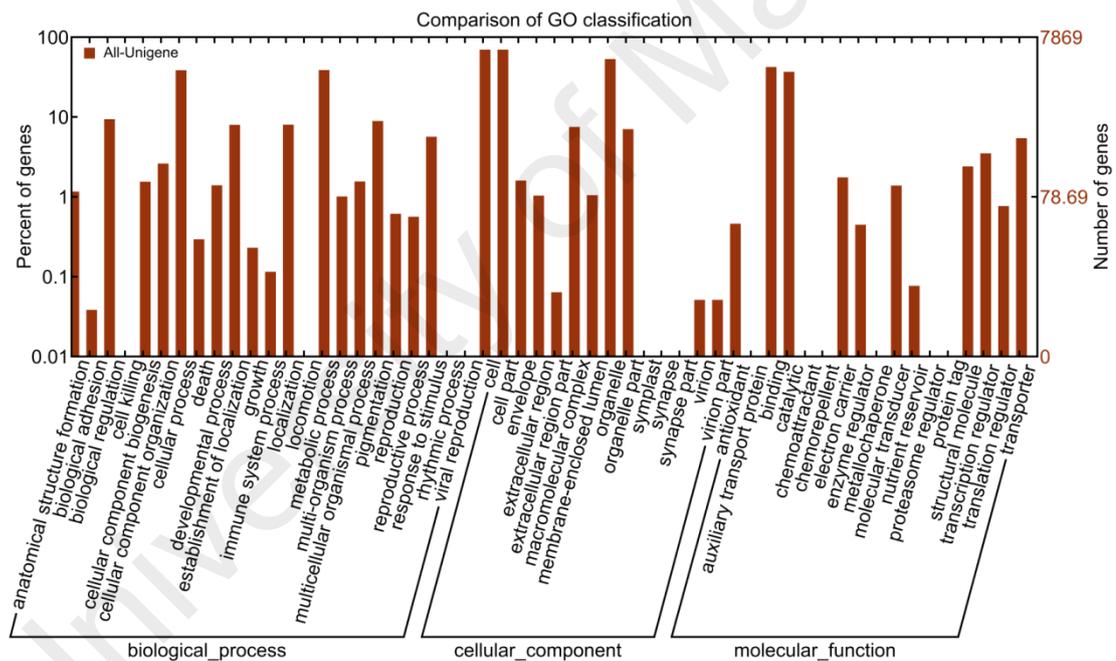


Figure 4.9: Histogram presentation of unigene distributions in GO functional classification. Unigenes were further classified into sub-groups in biological process, cellular component and molecular function.

4.3.4 Differentially expressed unigenes analysis

Unigene expression was calculated using reads per kb per million reads (RPKM) method. However, in order to distinguish between significant and non-significant DEGs, additional equations were employed. DEGs were determined using Poisson distribution equation, with the set threshold of False Discovery Rate (FDR) lower or equal to 0.001 and the absolute value of \log_2 (Ratio of RPKM of phenylalanine treated over RPKM of control) more or equal to 1. Through the calculation of \log_2 Ratio, up- and down-regulated transcripts were determined by comparing the fold changes of phenylalanine treated unigenes to their respective control unigenes.

The example of RPKM calculation was shown in Figure 4.10 to calculate the expression transcript of Unigene51701_All in control and phenylalanine treated samples. From RNA-Seq data, the unique mapped reads of control and phenylalanine treated samples were 53 and 203, respectively. The total number of reads for control and phenylalanine treated samples were 22,904,031 and 21,262,938, respectively. The length of Unigene51701_All was 984 bp. Figure 4.11 shows the fold changes of Unigene51701_All in phenylalanine treated sample compared to control which calculated using \log_2 Ratio equation. Since the absolute value of \log_2 Ratio calculation was more than one, Unigene51701_All was significantly up-regulated which shown 2.0447 higher fold changes in phenylalanine treated sample compared to the control. Besides, p value and FDR for Unigene51701_All were 6.20E-14 and 7.56E-15, respectively.

$$RPKM = \frac{10^6 C}{NL/10^3}$$

$$a) RPKM = \frac{10^6(53)}{[(22,904,031)(984)]/10^3}$$

$$RPKM = \frac{10^6(53)}{(2.25376 \times 10^{10})/10^3}$$

$$= 2.3516$$

$$b) RPKM = \frac{10^6(203)}{[(21,262,938)(984)]/10^3}$$

$$RPKM = \frac{10^6(53)}{(2.09227 \times 10^{10})/10^3}$$

$$= 9.7024$$

Figure 4.10: RPKM calculations for Unigene51701_All from control (a) and phenylalanine treated (b) samples.

C is the number of reads that uniquely mapped to unigene A;

N is the number of reads that uniquely mapped to all unigenes in the sample experiment;

L is the length of unigene A in basepair.

$$\begin{aligned} \text{Fold change} &= \log_2(RPKM \text{ phenylalanine treated} / RPKM \text{ control}) \\ &= \log_2(9.7024/2.3516) \\ &= \log_2(4.12587) \\ &= 2.0447 \end{aligned}$$

Figure 4.11: Fold changes calculation of Unigene51701_All in phenylalanine treated sample compared to control.

Figure 4.12 show the expression level of control and phenylalanine treated unigenes. In total, there were 14,644 up-regulated and 14,379 down-regulated unigenes showing significant differential expression (Table 4.4).

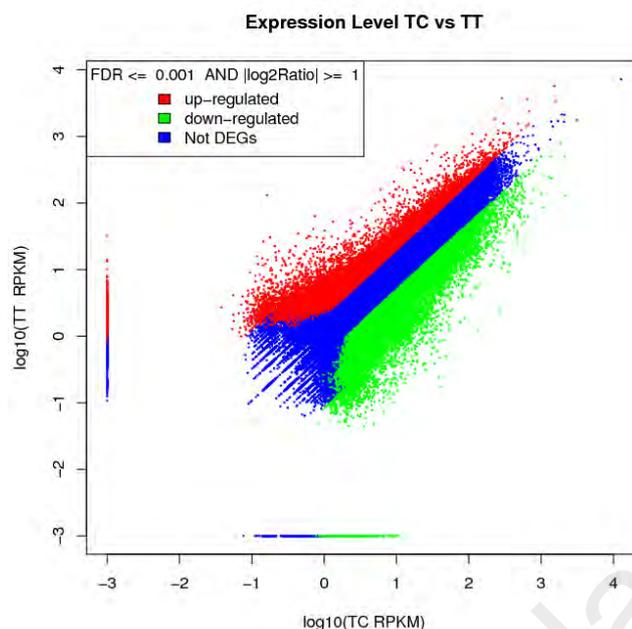


Figure 4.12: Expression level of unigenes that show significant differentially expressed in control (TC) and phenylalanine treated (TT) samples. Up-regulated and down-regulated genes were denoted by red and green spots respectively, while not differentially expressed genes were denoted as blue spots.

Table 4.4: Summary of DEGs expression levels in *B. rotunda* transcriptome data.

	Total	Up-regulated	Down-regulated
Total DEGs	100,869	47,451	53,418
Significant DEGs	29,023	14,644	14,379
-with annotation	16,018	6,104	9,914
-without annotation	13,005	8,540	4,465

4.4 Pathway analysis

Pathway-based analysis provides information and further understanding on how *B. rotunda* regulates biological functions and synthesizes secondary metabolites in response to phenylalanine at the molecular level. In total, there were 24,316 unigenes that mapped to the KEGG plant database using BlastX homology search. These unigenes were classified under 166 KEGG pathways in five main categories in KEGG which included Metabolism, Genetic Information Processing, Environmental Information Processing, Cellular Processes and Organismal Systems (Table 4.5). A single EC number may contain one or multiple unigenes. However, only 7,931 unigenes that were DEGs, significantly

up- or down-regulated, were mapped in the KEGG pathways. The total distribution of DEGs is represented in Table 4.5. Out of the 116 pathways, 16 pathways were significantly enriched with DEGs (Q value ≤ 0.05) (Table 4.6). The comparison between all unigenes and DEGs that mapped in DEG significantly enriched pathways also shown in Table 4.6.

Table 4.5: Distributions of all unigenes and DEGs in KEGG database classification.

There were five major categories which include Metabolism, Genetic Information Processing, Environmental Information Processing, Cellular Processes and Organismal Systems.

Category	Sub-Category	All genes with pathway annotation	DEGs genes with pathway annotation
Metabolism	Carbohydrate Metabolism	3,353	1,297
	Energy Metabolism	937	352
	Lipid Metabolism	1,445	514
	Nucleotide Metabolism	2,290	721
	Amino acid Metabolism	2,139	736
	Metabolism of Other Amino Acids	581	213
	Glycan Biosynthesis and Metabolism	278	86
	Metabolism of Cofactors and Vitamins	503	143
	Metabolism of Terpenoids and Polyketides	622	235
	Biosynthesis of Other Secondary Metabolites	1,092	385
	<i>Total</i>	<i>13,240</i>	<i>4,682</i>
Genetic Information Processing	Transcription	2,833	953
	Translation	755	184
	Folding, Sorting and Degradation	2,062	729
	Replication and Repair	923	275
	<i>Total</i>	<i>6,573</i>	<i>2,141</i>
Environmental Information Processing	Membrane Transport	224	116
	Signal Transduction	288	94
	<i>Total</i>	<i>512</i>	<i>210</i>
Cellular Processes	Transport and Catabolism	1,251	452
	<i>Total</i>	<i>1,251</i>	<i>452</i>
Organismal Systems	Immune System	140	36
	Environmental Adaptation	2,122	675
	<i>Total</i>	<i>2,262</i>	<i>711</i>
TOTAL UNIGENES		24,316	7,931

Table 4.6: Summary of unigene distributions in KEGG pathways that has significant differential expression of genes. Pathways in KEGG that has significant DEGs was determined by Q value ≤ 0.05 .

Category	Sub-Category	Pathway	DEGs with pathway annotation	Up-regulated unigenes	Down-regulated unigenes	P-value	Q-value
Metabolism	Carbohydrate Metabolism	Citrate cycle (TCA cycle)	102	14	88	4.07E-06	9.60E-05
		Galactose metabolism	73	18	55	1.41E-03	1.43E-02
		Glycolysis/ Gluconeogenesis	155	33	122	2.42E-03	2.20E-02
		Amino sugar and nucleotide sugar metabolism	151	56	95	4.39E-03	3.04E-02
		Pyruvate metabolism	134	18	116	3.67E-03	3.04E-02
		Glyoxylate and dicarboxylate metabolism	43	7	36	4.20E-03	3.04E-02
	Energy Metabolism	Nitrogen metabolism	75	25	50	2.41E-04	3.56E-03
	Amino Acid Metabolism	Phenylalanine metabolism	92	51	41	1.30E-03	1.43E-02
		Alanine, aspartate and glutamate metabolism	80	20	60	4.38E-03	3.04E-02
		Valine, leucine and isoleucine biosynthesis	45	7	38	6.42E-03	4.21E-02
	Metabolism of Terpenoid and Polyketides	Terpenoid backbone biosynthesis	75	4	71	4.82E-08	1.90E-06
Biosynthesis of Other Secondary Metabolites	Phenylpropanoid biosynthesis	163	68	95	1.46E-03	1.43E-02	
Genetic Information Processing	Translation	Aminoacyl-tRNA biosynthesis	96	10	86	1.45E-04	2.44E-03

Table 4.6, continued

	Folding, Sorting and Degradation	Protein processing in endoplasmic reticulum	241	53	188	1.30E-03	1.43E-02
Environmental Information Processing	Membrane Transport	ABC transporters	116	17	99	2.06E-09	1.22E-07
Cellular Processes	Transport and Catabolism	Endocytosis	208	44	164	4.85E-05	9.54E-04

4.4.1 Representation of genes regulation in phenylpropanoid pathway and flavonoid pathway.

In the transcriptome data, there were 411 unique unigenes were mapped to the phenylpropanoid pathway while 211 unigenes were mapped to the flavonoid pathway. In the phenylpropanoid pathway, 68 unigenes were up-regulated while 95 unigenes were down-regulated. Whereas in the flavonoid pathway, 11 unigenes were up-regulated and 42 unigenes were down-regulated. One unigene may map to more than one enzyme in the pathway (Table 4.7 and Table 4.8). Table 4.7 and Table 4.8 show the unigenes that might be involved in panduratin A biosynthesis and the number of up- and down-regulated unigenes with their respective gene regulation patterns (Figure 4.13). The gene regulation patterns in phenylpropanoid and flavonoid pathway, respectively were shown in Appendix D.6 & Appendix D.7. The most abundant unigenes that were mapped to the phenylpropanoid pathway was peroxidases (EC: 1.11.1.7) with a total of 90 unigenes. There were 40 unigenes that showed up-regulation while only 8 unigenes were down-regulated (Table 4.7).

Table 4.7: Unigenes potentially related to panduratin A biosynthesis in phenylpropanoid pathway.

One unigene may map to more than one enzyme in the pathway. The table show the total unigene that mapped to specific enzyme and gene regulation patterns either up-, down-regulated or both.

Enzyme Name	Abbreviations in flavonoid pathway (Figure 4.13)	EC Number	Enzyme Class	Total unigene	Up-regulated unigene	Down-regulated unigene
<i>Phenylpropanoid pathway</i>						
cinnamyl-alcohol dehydrogenase	-	1.1.1.195	Oxidoreductase	10	0	6
peroxidase	-	1.11.1.7	Oxidoreductase	90	40	8
ferulate-5-hydroxylase	-	1.14.-.-	Oxidoreductase	4	0	0
p-coumarate 3-hydroxylase	-	1.14.13.-	Oxidoreductase	5	0	2
trans-cinnamate 4-monooxygenase	C4H	1.14.13.11	Oxidoreductase	14	1	3
cinnamoyl-CoA reductase	-	1.2.1.44	Oxidoreductase	18	0	6
coniferyl-aldehyde dehydrogenase	-	1.2.1.68	Oxidoreductase	6	0	3
putative caffeoyl-CoA 3-O-methyltransferase	-	2.1.1.-	Transferase	2	0	0
caffeoyl-CoA O-methyltransferase	-	2.1.1.104	Transferase	2	1	0
caffeic acid 3-O-methyltransferase	-	2.1.1.68	Transferase	14	0	6
shikimate O-hydroxycinnamoyltransferase	-	2.3.1.133	Transferase	43	4	12
sinapoylglucose-choline O-sinapoyltransferase	-	2.3.1.91	Transferase	13	0	0
sinapoylglucose-malate O-sinapoyltransferase	-	2.3.1.92	Transferase	7	0	0
coniferyl-alcohol glucosyltransferase	-	2.4.1.111	Transferase	24	0	0
sinapate 1-glucosyltransferase	-	2.4.1.120	Transferase	29	1	5
beta-glucosidase	-	3.2.1.21	Hydrolase	76	9	17
phenylalanine ammonia-lyase	PAL	4.3.1.24	Lyase	14	2	5
phenylalanine/tyrosine ammonia-lyase	-	4.3.1.25	Lyase	1	0	0
4-coumarate--CoA ligase	4CL	6.2.1.12	Ligase	44	3	15

Table 4.8: Unigenes potentially related to panduratin A biosynthesis in flavonoid pathway.

One unigene may map to more than one enzyme in the pathway. The table show the total unigene that mapped to specific enzyme and gene regulation patterns either up-, down-regulated or both.

Enzyme Name	Abbreviations in flavonoid pathway (Figure 4.13)	EC Number	Enzyme Class	Total unigene	Up-regulated Unigene	Down-regulated unigene
<i>Flavonoid pathway</i>						
bifunctional dihydroflavonol 4-reductase/ flavanone 4-reductase	DFR	1.1.1.219/ 1.1.1.234	Oxidoreductase	15	0	4
leucoanthocyanidin dioxygenase/ anthocyanin synthase	ANS	1.14.11.19	Oxidoreductase	16	3	4
flavone synthase	FS1/FS2	1.14.11.22	Oxidoreductase	0	0	0
flavonol synthase	FLS	1.14.11.23	Oxidoreductase	29	5	6
naringenin 3-dioxygenase/ flavanone-3-hydroxylase	F3H	1.14.11.9	Oxidoreductase	20	2	5
p-coumarate 3-hydroxylase	-	1.14.13.-	Oxidoreductase	5	0	2
trans-cinnamate 4-monooxygenase	C4H	1.14.13.11	Oxidoreductase	14	1	3
flavonoid 3'-monooxygenase	-	1.14.13.21	Oxidoreductase	15	0	5
cytochrome P450, family 75, subfamily A (flavonoid 3',5'-hydroxylase)	-	1.14.13.88	Oxidoreductase	6	0	2
leucoanthocyanidin reductase	LAR	1.17.1.3	Oxidoreductase	4	0	1
anthocyanin reductase	ANR	1.3.1.77	Oxidoreductase	0	0	0
caffeoyl-CoA O-methyltransferase	-	2.1.1.104	Transferase	2	1	0
shikimate O-hydroxycinnamoyltransferase	-	2.3.1.133	Transferase	43	4	12
6'-deoxychalcone synthase	-	2.3.1.170	Transferase	15	0	0
chalcone synthase	CHS	2.3.1.74	Transferase	25	1	7
chalcone isomerase	CHI	5.5.1.6	Isomerase	2	0	0

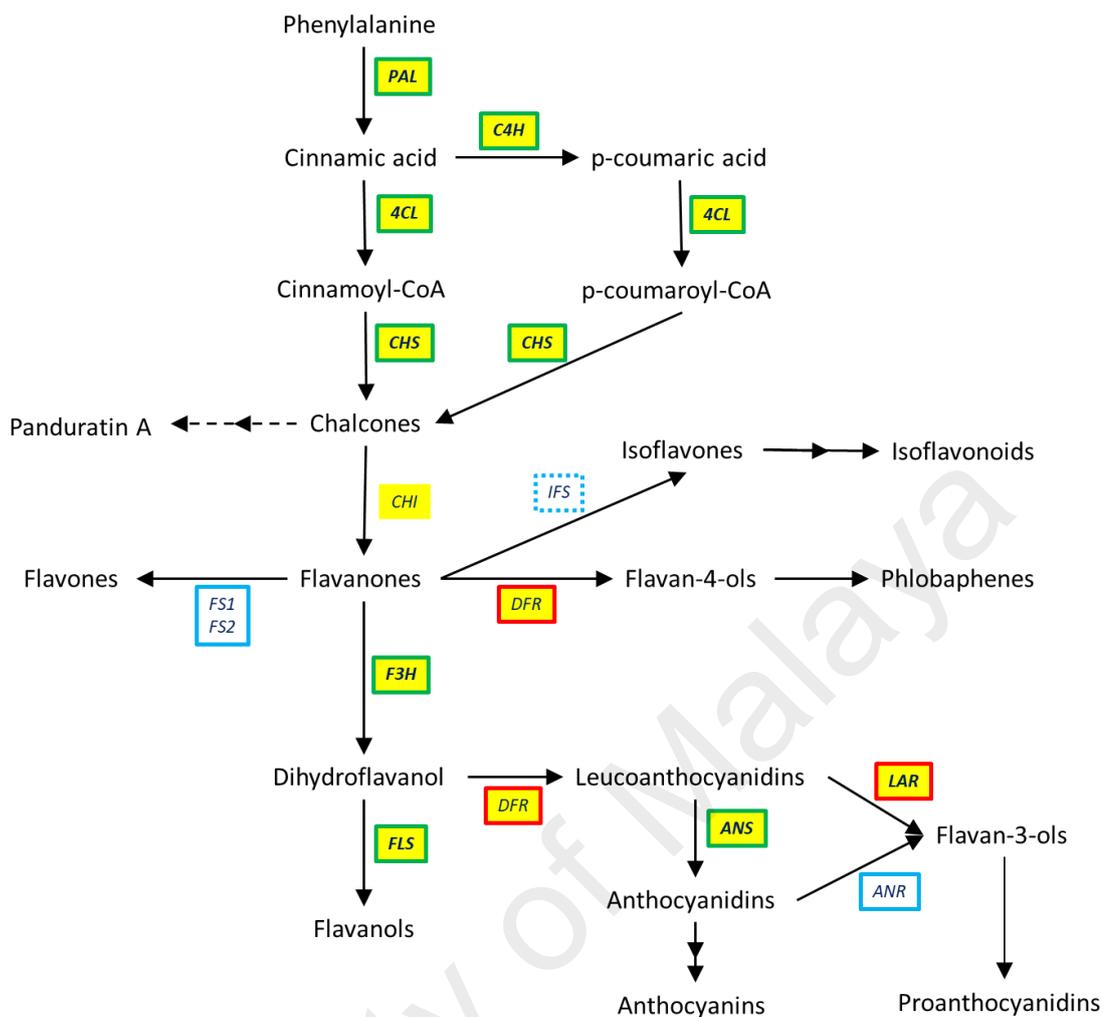


Figure 4.13: General flavonoid biosynthetic pathway (adapted from Bowsher et al., 2008).

The pathway showing synthesis of major flavonoid groups which include chalcones, flavanones, flavones, flavan-4-ols, flavan-3-ols, flavanols, isoflavones and anthocyanins. Panduratin A, a chalcone-derived compound was proposed to be derived from pinocembrin chalcone with dotted arrows. Abbreviations: PAL, phenylalanine ammonia lyase; C4H, cinnamate-4-hydroxylase; 4CL, 4-coumaroyl:coenzyme A ligase; CHS, chalcone synthase; CHI, chalcone isomerase; FS1/FS2, flavone synthase 1 and 2; IFS, isoflavone synthase; DFR, dihydroflavonol 4-reductase; F3H, flavanone-3-hydroxylase; FLS, flavonol synthase; ANS, anthocyanidin synthase; LAR, leucoanthocyanidin reductase and ANR, anthocyanidin reductase. Yellow boxes indicate that the unigenes mapped to the corresponding enzymes. The yellow boxes with green border consist of both up- and down-regulated unigene while yellow boxes with red border indicate down-regulated unigene. Boxes with blue border indicate that there were no unigene mapped to the corresponding enzymes, whereas box with dotted blue border indicate that the enzyme or the entire isoflavonoid pathway does not exist in *B. rotunda*.

An overview of the flavonoid biosynthetic pathway leading to the synthesis of major flavonoid groups and the proposed panduratin A biosynthesis is shown in Figure 4.13. Figure 4.13 shows the regulation patterns of unigenes that were mapped to the main enzymes in the flavonoid pathway. Only PAL, C4H, 4CL, CHS, CHI, F3H, FLS, DFR, ANS and LAR were mapped to the *B. rotunda* transcriptome unigenes. However, no gene regulation was detected for CHI (Table 4.8). In contrast, no unigene matched for FS1/FS2 and ANR (Table 4.8). Additionally, the isoflavone biosynthetic pathway map was not present in the *B. rotunda* system. In total, there were 14 unigenes mapped to PAL but only 2 of them showed up-regulation while 5 were down-regulated (Table 4.7). Out of 14 unigenes assigned as C4H, only 1 unigene was up-regulated and the other 3 unigenes were down-regulated. The most abundant unigene was assigned as 4CL with 44 unigenes. However, only 3 were up-regulated, while 15 others were down-regulated. A further 25 unigenes were assigned as CHS but only one unigene was up-regulated while 7 were down-regulated during production of chalcones in the flavonoid pathway. There was no gene regulation pattern observed during flavanone production from chalcones by CHI. A non-enzymatic reaction is suggested to be involved in this step. Subsequently, only two F3H and five FLS unigenes were up-regulated to form flavonols from flavanones. In contrast, there were four down-regulated DFR, three up-regulated ANS and one down-regulated LAR unigenes involved in anthocyanin and proanthocyanidin production.

Based on the gene regulation patterns in the flavonoid biosynthetic pathway, the highest up-regulated expression level was Unigene49558_All with 3.8 fold higher compared to control (Table 4.9). This unigene was mapped to F3H, FLS and ANS in flavonoid pathway. The second highest up-regulated gene was Unigene41852_All, which annotated as 4CL which showed a 2.2 fold change. The rest of the unigenes in the flavonoid pathway were 1 to 1.8 fold up-regulated. The most down-regulated unigenes in the flavonoid pathway was 4CL with seven out of fifteen unigenes showing expression

levels between 2 to 3.3 fold lower than the control (Table 4.9). The unigenes included Unigene88072_All, Unigene37844_All, Unigene68813_All, Unigene44539_All, Unigene51006_All, Unigene32973_All and Unigene85725_All. Three other unigenes that showed more than 2 fold down-regulation was C4H (Unigene17324_All) with 2.5 fold, CHS (Unigene35484_All) with 2.2 fold and DFR (Unigene100192_All) with 2.8 fold.

University of Malaya

Table 4.9: Gene regulation patterns in flavonoid pathway.

The genes include phenylalanine ammonia lyase (PAL), cinnamate-4-hydroxylase (C4H), 4-coumarate-CoA ligase (4CL), chalcone synthase (CHS), flavanone-3-hydroxylase (F3H), flavonol synthase (FLS), dihydroflavonol-4-reductase (DFR), anthocyanin synthase (ANS) and leucoanthocyanidin reductase (LAR).

Enzyme	EC Number	Up-regulated		Down-regulated	
		Unigene ID	Fold Change	Unigene ID	Expression Level Fold
PAL	4.3.1.24	Unigene10327_ All	1.1	Unigene83336_ All	-1.9
		Unigene89418_ All	1	Unigene56631_ All	-1.9
				Unigene64872_ All	-1.6
				Unigene619_ All	-1.3
				Unigene9322_ All	-1
C4H	1.14.13.11	Unigene67845_ All	1.4	Unigene17324_ All	-2.5
				Unigene11543_ All	-1.7
				Unigene93243_ All	-1.1
4CL	6.2.1.12	Unigene41852_ All	2.2	Unigene88072_ All	-3.3
		Unigene36813_ All	1.2	Unigene37844_ All	-3
		Unigene3277_ All	1.1	Unigene68813_ All	-2.8
				Unigene44539_ All	-2.3
				Unigene51006_ All	-2.2
				Unigene32973_ All	-2.1
				Unigene85725_ All	-2
				Unigene28297_ All	-1.9
				Unigene57823_ All	-1.8
				Unigene520_ All	-1.7
				Unigene19555_ All	-1.7
				Unigene20812_ All	-1.5
				Unigene10021_ All	-1.4
		Unigene6803_ All	-1.4		
		Unigene20574_ All	-1.4		

Table 4.9, continued

CHS	2.3.1.74	Unigene1735_A ll	1.5	Unigene35484_ All	-2.2
				Unigene31906_ All	-1.8
				Unigene37184_ All	-1.4
				Unigene33635_ All	-1.3
				Unigene63145_ All	-1.3
				Unigene55042_ All	-1.1
				Unigene29406_ All	-1.1
F3H	1.14.11.9	Unigene49558_ All	3.8	Unigene100816 All	-1.6
		Unigene5973_A ll	1.4	Unigene4657_A ll	-1.6
				Unigene22973_ All	-1.4
				Unigene4884_A ll	-1.3
				Unigene23932_ All	-1.1
FLS	1.14.11.23	Unigene49558_ All	3.8	Unigene100816 All	-1.6
		Unigene89505_ All	1.8	Unigene4657_A ll	-1.6
		Unigene56837_ All	1.8	Unigene22973_ All	-1.4
		Unigene26406_ All	1.6	Unigene4884_A ll	-1.3
		Unigene5973_A ll	1.4	Unigene23932_ All	-1.1
				Unigene33774_ All	-1
DFR	1.1.1.219			Unigene100192 All	-2.8
				Unigene40110_ All	-1.9
				Unigene84008_ All	-1.7
				Unigene49734_ All	-1.3
ANS	1.14.11.19	Unigene49558_ All	3.8	nigene100816_ All	-1.6
		Unigene89505_ All	1.8	Unigene4657_A ll	-1.6
		Unigene56837_ All	1.8	Unigene4884_A ll	-1.3
				Unigene30270_ All	-1.1
LAR	1.17.1.3			Unigene73982_ All	-1

4.4.2 Representation of genes regulation in proposed pathway.

In this thesis, as presented in Figure 4.14, it is proposed that flavonoid O-methyltransferase (ChOMT1, ChOMT2, 5-OMT and 7-OMT) and prenyltransferase (PT1 and PT2) are two important enzymes for producing the bioactive compounds in *B. rotunda*.

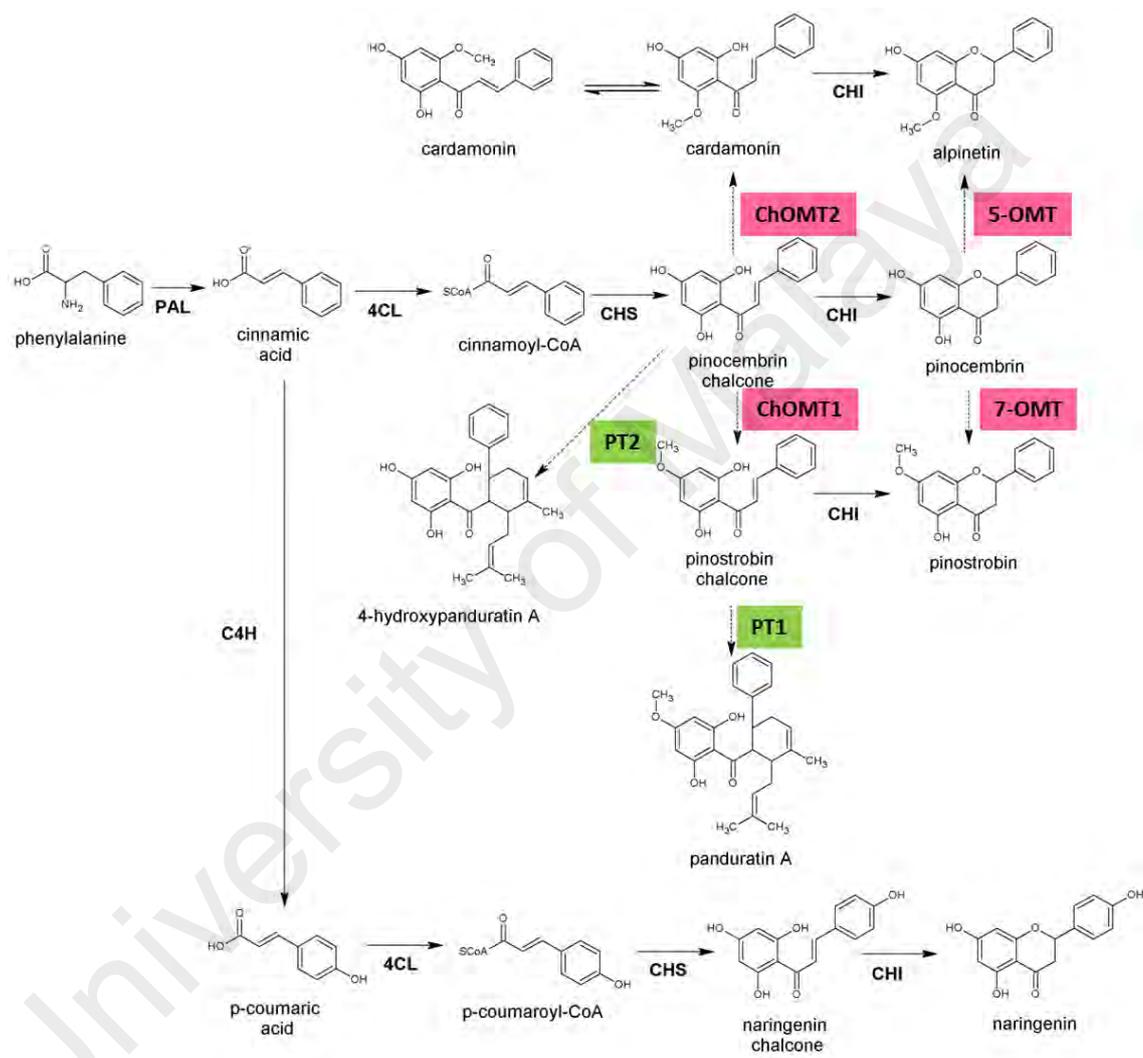


Figure 4.14: Proposed pathway leading to cardamonin, alpinetin, pinocembrin, pinostrobin, 4-hydroxypanduratin A and panduratin A production (adapted from Bowsher et al., 2008).

Flavonoid O-methyltransferase was highlighted with pink box whereas prenyltransferase was highlighted with green box. Abbreviations: PAL, phenylalanine ammonia lyase; C4H, cinnamate-4-hydroxylase; 4CL, 4-coumaroyl:coenzyme A ligase; CHS, chalcone synthase; CHI, chalcone isomerase; PT1, prenyltransferase1; PT2, prenyltransferase2 5-OMT, pinocembrin 5-O-methyltransferase, 7-OMT, pinocembrin 7-O-methyltransferase, ChOMT1, pinocembrin chalcone 4'-O-methyltransferase and ChOMT2, pinocembrin chalcone 2'-O-methyltransferase. Double arrows indicate reversible reactions. Dotted arrows indicate the proposed pathway.

4.4.2.1 Flavonoid O-methyltransferase analysis

From the *B. rotunda* transcriptome data, there were 12 unigenes annotated by NR database as flavonoid O-methyltransferase. After eliminating 4 unigenes that have gaps in the nucleotide sequence, 8 unigenes were further analyzed for the domain region. There were 2 unigenes that contain both O-methyltransferase and dimerization domains, Unigene891_All and Unigene25107_All (Appendix I.2). However, only Unigene891_All has complete ORF with 362 amino acid in length (Appendix H.2). The expression fold of Unigene891_All in the phenylalanine treated sample was slightly up-regulated by 0.1 fold higher as compared to the control sample, whereas Unigene25107_All was slightly down-regulated by 0.9 fold lower than control sample. However, the fold changes for both unigenes were not significant.

Subsequently, the protein Blast results showed that Unigene891_All has the highest score to tricetin 3',4',5'-O-trimethyltransferase-like protein from *Musa acuminata* subsp. *malaccensis* with 81% and 99% coverage. In contrast, Unigene25107_All showed only 29 – 43% identity to other plant flavonoid O-methyltransferases (Table 4.10). It has the highest score to trans-resveratrol di-O-methyltransferase- like from *Prunus mume* with 46% identity and 96% coverage.

Multiple sequence alignment of Unigene891_All and Unigene 21507_All with other flavonoid prenyltransferase proteins was done using Clustal Omega (Figure 4.15). From the alignment, it showed that the dimerization domain was located at the N-terminal region, whereas longer O-methyltransferase domain located at C-terminal region. Histidine residue was shown to be conserved across the plant flavonoid O-methyltransferase proteins. Histidine residue was located at position 268 amino acid within the O-methyltransferase domain region. It serves as a base for the O-methylation reaction. Subsequently, a phylogenetic tree was constructed to predict Unigene891_All functions

(Figure 4.16). From the results, the Unigene891_All were clustered under the same clades with flavone O-methyltransferase-like and tricetin 3',4',5'-O-trimethyltransferase-like proteins from *Musa acuminata* subsp. *malaccensis*. Whereas, Unigene25107_All shared the same clade with flavone 7-O-methyltransferase from *Hordeum vulgare* subsp. *vulgare*.

University of Malaya

Table 4.10: Summary of homology search using Protein blast of two unigenes annotated as flavonoid O-methyltransferase.

Unigene ID	Nucleotide length (bp)	Amino acid length	Expression Level Fold	Description	Max score	Total score	Query cover	E value	Identity	Accession
Unigene891_All	1320	362	Non-significantly regulated	PREDICTED: tricetin 3',4',5'-O-trimethyltransferase-like [<i>Musa acuminata</i> subsp. <i>malaccensis</i>]	424	424	99%	6.00E-144	81%	XP_009418616.1
				PREDICTED: flavone O-methyltransferase 1-like [<i>Musa acuminata</i> subsp. <i>malaccensis</i>]	420	420	99%	3.00E-142	83%	XP_009392797.1
				PREDICTED: flavone O-methyltransferase 1-like [<i>Musa acuminata</i> subsp. <i>malaccensis</i>]	409	409	99%	5.00E-137	76%	XP_009395003.1
				PREDICTED: tricetin 3',4',5'-O-trimethyltransferase [<i>Phoenix dactylifera</i>]	404	404	95%	1.00E-135	78%	XP_008788729.1
				PREDICTED: caffeic acid 3-O-methyltransferase-like [<i>Elaeis guineensis</i>]	401	401	95%	9.00E-135	78%	XP_010934914.1
				PREDICTED: tricetin 3',4',5'-O-trimethyltransferase-like [<i>Elaeis guineensis</i>]	396	396	95%	8.00E-133	76%	XP_010934913.1
				PREDICTED: tricetin 3',4',5'-O-trimethyltransferase-like [<i>Phoenix dactylifera</i>]	395	395	95%	2.00E-132	74%	XP_008775842.1

Table 4.10, continued

Unigene25107_All	817	265	Non-significantly regulated	PREDICTED: trans-resveratrol di-O-methyltransferase- like [<i>Prunus mume</i>]	284	284	96%	1.00E-90	46%	XP_008229108.1
				o-methyltransferase, putative [<i>Ricinus communis</i>]	283	283	86%	6.00E-90	29%	XP_002525818.1
				PREDICTED: caffeic acid 3-O-methyltransferase 1 [<i>Nelumbo nucifera</i>]	283	283	86%	6.00E-90	31%	XP_010241050.1
				caffeic acid/5-hydroxyferulic acid O-methyltransferase [<i>Arabidopsis thaliana</i>]	282	282	86%	8.00E-90	31%	NP_200227.1
				O-methyltransferase 1 [<i>Arabidopsis lyrata</i> subsp. <i>lyrata</i>]	282	282	86%	1.00E-89	31%	XP_002864309.1
				PREDICTED: caffeic acid 3-O-methyltransferase [<i>Prunus mume</i>]	282	282	86%	1.00E-89	30%	XP_008234634.1
				O-methyltransferase [<i>Populus trichocarpa</i> x <i>Populus deltoides</i>]	281	281	97%	1.00E-89	29%	AAF60951.1
				RecName: Full=Quercetin 3-O-methyltransferase 2; AltName: Full=Flavonol 3-O-methyltransferase 2 [<i>Chrysosplenium americanum</i>]	280	280	86%	2.00E-89	29%	Q42653.1

Table 4.10, continued

				RecName: Full=Quercetin 3-O-methyltransferase 1; AltName: Full=Flavonol 3-O-methyltransferase 1 [<i>Chrysosplenium americanum</i>]	280	280	86%	3.00E-89	29%	P59049.1
				Flavonoid o-methyltransferase related [<i>Theobroma cacao</i>]	268	268	96%	3.00E-84	43%	XP_007043718.1
				PREDICTED: flavone 3'-O-methyltransferase 1-like [<i>Camelina sativa</i>]	260	260	86%	3.00E-81	29%	XP_010433307.1



Figure 4.15: Comparison of the deduced amino acid sequence of Unigene891_All and Unigene25107_All with other flavonoid O-methyltransferases from other plants such as *Oryza sativa* (OsNOMT), *Medicago sativa* (ChOMT), *Hordeum vulgare* subsp. *vulgare* (Fl-OMT) and *Musa acuminata* subsp. *malaccensis* (MaFl-OMT1-like_1, MaFl-OMT1-like_2 and MaFl-OMT1-like_3).

Dimerization and O-methyltransferase domain regions were denoted by green and orange arrows, respectively. Blue triangle indicates Histidine residue. The accession numbers of the flavonoid O-methyltransferases from other plants to construct the multiple sequence alignment are shown in the Appendix J.

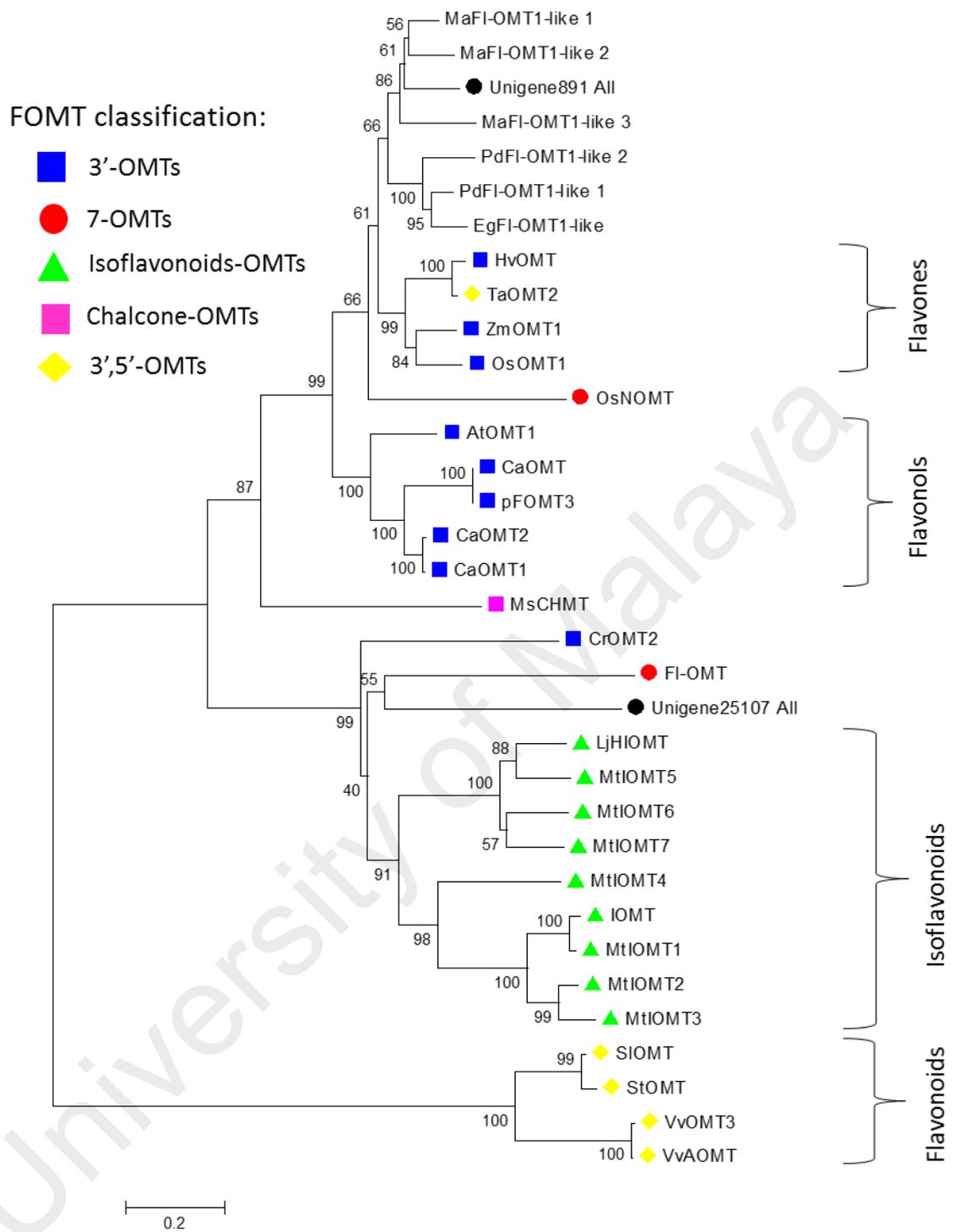


Figure 4.16: Phylogenetic relationship of Unigene891_All and Unigene25107_All from *B. rotunda* with flavonoid O-methyltransferase (FOMT) proteins from other plants. Topology tree was performed using the neighbour-joining method by the MEGA 5.10 version. The numbers next to the nodes are bootstrap values from 1000 replicates. Flavonoid O-methyltransferase protein from *B. rotunda*, Unigene891_All were indicated as black circle. Plant FOMT classification was indicated by red circle, green triangle, blue rectangle and yellow diamond. The accession numbers of the flavonoid O-methyltransferase from other plants to construct the phylogenetic tree are shown in the Appendix J.

4.4.2.2 Prenyltransferase analysis

From the *B. rotunda* transcriptome data, there were only Unigene31983_All annotated by the NR database as flavonoid prenyltransferase. The expression fold of Unigene31983_All in the phenylalanine treated sample was slightly up-regulated by 0.6 folds higher as compared to the control sample. However, the fold changes were not significantly regulated. The predicted ORF length of Unigene31983_All was 391 amino acids. Open reading frame prediction for Unigene31983_All is shown in Appendix H.3. The unigene contained the PT-UbiA superfamily conserved domain region (Appendix I.3). Subsequently, the protein Blast result showed that Unigene31983_All has the highest score to homogentisate phytyltransferase 1 protein from *Musa acuminata* subsp. *malaccensis* with 75% and 100% coverage (Table 4.11). The Unigene31983_All ORF also showed 45-66% identity to other flavonoid prenyltransferase proteins (Table 4.11).

Multiple sequence alignment of Unigene31983_All (391 amino acid length) with other flavonoid prenyltransferase proteins was done using Clustal Omega (Figure 4.17). From the alignment, it showed that NQxxDxxxD and KDxxDx[E/D]GD motifs were conserved across flavonoid prenyltransferase proteins. However, the Unigene31983_All do not have the full ORF, lacking a few amino acids at the 3' terminal region. A phylogenetic tree was constructed using partial Unigene31983_All ORF sequence with other plant aromatic and homogentisate prenyltransferases (Figure 4.18). Aromatic prenyltransferases include flavonoid, coumarin and phloroglucinol prenyltransferases. Surprisingly, Unigene31983_All was not clustered under flavonoid prenyltransferases but appeared in the same clade with tocopherol producing homogentisate phytyltransferases.

Table 4.11: Summary of homology search using Protein blast of Unigene31983_All annotated as prenyltransferase.

Nucleotide length (bp)	Amino acid length	Expression Level Fold	Description	Max score	Total score	Query cover	E value	Identity	Accession
1281	391	Non-significantly regulated	PREDICTED: probable homogentisate phytyltransferase 1, chloroplastic [<i>Musa acuminata</i> subsp. <i>malaccensis</i>]	318	318	100%	2.00E-101	75%	XP_009396234.1
			Chlorophyll synthase [<i>Musa balbisiana</i>]	308	308	100%	2.00E-97	75%	CBW30171.1
			Chlorophyll synthase [<i>Musa balbisiana</i>]	306	306	100%	1.00E-96	75%	CBW30209.1
			PREDICTED: probable homogentisate phytyltransferase 1, chloroplastic isoform X1 [<i>Musa acuminata</i> subsp. <i>malaccensis</i>]	300	300	100%	3.00E-94	74%	XP_009394579.1
			homogentisate phytyltransferase [<i>Elaeis oleifera</i>]	288	288	100%	7.00E-90	71%	AHL26475.1
			probable homogentisate phytyltransferase 1, chloroplastic [<i>Elaeis guineensis</i>]	288	288	100%	8.00E-90	71%	NP_001291355.1
			PREDICTED: probable homogentisate phytyltransferase 1, chloroplastic isoform X2 [<i>Musa acuminata</i> subsp. <i>malaccensis</i>]	284	284	93%	2.00E-88	75%	XP_009394580.1
			PREDICTED: probable homogentisate phytyltransferase 1, chloroplastic isoform X2 [<i>Elaeis guineensis</i>]	284	284	99%	2.00E-88	70%	XP_010928674.1

Table 4.11, continued

Nucleotide length (bp)	Amino acid length	Expression Level Fold	Description	Max score	Total score	Query cover	E value	Identity	Accession
			aromatic prenyltransferase [<i>Epimedium acuminatum</i>]	246	246	100%	1.00E-73	66%	AEZ53107.1
			isoliquiritigenin dimethylallyltransferase [<i>Sophora flavescens</i>]	229	229	81%	6.00E-67	56%	BAK52290.1
			flavonoid prenyltransferase [<i>Sophora flavescens</i>]	222	222	79%	4.00E-64	54%	BAG12674.1
			flavone prenyltransferase [<i>Glycyrrhiza uralensis</i>]	220	220	80%	2.00E-63	56%	AIT11912.1
			genistein 6-dimethylallyltransferase [<i>Sophora flavescens</i>]	217	217	80%	2.00E-62	53%	BAK52291.1
			RecName: Full=Naringenin 8-dimethylallyltransferase 1, chloroplastic; Short=SfN8DT-1; Flags: Precursor [<i>Sophora flavescens</i>]	217	217	85%	4.00E-62	51%	B1B3P3.1
			flavonoid prenyltransferase [<i>Sophora flavescens</i>]	215	215	80%	2.00E-61	52%	AHA36633.1
			8-dimethylallyltransferase [<i>Sophora flavescens</i>]	215	215	81%	2.00E-61	53%	BAK52289.1
			glycinol 4-dimethylallyltransferase [<i>Glycine max</i>]	208	208	80%	5.00E-59	50%	NP_001235990.1
			genistein 3'- [<i>Lupinus albus</i>]	202	202	87%	1.00E-56	45%	AER35706.1

	KDxxDx[E/D]GD	
Unigene31983_A11	MSFFSIVIALFKDIPDIEGDIIFGIRSFVSLRGQKRVFNICVYLLEMAYGVAMAVGITSS	342
LaPT1	LSIFCIVISMKDIPDMEGDIEKFGIKSFALSLGQKRVFSICISLLQMSYGVGILVGATSP	350
Sfn8DT-2	VSIYAIVIALFKDIPDMEGDIEKFGIQSLSLRLGPKRVFNICVSLLEMAYGVITILVGATSP	351
SffPT	VSIYAIVIALFKDIPDMEGDIEKFGIQSLSLRLGPKRVFNICVSLLEMAYGVITILVGATSP	351
Sfn8DT-3	VSIYAIVIALFKDIPDMEGDIEKFGIQSLSLRLGPKRVFNICVSLLEMAYGVITILVGATSP	354
Sfn8DT-1	VSIYAIVIALFKDIPDMEGDIEKFGIQSLSLRLGPKRVFNICVSLLEMAYGVITILVGATSP	354
SfiLDT	LSLFFVIALFKDIPDIEGDIKFGVQSLAVRLGQKRVFNICISLLEMAYGVITILVGATSP	335
SfG6DT	LSLFFVIALFKDIPDIEGDIKFGIRSLSAQLGQKRVFNICISLLQMSYGVITILVGATSP	351
GmG4DT	MTFYSLGLALFKDIPDVEGDIKHGIDSFVRLGQKRAFNICVFFEMAFGVGILAGASCS	353
GuA6DT	MSFFSLVIALFKDIPDIEGDIKAFGVQSFASLGGKRVFNICVSLLETAYGVALLMGATSS	356
	::::: : :: : ***** * . : * : : * * * : : : : : * : .	
Unigene31983_A11	CNWSKLVITVLGHAVLALILWNRKSVNLMGKPAITSFYMFIVQ-LFYAEY-----	391
LaPT1	YLWSKIITVVGHAIALALVLQYRAKSVDPKSKDSVQSFYMFIVKLFIAECLLLPLFRS	408
Sfn8DT-2	ILWSKIITVLGHAILASVLWYHAKSTDLTNSVVLQSFYMFIVK-LHTAEYCLIPFR-	407
SffPT	ILWSKIITVLGHAVLASVLWYHAKSVDLTNSVVLQSFYMFIVK-LHTADYFLIPFR-	407
Sfn8DT-3	ILWSKIITVLGHAVLASVLWYHAKSVDLTNSVVLQSFYMFIVK-LHTAEYFLIPFR-	410
Sfn8DT-1	ILWSKIITVLGHAVLASVLWYHAKSVDLTNSVVLHVSFYMFIVK-LHTAEYFLIPFR-	410
SfiLDT	FLWSKISTGLGHAVLASIVWNRKSVDLKNDKSYKSFYMFIVK-LICAEYCLIPFR-	391
SfG6DT	FLWSKISMVLGHAILASILGYQVKSVDLKNNDALQSFYLFIVK-LLTVEYCLIPFR-	407
GmG4DT	HFWTKIFITGMGNVAVLASILWYQAKSVLSDKASTGFSFYMFIVK-LLYAGFFMALIR-	409
GuA6DT	CLWSKIITVLGHAILALVLFYRAKSVNLMGKPAITSFYMFIVK-LLYAEYFLVPLVR-	412
	* : * : : * . * * * : : . * * : : : * : : * : * : * : *	

Figure 4.17, continued

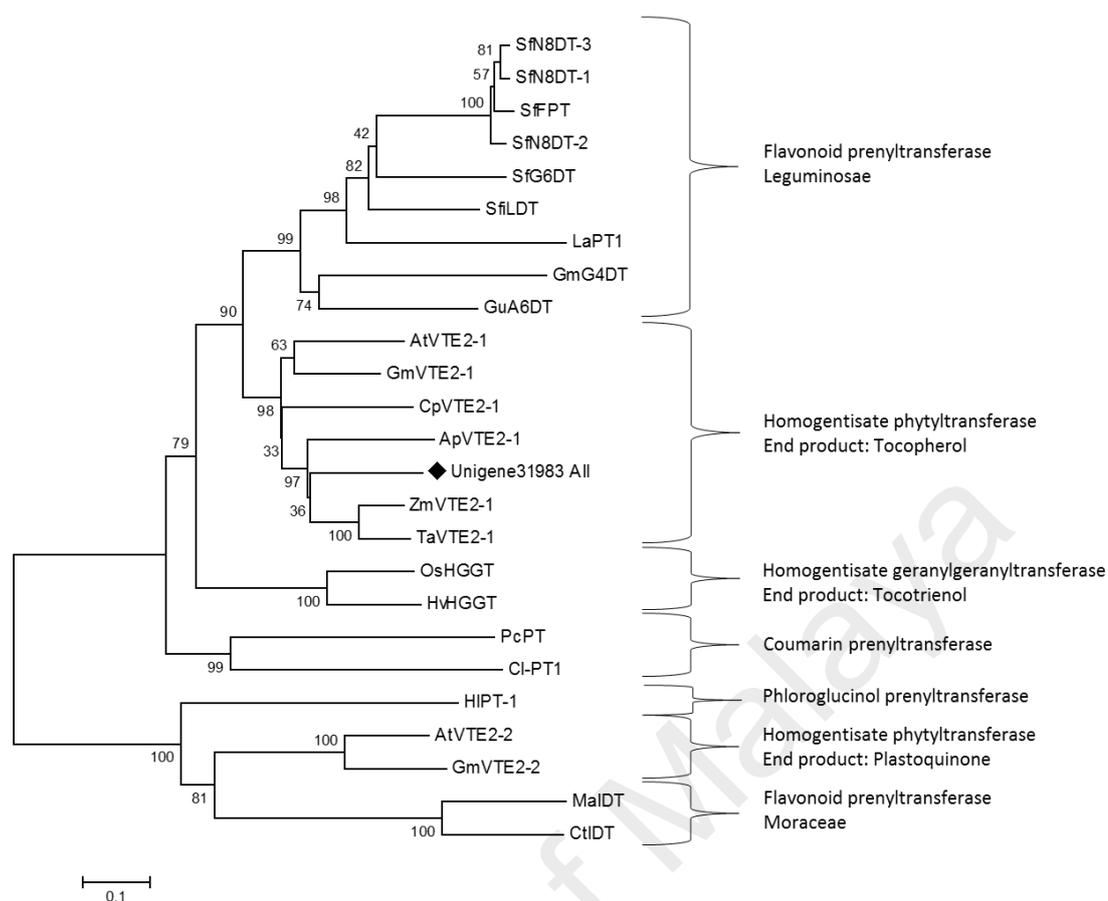


Figure 4.18: Phylogenetic tree relationship between Unigene31983_All with plant aromatic and homogentisate prenyltransferases.

Topology tree was performed using the neighbour-joining method by the MEGA 5.20 version. The numbers next to the nodes are bootstrap values from 1000 replicates. Flavonoid prenyltransferase protein from *B. rotunda*, Unigene31983_All were indicated as diamond. The accession numbers of the aromatic and homogentisate prenyltransferases from other plants to construct the phylogenetic tree are shown in the Appendix J.

4.5 Experimental validation

A single, tight peak on a dissociation curve on the qPCR indicated a very specific amplification for all 9 randomly selected unigenes (Figure A.14 in Appendix M.2). The generated data from the relative quantification was analysed with Quant Studio software (Table A.30 and Table A.31 in Appendix M.1). The qPCR results of 9 randomly selected unigenes showed general agreement with their transcript abundance changes as determined by RNA-Seq, suggesting the reliability of the transcriptome profiling data (Figure 4.19). For the unigenes tested only two showed some discrepancies although both

were similarly up-regulated i.e Unigene58054_All showed higher expression level (about 83.361 fold changes) while Unigene1735_All had moderate expression level (4.776 fold changes) in qPCR as compared to the RNA-Seq results, respectively.

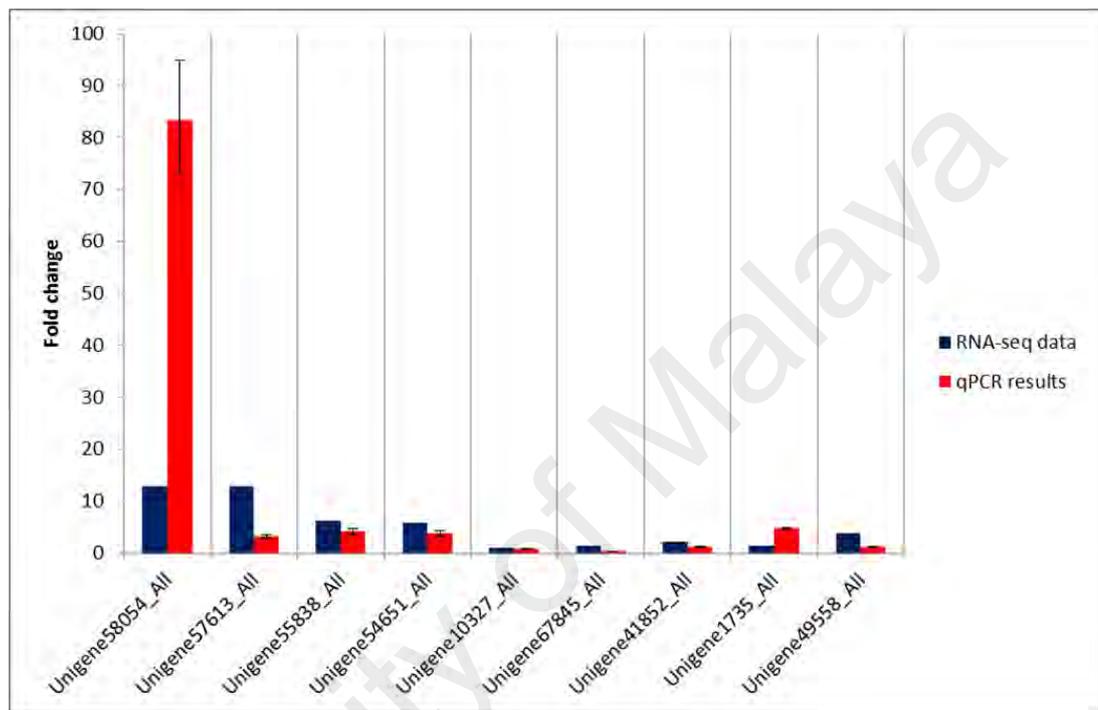


Figure 4.19: Expression pattern validation of selected unigenes by qPCR. Changes in transcript levels of 9 selected unigenes. Y-axis shows –fold changes in transcript abundance of unigenes compared to their respective control. Blue bar indicates transcript abundance changes calculated by the RPKM method. Red bar with associated standard error bar represents relative expression level determined by qPCR using $2^{-\Delta\Delta CT}$ method. Results represent mean standard deviations (\pm SD) of four experimental replicates.

4.6 Analysis of transcription factors and transcription regulators

Transcription factors (TFs) and transcription regulators (TRs) play essential roles in regulating differentially expressed genes in both a spatial and temporal manner. In total, 139 transcription factors that were found in *B. rotunda* can be further classified under 35 transcription factor families (Table 4.12). Based on the iTAK rice transcription factor database, 21 rice TFs were not found in *B. rotunda*. The most abundant TFs found in *B. rotunda* was C3H (17), followed by MYB (16), NAC (13), WRKY (9), bZIP (8) and AP2-

EREBP (7). In response to phenylalanine treatment, eight TFs were up-regulated while twenty six TFs were down-regulated. Up-regulated TFs included MYB, NAC, WRKY, bZIP, AP2-EREBP, G2-like, GRAS and C2C2-CO-like transcription factors.

Subsequently, 46 transcription regulators are classified under 15 families found in *B. rotunda* (Table 4.13). The most abundant TRs found in *B. rotunda* was orphan (9), followed by AUX/IAA (8) and SET (5). In total, 4 TRs were up-regulated and 9 were down-regulated in response to phenylalanine.

Table 4.12: Transcription factors (TFs) identified in *B. rotunda* by mapping to transcription factors in rice database using iTAK software.

	Transcription factors (<i>B. rotunda</i>)	Number of genes	Up-regulated TFs	Down-regulated TFs	Comparative transcription factors (<i>Oryza sativa</i>)	Number of genes
1	C3H	17	0	6	C3H	70
2	MYB	16	1	4	MYB	184
3	NAC	13	1	5	NAC	143
4	WRKY	9	1	2	WRKY	98
5	bZIP	8	1	1	bZIP	91
6	AP2-EREBP	7	1	2	AP2-EREBP	164
7	C2H2	5	0	0	C2H2	123
8	G2-like	5	1	1	G2-like	45
9	HB	5	0	1	HB	94
10	TUB	5	0	0	TUB	15
11	bHLH	4	0	0	bHLH	135
12	GRAS	4	1	0	GRAS	60
13	Tify	4	0	1	Tify	17
14	FAR1	3	0	0	FAR1	8
15	LOB	3	0	0	LOB	36
16	MADS	3	0	0	MADS	69
17	PBF-2-like	3	0	0	PBF-2-like	2
18	ABI3VP1	2	0	0	ABI3VP1	55
19	Alfin-like	2	0	0	Alfin-like	9
20	BBR/BPC	2	0	0	BBR/BPC	4
21	BSD	2	0	0	BSD	10
22	C2C2-GATA	2	0	0	C2C2-GATA	25
23	SRS	2	0	0	SRS	5
24	Trihelix	2	0	0	Trihelix	26
25	ARF	1	0	1	ARF	27
26	BES1	1	0	0	BES1	6
27	C2C2-CO-like	1	1	0	C2C2-CO-like	8

Table 4.12, continued

28	C2C2-YABBY	1	0	1	C2C2-YABBY	8
29	CCAAT	1	0	0	CCAAT	51
30	EIL	1	0	0	EIL	9
31	GeBP	1	0	0	GeBP	13
32	GRF	1	0	0	GRF	12
33	mTERF	1	0	1	mTERF	34
34	OFP	1	0	0	OFP	31
35	zf-HD	1	0	0	zf-HD	14
36	ARR-B	0	0	0	ARR-B	9
37	C2C2-Dof	0	0	0	C2C2-Dof	30
38	CAMTA	0	0	0	CAMTA	6
39	CPP	0	0	0	CPP	11
40	CSD	0	0	0	CSD	2
41	DBP	0	0	0	DBP	3
42	E2F-DP	0	0	0	E2F-DP	8
43	FHA	0	0	0	FHA	18
44	HRT	0	0	0	HRT	1
45	HSF	0	0	0	HSF	25
46	LFY	0	0	0	LFY	2
47	LIM	0	0	0	LIM	6
48	PLATZ	0	0	0	PLATZ	15
49	RWP-RK	0	0	0	RWP-RK	13
50	S1Fa-like	0	0	0	S1Fa-like	2
51	SBP	0	0	0	SBP	19
52	Sigma70-like	0	0	0	Sigma70-like	6
53	TAZ	0	0	0	TAZ	6
54	TCP	0	0	0	TCP	21
55	ULT	0	0	0	ULT	2
56	VOZ	0	0	0	VOZ	2
	Total	139	8	26	Total	1908

Table 4.13: Transcription regulators identified in *B. rotunda* using transcription regulators in rice database using iTAK software.

	Transcriptional regulator (<i>B. rotunda</i>)	Number of gene	Up-regulated	Down-regulated	Transcription regulator (<i>Oryza sativa</i>)	Number of gene
1	Orphans	9	2	3	Orphans	79
2	AUX/IAA	8	1	1	AUX/IAA	32
3	SET	5	1	0	SET	41
4	SNF2	4	0	2	SNF2	39
5	TRAF	4	0	1	TRAF	59
6	RB	3	0	0	RB	2
7	SWI/SNF-BAF60b	3	0	0	SWI/SNF-BAF60b	11
8	MED6	2	0	0	MED6	1
9	PHD	2	0	1	PHD	39
10	GNAT	1	0	0	GNAT	35
11	HMG	1	0	0	HMG	9
12	Jumonji	1	0	0	Jumonji	14
13	Rcd1-like	1	0	0	Rcd1-like	5

Table 4.13, continued

14	SOH1	1	0	0	SOH1	2
15	SWI/SNF-SWI3	1	0	1	SWI/SNF-SWI3	4
16	ARID	0	0	0	ARID	6
17	Coactivator p15	0	0	0	Coactivator p15	3
18	DDT	0	0	0	DDT	7
19	IWS1	0	0	0	IWS1	17
20	LUG	0	0	0	LUG	6
21	MBF1	0	0	0	MBF1	2
22	MED7	0	0	0	MED7	1
23	Pseudo ARR-B	0	0	0	Pseudo ARR-B	5
	Total	46	4	9	Total	419

4.7 R2R3 MYB transcription factor analysis

From *B. rotunda* transcriptome data, there were 145 unigenes annotated by NR database as R2R3 MYB transcription factor. As summarized in Table 4.14, only 21 unigenes consist of R2 and R3 repeat MYB domain. However, through multiple sequence alignment analysis, only 13 unigenes have the complete ORF (Figure 4.20). From the alignment, conserved R2 and R3 repeat MYB domains were determined. Further protein motif identification were done using MEME tool as presented in Figure 4.21. It can be concluded that R2 and R3 repeats were highly conserved in *B. rotunda* R2R3 MYB transcription factors. Phylogenetic tree was constructed to predict the MYB functions using 13 complete ORF MYB unigene sequences (Figure 4.22). From the results, the unigenes were classified into 7 clades which include abiotic/biotic stress, transcriptional repressor, anthocyanin biosynthesis, lignin biosynthesis, plant development and cell wall thickening clade (Table 4.15).

However, not all unigenes exhibit the subgroup-specific motifs. Six out of thirteen unigenes such as Unigene4579_All; Unigene51701_All, Unigene4347_All, Unigene70854_All, Unigene39151_All and Unigene4817_All possess the subgroup-specific motifs (Table 4.15). Unigene4579_All was classified under subgroup 1 which

are responsible for abiotic or biotic stress in the plant by harbouring one of two subgroup-specific motifs, YASSxxNI. Subsequently, four unigenes (Unigene51701_All, Unigene4347_All, Unigene20944_All, Unigene70854_All and Unigene83922_All) were categorized under subgroup 4 transcriptional repressor R2R3 MYBs. While subgroup 21 R2R3 MYBs consist of Unigene39151_All, Unigene29311_All and Unigene4817_All. Although none subgroup-specific motifs were detected on C-terminal region of Unigene19149_All, Unigene85656_All, Unigene69184_All, through phylogenetic tree analysis, the unigenes were clustered into anthocyanin regulating subgroup 6, lignin regulating subgroup 13 and plant developmental regulating subgroup 14 R2R3 MYBs, respectively.

Table 4.14: Summary of manually analysed R2R3 MYB transcription factor in *B. rotunda* transcriptome data, which included the total number of R2R3 MYB transcription factor annotation, unigene that had complete nucleotide sequence without gaps and the unigene that contained the domain region.

	No. of Unigenes
Total Unigenes	145
Unigene that have complete nucleotide sequence without gaps	95
Unigene that contain MYB domain	
- 1 partial domain	8
- 1 domain	23
- 1 domain and 1 partial domain	2
- 2 domains	21
Complete ORF	13



Figure 4.20: Multiple sequence alignment of 13 R2R3 MYB transcription factors ORF. MYB DNA binding regions were indicated by green arrows; R2R3 repeat domains.

Unigene39151_All	-----FSVDQP-KVRL---FNQS---SLISF-PPQQSSHVSANE-----TPPF-	215
Unigene29311_All	-----	173
Unigene4817_All	---TSTTFP-AYSMINWFDGKVFTRAKSF-LPGGQSPPEASA-----NPPVT	256
Unigene19149_All	---ETEG---LQPRSIFFDLEFFEQDAGI-WGGADSHVY-----	222
Unigene69184_All	---YNPVN---PFSFYNVSPAALWPKHHSQIQVCGPST-IPKLLKQQQITSL	215
Unigene29913_All	-----LDVSCLEFQEYS---D--YP-----	198
Unigene4579_All	LQDWMKPPPLTKRSTDPANSAVFFESSTSVASATAISSLLMLETT--RAKASVGTKIPTP	241
Unigene85656_All	-----	162
Unigene83922_All	-C--FCYRLGFQGSSEACSCHKSYAKSFVSVFQYPPLEEGQA--	223
Unigene51701_All	-T--KLPDLNLELSI-----GPPL--PQQLEANPMLQVNSRAEEFKHARRIPWP	212
Unigene20944_All	-A--ENLELDLNLISI-----GLPHCSFSLVHVHCLI-----	184
Unigene70854_All	-R--FYLDLNLDLTI-----SLPYN-----	190
Unigene4347_All	-P--KLPDLNLELCI-----EPFISIPRQLEGDFMSTKEEED--QEQRRIE--	220
Unigene39151_All	-----HHDNEI--TNRDAPPS-----FIDF--LGVGPG	239
Unigene29311_All	-----	173
Unigene4817_All	YQISNAWLHGERQH--GRKKIMP-----FIDF--LGVGAT	288
Unigene19149_All	-----FGGD-----LFM-----	229
Unigene69184_All	VSML-----	219
Unigene29913_All	-----	198
Unigene4579_All	LSILETWLLNESTEQGGTPSIDTPLDDNVALFQFNSVNVNP-	282
Unigene85656_All	-----	162
Unigene83922_All	-----	223
Unigene51701_All	QNI-----ELN-----	218
Unigene20944_All	-----	184
Unigene70854_All	-----	190
Unigene4347_All	-GF-----CFSC-----	226

Figure 4.20, continued

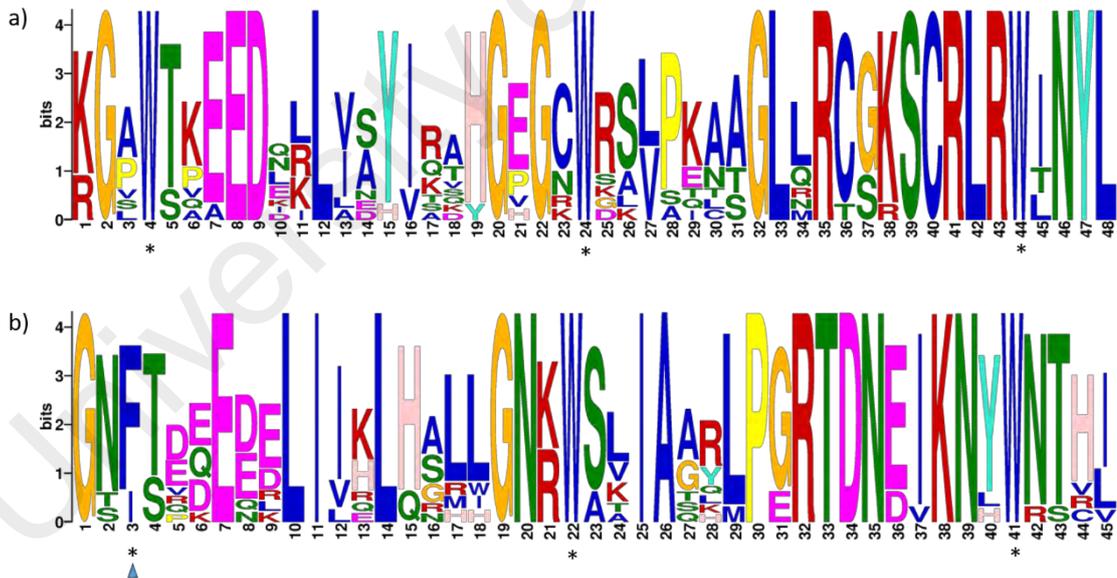


Figure 4.21: R2 and R3 MYB repeats were highly conserved in selected 11 R2R3 MYB proteins in *B. rotunda*.

These included Unigene4347_All, Unigene51701_All, Unigene20944_All, Unigene70854_All, Unigene100745_All, Unigene839222_All, Unigene85656_All, Unigene19149_All, Unigene15802_All and Unigene4579_All, Unigene91138_All. The sequence logo of a) R2 and b) R3 MYB repeats were based on domain region of selected 11 R2R3 MYB proteins in *B. rotunda*. Conserved tryptophan amino acids in the MYB domain were labelled with asterisks, blue triangle shows changed residue in the conserved R2R3 MYB domain.

Table 4.15: Summary of phylogenetic tree analysis of 13 *B. rotunda* R2R3 MYB proteins, which classify the R2R3 proteins into 7 subgroups with the predicted functions.

Functions	R2R3-MYB Subgroup	Unigene ID	No. of fold in expression level	Length of amino acid (bp)	Presence of motif	Motif on C-terminus
Abiotic/biotic stress	1	Unigene4579_All	Up-regulated: 1.8	282	Yes	YASSxxNI, SL[F/I]EKWLF[D/E]
Transcriptional repressor	4	Unigene51701_All	Up-regulated: 2	218	Yes	LNL[E/D]L
		Unigene4347_All	Non-significantly regulated	226	Yes	LNL[E/D]L
		Unigene20944_All	Non-significantly regulated	229	No	LNL[E/D]L
		Unigene70854_All	Non-significantly regulated	190	Yes	LNL[E/D]L
		Unigene83922_All	Non-significantly regulated	223	No	LNL[E/D]L
Anthocyanin biosynthesis	6	Unigene19149_All	Non-significantly regulated	229	No	KPRPR[S/T]F
Lignin biosynthesis	13	Unigene85656_All	Non-significantly regulated	162	No	DVFxKDLQRMA
Plant development	14	Unigene69184_All	Down-regulated: -1.4	219	No	SFSQLLLDPN, TSTSADQSTISWEDI
Cell wall thickening	21	Unigene39151_All	Non-significantly regulated	239	Yes	FxDfL
		Unigene29311_All	Non-significantly regulated	173	No	FxDfL
		Unigene4817_All	Up-regulated: 1	288	Yes	FxDfL
Unknown	28	Unigene29913_All	Non-significantly regulated	198	No	unknown

4.8 Molecular cloning and characterization of putative *BrMYB1*, *BrMYB2* and *BrMYB3* genes in *B. rotunda*

4.8.1 Genomic DNA extraction and total RNA extraction

B. rotunda genomic DNA and total RNA were successfully extracted (Figure 4.23). The intact 26S and 18S rRNA bands indicated that RNA was extracted. The concentration and purity of the DNA and RNA extracted are shown in Table 4.16.

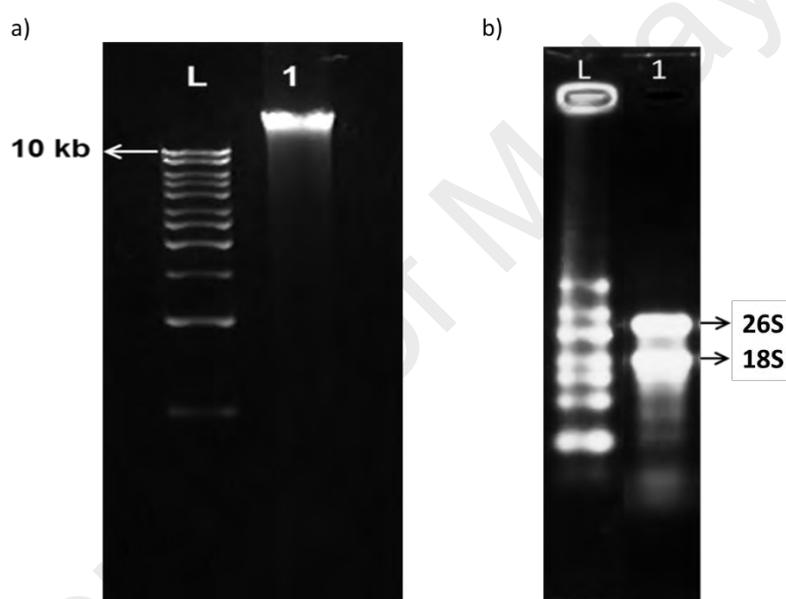


Figure 4.23: DNA and RNA extraction of *B. rotunda* callus.

a) Genomic DNA of *B. rotunda*. Lane L: Perfect 1kb DNA ladder (EURx, Poland); Lane 1: *B. rotunda* genomic DNA. b) Total RNA extraction of *B. rotunda* 14 days after phenylalanine treated callus. Lane L: RiboRuler High Range RNA ladder (Thermo Scientific, USA); Lane 1: total RNA extracted from phenylalanine treated callus.

Table 4.16: Concentration and purity of DNA and total RNA of *B. rotunda*

Sample	Concentration ng/ μ l	A260/A280
DNA	724	1.94
RNA	554.9	2.02

4.8.2 PCR verification

PCR verification was done prior to *BrMYB1*, *BrMYB2* and *BrMYB3* putative genes amplification to ensure total genomic DNA elimination in the prepared cDNA template. For verification, the cDNA template was prepared with and without the reverse transcriptase from the same RNA source. The PCR was done by employing the *BrMYB1* partial gene primers. The expected size for the putative *BrMYB1* partial gene was 416 bp. The cDNA template prepared to amplify the putative *BrMYB1*, *BrMYB2* and *BrMYB3* genes were free from genomic DNA traces as shown in Figure 4.24. There was no putative *BrMYB1* partial gene amplified product from the cDNA template that was prepared without the reverse transcriptase.

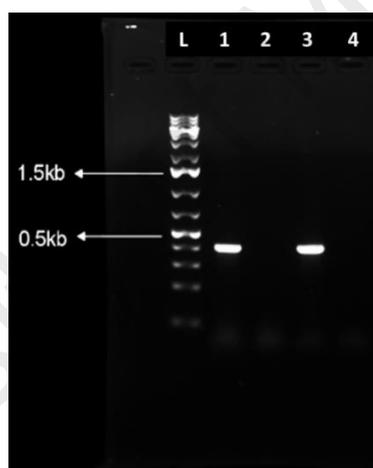


Figure 4.24: PCR verification of total RNA extracted using putative *BrMYB1* partial gene primers.

Lane L: 1kb Plus DNA ladder (Thermo Scientific, USA); Lane 1: PCR fragment using cDNA template; Lane 2: non-reverse transcriptase control; Lane 3: positive control using *B. rotunda* DNA template and Lane 4: negative control using distilled water to replace cDNA.

4.8.3 Putative *BrMYB1*, *BrMYB2* and *BrMYB3* genes amplification

Based on mapping the *B. rotunda* transcriptome data to rice transcription factor database revealed that only one MYB transcription factor was up-regulated upon phenylalanine treatment (Table 4.12). The unigene annotated as the up-regulated MYB

was identified as Unigene14544_All. Therefore, Unigene14544_All was selected for further cloning and molecular characterization in this study. The putative gene amplified based on transcriptomic Unigene14544_All reference sequence was later denoted as *BrMYB1*. Subsequently, based on phylogenetic tree analysis, Unigene51701_All and Unigene19149_All were classified as subgroup 4 and subgroup 6 R2R3 MYB, respectively (Figure 4.22). From this findings, both Unigene51701_All and Unigene19149_All were selected for further cloning and molecular characterization denoted as *BrMYB2* and *BrMYB3*, respectively.

Putative *BrMYB1*, *BrMYB2* and *BrMYB3* genes were amplified from both genomic DNA and cDNA template. The PCR was performed using the optimized annealing temperature that was selected through gradient PCR as shown in the Appendix E. The annealing temperatures for *BrMYB1*, *BrMYB2* and *BrMYB3* were 52.9°C, 51.2°C and 52.9°C; respectively. Figure 4.25 shows the MYB amplifications. The expected size for *BrMYB1*, *BrMYB2* and *BrMYB3* were 530 bp, 740 bp and 807 bp; respectively. However, *BrMYB2* and *BrMYB3* that were amplified from genomic DNA showed larger fragments as compared to the putative genes that were amplified from the cDNA template. Subsequently, a clean and pure target putative gene fragment was cloned into TOPO TA vector and introduced into Top10 competent cells. Then, the positive clones harbouring the target putative genes from both DNA and cDNA were cultured overnight. Subsequently, the plasmids harbouring target genes were isolated and sent for sequencing.

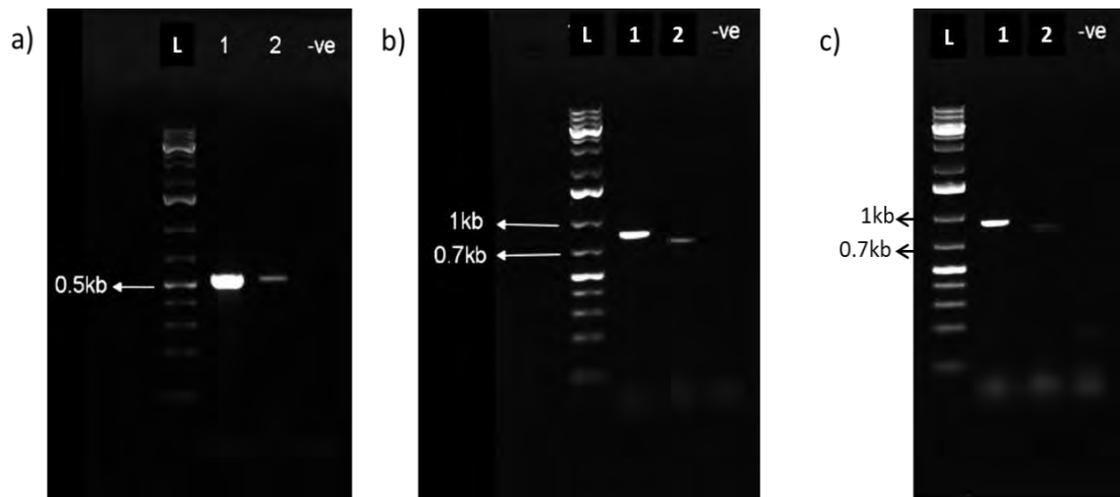


Figure 4.25: MYB open reading frame gene amplification of a) *BrMYB1*, b) *BrMYB2* and c) *BrMYB3*.

Lane L: 1kb Plus DNA ladder (Thermo Scientific, USA); Lane 1: gene amplification from DNA template; Lane 2: gene amplification from cDNA template and -ve: negative control using distilled water to replace DNA or cDNA.

4.8.4 Gel purification and plasmid prep

Prior to gel electrophoresis, gel purification was done to ensure that only one band fragment consisting of the target putative gene was excised and purified. Figure 4.26 show an intact band fragment after gel purification. The PCR library colony screening was shown in Figure A.12, Appendix F. The plasmid harbouring putative *BrMYB1*, *BrMYB2* and *BrMYB3* genes were successfully extracted (Figure 4.27).

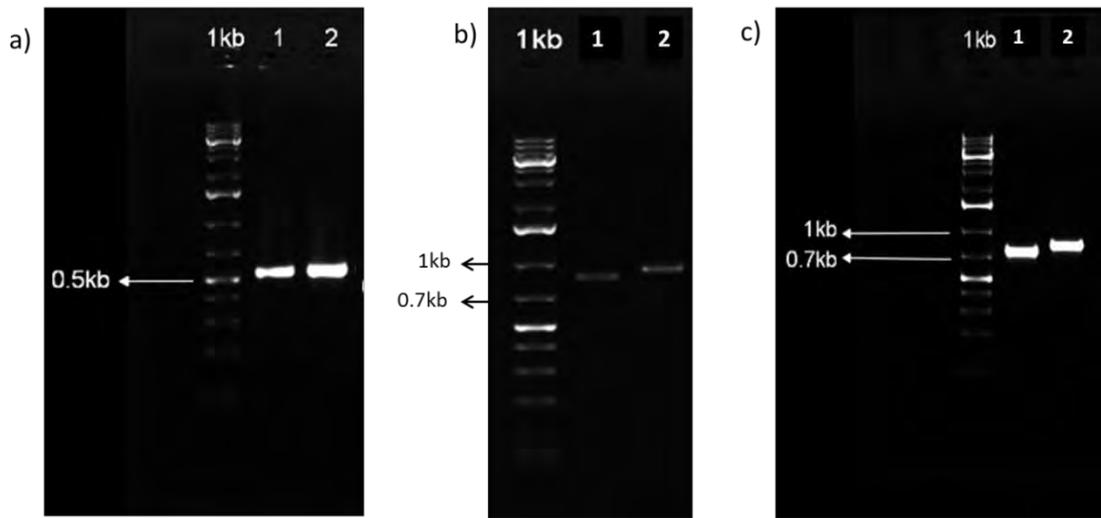


Figure 4.26: Gel purification of *BrMYB1*, *BrMYB2* and *BrMYB3* fragments.

a) *BrMYB1* fragment. Lane L: 1kb Plus DNA ladder (Thermo Scientific, USA); Lane 1: *BrMYB1* gene fragment from DNA template and Lane 2: *BrMYB1* gene fragment from cDNA template. b) DNA fragment of *BrMYB2* and *BrMYB3*. Lane L: 1kb Plus DNA ladder (Thermo Scientific, USA); Lane 1: *BrMYB2* and Lane 2: *BrMYB3*. c) cDNA fragment of *BrMYB2* and *BrMYB3*. Lane L: 1kb Plus DNA ladder (Thermo Scientific, USA); Lane 1: *BrMYB2* and Lane 2: *BrMYB3*.

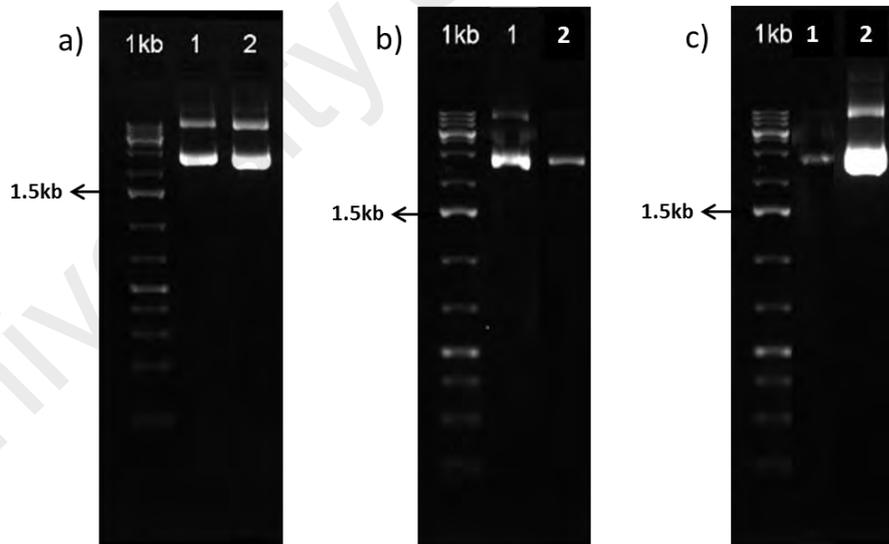


Figure 4.27: Plasmid isolation for a) *BrMYB1*, b) *BrMYB2* and c) *BrMYB3* sequencing.

Lane L: 1kb Plus DNA ladder (Thermo Scientific, USA); Lane 1: plasmid harbouring putative gene fragment from DNA template and Lane 2: plasmid harbouring gene fragment from cDNA template.

4.8.5 Sequence analysis of *BrMYB1*, *BrMYB2* and *BrMYB3* genes

4.8.5.1 Putative *BrMYB1* sequence features

The complete putative *BrMYB1* sequence from DNA and cDNA sequences of the respective clones were obtained. Putative *BrMYB1* that was amplified from DNA and cDNA template had shown clear chromatogram peaks indicating good sequencing results (Appendix K.1). The length of putative *BrMYB1* generated from DNA and cDNA were 526 and 530 bp, respectively. Putative *BrMYB1* contains an ORF of 420 bp which encodes 139 amino acids. The multiple sequence alignment between transcriptomic sequence of Unigene14544_All, putative *BrMYB1* from DNA template and putative *BrMYB1* from cDNA template were carried out using Clustal Omega. (Figure 4.28). From the results obtained, there were 4 nucleotide deletions observed in the DNA sequence. Additionally by comparing the DNA and cDNA sequence of putative *BrMYB1* with the reference transcriptomic Unigene14544_All, there were 3 nucleotide substitutions in the DNA sequence and 2 nucleotide substitutions in cDNA sequence.

The Blastx homology alignment search result shows that putative *BrMYB1* from cDNA template has one domain region (Appendix L.1). It has the highest score to MYB59-like isoform X2 from *Musa acuminata* subsp. *malaccensis* with 53% identity and 59% coverage (Table 4.17). Besides, putative *BrMYB1* also showed high score with other MYB from *Musa acuminata* subsp. *malaccensis* such as MYBAS1, MYBAS1-like isoform X3, MYBAS1-like isoform X1 and MYBAS1-like isoform X2. In addition, Blastp homology search of putative *BrMYB1* using only the ORF region showed the highest score to MYBAS2 isoform X2 from *Elaeis guineensis* with 47% identity and 79% coverage (Table 4.18). Overall, Blastx and Blastp homology search of sequenced putative *BrMYB1* annotated the Unigene14544_All as MYB-related transcription factor.

In order to classify the putative BrMYB1, phylogenetic tree using the MYB-related proteins and R3-MYB proteins from other plants were constructed. The selection of MYB proteins from other plants for phylogenetic tree construction were based on the preliminary characterization of putative BrMYB1 using blastx and blastp homology search, which showed that the putative BrMYB1 consist only one MYB domain. From the phylogenetic tree constructed, putative BrMYB1 was clustered together with MYB-related proteins and falls under the same clade with MaMYB59-like isoform X2 and MaMYBAS1-like isoform X3 proteins from *Musa acuminata* subsp. *malaccensis*; and EgMYBAS2 isoform X2 from *Elaeis guineensis* (Figure 4.29). Other R3-MYB proteins from Arabidopsis such as AtCPC, AtTRY, AtTCL1, AtETC1, AtETC2, AtETC3 and AtTCL2/CPL4 were branched into different cluster.

To further characterize the putative BrMYB1, multiple sequence alignment were conducted using putative BrMYB1 ORF region (140 amino acid length) with other MYB-related proteins from other plants such as MtMYB, ZmMYB isoform X2, OsMYBAS1-1, LuMYB1-1, LuMYB1-2, MaMYBAS1, MaMYBAS1-like isoform X2, MaMYBAS1-like isoform X3, EgMYBAS2 isoform X1 and EgMYBAS2 isoform X2 (Figure 4.30). In agreement with blastx and blastp homology search results, the protein sequence alignment verified that putative BrMYB1 exhibited only one MYB domain, the R3 repeat domain and thus classified the BrMYB1 as a R3-type MYB transcription factor (Figure 4.30). Although the putative BrMYB1 sequence alignment has shown highly conserved R3 domain at N-terminal region, putative BrMYB1 exhibited different first alpha-helix compared to other MYBAS proteins (Figure 4.30). In addition to R3 repeat domain region, putative BrMYB1 exhibited additional Motif1 and Motif2 which was located at the C-terminal region and were also found to be conserved across all MYBSA proteins. Additionally, the location of two nuclear localization signals; NLS1 and NLS2 were also

identified in the protein alignment. However, putative BrMYB1 only exhibited the NLS1 motif.



Figure 4.28: Nucleotide sequencing alignment results of transcriptomic generated Unigene14544_All sequence, putative *BrMYB1* sequence from DNA template and putative *BrMYB1* sequence from cDNA template. Forward and reverse primers are indicated by the blue arrows. For nucleotide sequence, the start and stop codons are indicated by a red arrow; with ATG as start codon and TAA as stop codon, while in the protein sequence, the start and stop codons are indicated by a red arrow (“M” as start codon and “X” as stop codon). Nucleotide deletions in the DNA sequence are indicated by a brown box. For nucleotide sequence alignment, nucleotide substitutions in DNA and cDNA sequences are denoted by brown and yellow triangle, respectively. While for protein sequence alignment, amino acid substitution are denoted by a colon (:).

Table 4.17: The percentage identity of putative *BrMYB1* cDNA with other plant MYB transcription factor at the nucleotide level over the entire sequence using NCBI blastx.

Species	Description	Max score	Total score	Query cover	E value	Identity	Accession
<i>Musa acuminata</i> subsp. <i>malaccensis</i>	PREDICTED: transcription factor MYB59-like isoform X2	89.7	89.7	59%	1.00E-18	53%	XP_009404618.1
<i>Musa acuminata</i> subsp. <i>malaccensis</i>	PREDICTED: myb-related protein MYBAS1	89.4	89.4	62%	3.00E-18	53%	XP_009421350.1
<i>Musa acuminata</i> subsp. <i>malaccensis</i>	PREDICTED: myb-related protein MYBAS1-like isoform X3	88.6	88.6	65%	3.00E-18	46%	XP_009402365.1
<i>Musa acuminata</i> subsp. <i>malaccensis</i>	PREDICTED: myb-related protein MYBAS1-like isoform X1	89	89	65%	5.00E-18	46%	XP_009402363.1
<i>Musa acuminata</i> subsp. <i>malaccensis</i>	PREDICTED: myb-related protein MYBAS1-like isoform X2	88.6	88.6	65%	6.00E-18	46%	XP_009402364.1
<i>Elaeis guineensis</i>	PREDICTED: myb-related protein MYBAS2 isoform X2	87	87	60%	1.00E-17	51%	XP_010922397.1
<i>Elaeis guineensis</i>	PREDICTED: myb-related protein MYBAS2 isoform X1	87.8	87.8	60%	1.00E-17	51%	XP_010922396.1

Table 4.18: The percentage identity of putative BrMYB1 protein sequence (ORF sequence) with other plant MYB transcription factor at the protein level over the entire sequence using NCBI blastp.

Species	Description	Max score	Total score	Query cover	E value	Identity	Accession
<i>Elaeis guineensis</i>	PREDICTED: myb-related protein MYBAS2 isoform X2	95.9	95.9	79%	2.00E-21	47%	XP_010922397.1
<i>Elaeis guineensis</i>	PREDICTED: myb-related protein MYBAS2 isoform X1	96.3	96.3	79%	4.00E-21	47%	XP_010922396.1
<i>Musa acuminata</i> subsp. <i>malaccensis</i>	PREDICTED: myb-related protein MYBAS1-like isoform X3	92	92	83%	6.00E-20	45%	XP_009402365.1
<i>Musa acuminata</i> subsp. <i>malaccensis</i>	PREDICTED: myb-related protein MYBAS1	92.4	92.4	83%	6.00E-20	49%	XP_009421350.1
<i>Musa acuminata</i> subsp. <i>malaccensis</i>	PREDICTED: transcription factor MYB59-like isoform X2	92	92	75%	8.00E-20	57%	XP_009404618.1
<i>Zea mays</i>	PREDICTED: MYB transcription factor isoform X2	77.4	77.4	74%	1.00E-14	48%	XP_008672361.1
<i>Oryza sativa</i>	MYB transcription factor MYBAS1-1	75.5	75.5	79%	4.00E-14	46%	AAV97899.1
<i>Medicago truncatula</i>	myb transcription factor	75.9	75.9	74%	5.00E-14	44%	XP_003606617.1
<i>Medicago truncatula</i>	myb transcription factor	73.9	73.9	66%	2.00E-13	47%	XP_003606616.1
<i>Linum usitatissimum</i>	MYB1-1	73.9	73.9	69%	2.00E-13	43%	ACU86961.1
<i>Linum usitatissimum</i>	MYB1-2	73.9	73.9	69%	2.00E-13	48%	ACU86962.1

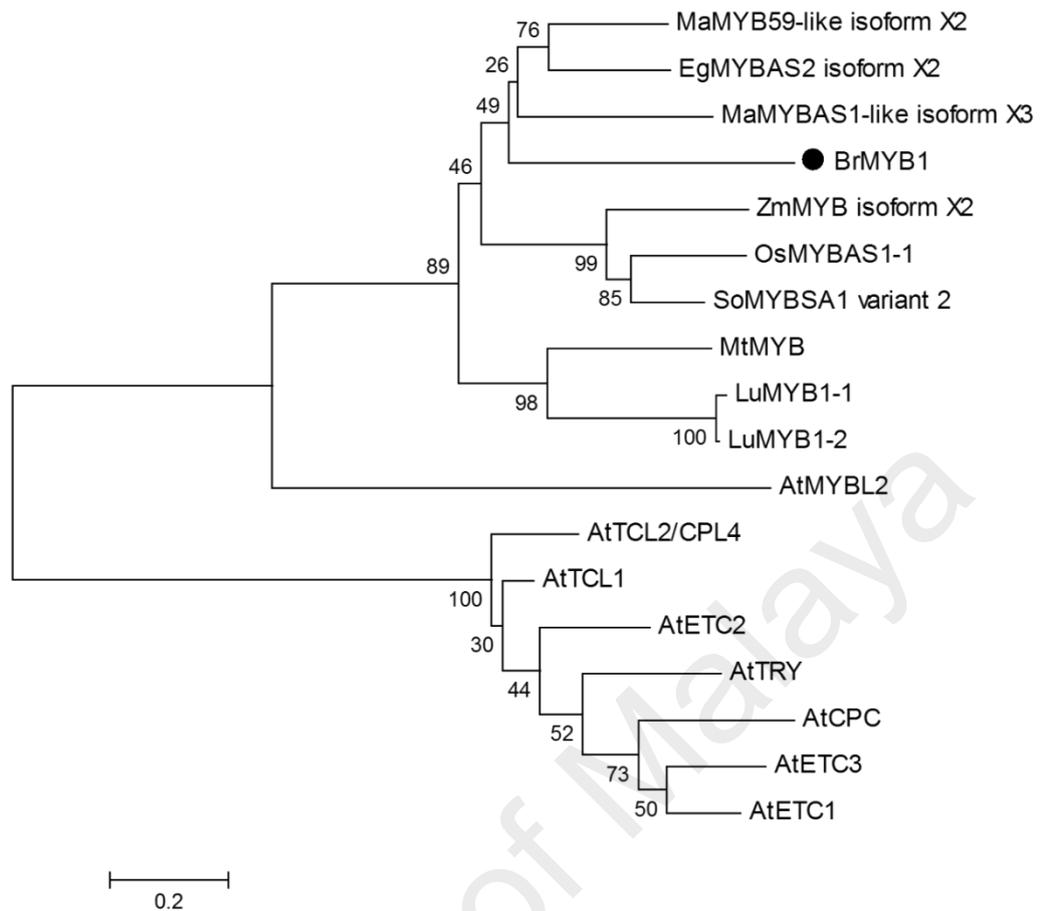


Figure 4.29: Phylogenetic tree analysis of putative BrMYB1 protein with other R3-type MYB in plants.

The tree was performed using the neighbour-joining method by the MEGA 5.10 version. The numbers next to the nodes are bootstrap values from 1000 replicates. The accession numbers of the R3 MYB proteins from other plants used to construct the phylogenetic tree are shown in the Appendix J.



Figure 4.30: Putative BrMYB1 protein multiple sequence alignment with other MYB-related proteins from other plant species.

The R3 repeat MYB DNA binding region is indicated by green arrow, while additional Motif 1 and 2 are indicated by the red boxes. The blue boxes indicate the nuclear localization signals (NLS1 and NLS2). The helices (H) are indicated with brown boxes. Conserved tryptophan amino acids in the MYB domain are labelled with red triangle, while yellow triangle indicates the tryptophan residue that was replaced by the start codon, methionine. The accession numbers of the MYB proteins from other plants to construct the multiple sequence alignment are shown in the Appendix J.

4.8.5.2 Putative *BrMYB2* sequence features

The complete putative *BrMYB2* sequence from DNA and cDNA sequences of the respective clones were obtained. Putative *BrMYB2* that was amplified from DNA and cDNA template had shown clear chromatogram peaks indicating good sequencing results (Appendix K.2). The length of putative *BrMYB2* generated from DNA and cDNA were 825 and 740 bp, respectively. Putative *BrMYB2* contains an ORF of 657 bp which encodes 218 amino acids. The multiple sequence alignment between transcriptomic sequence of Unigene51701_All, putative *BrMYB2* from DNA template and putative *BrMYB2* from cDNA template was carried out using Clustal Omega. (Figure 4.31). From the results obtained, there was an intron sequence present in the putative *BrMYB2* amplified from DNA, which indicated that putative *BrMYB2* consisted of 2 exons and one intron (Figure 4.31). There were two nucleotide substitutions in the putative *BrMYB2* generated from the cDNA template as compared to the reference Unigene51701_All sequence. In contrast, aside from the intron region, the nucleotide sequence of putative *BrMYB2* that was generated from DNA template showed 100 % nucleotide identity to the Unigene51701_All.

The Blastx homology alignment search result showed that putative *BrMYB2* from the cDNA template has two domain regions (Appendix L.2). It has the highest score to the MYB transcription factor from *Morus alba* var. *multicaulis* with 90% identity and 52% coverage (Table 4.19). Putative *BrMYB2* also showed high score with MYB from other plant species such as *Populus tomentosa*, *Chrysanthemum x morifolium*, *Populus trichocarpa*, *Gossypium arboreum* and *Leucaena leucocephala*. In addition, Blastp homology search of putative *BrMYB2* using the ORF region showed 65 – 67% identity to MYB transcriptional repressor from *Vitis vinifera*, *Brassica oleracea* var. *oleracea*, *Brassica rapa*, *Camelina sativa*, *Brassica napus*, and *Arabidopsis thaliana*; and 62 – 67 % identity to R2R3 MYB proteins from *Salvia miltiorrhiza*, *Triticum aestivum* and

Panicum virgatum (Table 4.20). Overall, Blastx and Blastp homology search of sequenced putative BrMYB2 annotated the Unigene51701_All as a R2R3 MYB transcriptional repressor protein.

A phylogenetic tree was constructed using other C2 transcriptional repressor R2R3 MYB proteins from other plant species. From the phylogenetic tree constructed, BrMYB2 was clustered together with subgroup 4 R2R3 MYB proteins such as TtMYB6 from *Tradescantia fluminensis*, AmMYB330 from *Antirrhinum majus*, AtMYB3, AtMYB6 and AtMYB8 from *Arabidopsis thaliana* (Figure 4.32). In addition, putative BrMYB2 shared the same clade with AmMYB330.

Protein multiple sequence alignments were done using the putative BrMYB2 ORF region with other selected C2 transcriptional repressor R2R3 MYB proteins from other plant species such as FaMYB1, AtMYB4, AtMYB7, AtMYB6, AtMYB8, AtMYB32, PgMYB5, AmMYB330, ZmMYB31, ZmMYB42, TaMYB4, PhMYB4, VvMYB4a, VvMYB4b and SmMYB39 (Figure 4.33). Based on the sequence alignments, the N-terminal region of repressor proteins were highly conserved as compared to the more divergent C-terminal region. The intron region of putative BrMYB2 was identified within the R3 repeat, located in between the first and second alpha-helices.

From the protein alignment, typical signature motifs for transcriptional repressors such as C1 (LlSrGIDPx[T/S]HRx[I/L]), C2 (pdLNL[D/E]Lxi[G/S]), Zinc finger or Zf (CX₁₋₂CX₇₋₁₂CX₂C) and C4 (FLGLX₄₋₇[V/L]L[D/G][F/Y][R/S]X₁LEMK) motifs were identified at the C-terminal region. Nevertheless, only two motifs were found in the BrMYB2 protein which included C1 and C2 whereas putative BrMYB2 lacked the Zf and C4 motifs. In addition to signature motifs, bHLH interaction motif ([D/E]LX₂[R/K]X₃LX₆LX₃R) which is located within the R3 repeat domain region was also found in most of the C2 repressors. The first four conserved residues were positioned

within the first α -helix, while the second last residues were located at the second α -helix of R3 domain. In BrMYB2 protein sequence, the last residue of bHLH interaction motif, arginine was replaced by lysine (Figure 4.33). Subsequently, element 3 motif (DNEI) also found within the R3 domain that was located at the α -helix 3. Similar to other C2 repressors, BrMYB2 also consisted of DNEI-type element 3 motif (Figure 4.33).

University of Malaya

```

MYB2_DNA -----
MYB2_cDNA -----
Unigene51701_All CCTTCCTCTCACCCCTCCTCTACCTTTTCTCTACCAATTCAATCTCCTCTTCTTTATAAC 60

MYB2_DNA -----
MYB2_cDNA -----
Unigene51701_All ACATAACCAGAGCTGATCCATGATCAGATTTCATCTCTTCAATCGAGGACTAGCTATATAI 120

MYB2_DNA -----TAGAGAGATTGTGTATGCACATGTGAGAGAGATGGGAGGTCGCCGTGCTG 51
MYB2_cDNA -----TAGAGAGATTGTGTATGCACATGTGAGAGAGATGGGAGGTCGCCGTGCTG 51
Unigene51701_All ATGTGTGAGTAGAGAGATTGTGTATGCACATGTGAGAGAGATGGGAGGTCGCCGTGCTG 180
*****

MYB2_DNA TGAGAAAGCTCACACGAACAAGGGAGCATGGACGAAGGAGGAAGACAATCGGCTAGTGGC 111
MYB2_cDNA TGAGAAAGCTCACACGAACAAGGGAGCATGGACGAAGGAGGAAGACAATCGGCTAGTGGC 111
Unigene51701_All TGAGAAAGCTCACACGAACAAGGGAGCATGGACGAAGGAGGAAGACAATCGGCTAGTGGC 240
*****

MYB2_DNA CTACATCCGCTCCCACGGCGAGGGGTGTGGCGGTTGCTGCCTGAGGCCGCCGCTTCT 171
MYB2_cDNA CTACATCCGCTCCCACGGCGAGGGGTGTGGCGGTTGCTGCCTGAGGCCGCCGCTTCT 171
Unigene51701_All CTACATCCGCTCCCACGGCGAGGGGTGTGGCGGTTGCTGCCTGAGGCCGCCGCTTCT 300
*****

MYB2_DNA CCGATGCGGCAAGAGCTGCCGGCTCCGCTGGATCAACTACCTCCGGCCTGACCTCAAGCG 231
MYB2_cDNA CCGATGCGGCAAGAGCTGCCGGCTCCGCTGGATCAACTACCTCCGGCCTGACCTCAAGCG 231
Unigene51701_All CCGATGCGGCAAGAGCTGCCGGCTCCGCTGGATCAACTACCTCCGGCCTGACCTCAAGCG 360
*****

MYB2_DNA CGGCAACTTCAGTGACGACGAGGACGAGCTGATCATCAAGCTCCACAGCCTCCTCGGCAA 291
MYB2_cDNA CGGCAACTTCAGTGACGACGAGGACGAGCTGATCATCAAGCTCCACAGCCTCCTCGGCAA 291
Unigene51701_All CGGCAACTTCAGTGACGACGAGGACGAGCTGATCATCAAGCTCCACAGCCTCCTCGGCAA 420
*****

MYB2_DNA CAAGTGTATATATATTCACAAATTGCTAACGATCACTCCAAATTATGCATAATTTGT 351
MYB2_cDNA CAA----- 294
Unigene51701_All CAA----- 423
***

MYB2_DNA TGTGTGTGAGGTTTTTTTAAAGTGCAGATGGTCTCTTATAGCTGCGAAATGCCAGGAA 411
MYB2_cDNA -----ATGGTCTCTTATAGCTGCGAAATGCCAGGAA 326
Unigene51701_All -----ATGGTCTCTTATAGCTGCGAAATGCCAGGAA 455
*****

MYB2_DNA GAACAGACAACGAGATCAAGAACTACTGGAACACTCATATCAGAAGAAAGTACAGAACA 471
MYB2_cDNA GAACAGACAACGAGATCAAGAACTACTGGAACACTCATATCAGAAGAAAGTACAGAACA 386
Unigene51701_All GAACAGACAACGAGATCAAGAACTACTGGAACACTCATATCAGAAGAAAGTACAGAACA 515
*****

MYB2_DNA GGGGAGTCGATCCGGTGACTACCCGGCCCTTACCAACGCCAATCCTTCTTCTTCTCCTT 531
MYB2_cDNA GGGGAGTCGATCCGGTGACTACCCGGCCCTTACCAACGCCAATCCTTCTTCTTCTCCTT 446
Unigene51701_All GGGGAGTCGATCCGGTGACTACCCGGCCCTTACCAACGCCAATCCTTCTTCTTCTCCTT 575
*****

MYB2_DNA CTAACTCGGAGCCAATTATCTCCTCCTTTCGTCGAAAGAACGGCAGAGCAGCAGCAGCG 591
MYB2_cDNA CTAACTCGGAGCCAATTATCTCCTCCTTTCGTCGAAAGAACGGCAGAGCAGCAGCAGCG 506
Unigene51701_All CTAACTCGGAGCCAATTATCTCCTCCTTTCGTCGAAAGAACGGCAGAGCAGCAGCAGCG 635
*****

```

Figure 4.31: Nucleotide sequencing alignment results of transcriptomic generated Unigene51701_All sequence, putative *BrMYB2* sequence from DNA template and putative *BrMYB2* sequence from cDNA template. Forward and reverse primers are indicated by blue arrows. For nucleotide sequence, the start and stop codons are indicated by a red arrow; with ATG as start codon and TAG as stop codon, while in the protein sequence, the start and stop codons are indicated by a red arrow (“M” as start codon and “X” as stop codon). In the nucleotide sequence, the intron is indicated by brown arrow. For nucleotide and protein sequence alignment, nucleotide and amino acid substitutions in cDNA sequences are denoted by a yellow triangle. In protein sequence alignment, the intron location is marked by a bold triangle.

```

MYB2_DNA      AGGAGGAGCCATCGTCGGCATGGCGGCTGACCAAGCTCCCGGATCTCAATCTTGAGCTCT 651
MYB2_cDNA     AGGAGGAGCCATCGTCGGCATGGCGGCTGACCAAGCTCCCGGATCTCAATCTTGAGCTCT 566
Unigene51701_A11 AGGAGGAGCCATCGTCGGCATGGCGGCTGACCAAGCTCCCGGATCTCAATCTTGAGCTCT 695
*****

MYB2_DNA      CCATTGGCCCTCCCTTGCCCGCAGCAATTAGAAGCTAATCCAATGCTGCAGGTC AATTCCA 711
MYB2_cDNA     CCATTGGCCCTCCCTTGCCCGCAGCAATTAGAAGCTAATCCAATGCTGCAGGTC AATTCCA 626
Unigene51701_A11 CCATTGGCCCTCCCTTGCCCGCAGCAATTAGAAGCTAATCCAATGCTGCAGGTC AATTCCA 755
*****

MYB2_DNA      GAGCAGAAGAATTTAAGCACGCGCGGAGGATTCCTTGGCCTCAAAACATAGA AACTCAACT 771
MYB2_cDNA     GAGCAGAAGAATTTAAGCACGCGCGGAGGATTCCTTGGCCTCAAAACATAGA AACTCAACT 686
Unigene51701_A11 GAGCAGAAGAATTTAAGCACGCGCGGAGGATTCCTTGGCCTCAAAACATAGA AACTCAACT 815
*****

MYB2_DNA      AGTCCTGTAGTTTAGTTAATTAGATGTGGATTTCTTCCTGTACTGTTCTGCTC----- 825
MYB2_cDNA     AGTCCTGTAGTTTAGTTAATTAGATGTGGATTTCTTCCTGTACTGTTCTGCTC----- 740
Unigene51701_A11 AGTCCTGTAGTTTAGTTAATTAGATGTGGATTTCTTCCTGTACTGTTCTGCTC----- 875
*****
<-----
MYB2_DNA      -----
MYB2_cDNA     -----
Unigene51701_A11 TTAATCTTTAAATTAATGCATGATGGTAGCTAAGCGATCTTAATGTGAAGAATGTGAT 935

MYB2_DNA      -----
MYB2_cDNA     -----
Unigene51701_A11 TATAGACTGTTTTGATACTCAATCTGTTTTGTGTTGTTGTTGTTGAGC 984

```

Figure 4.31, continued

Table 4.19: The percentage identity of putative *BrMYB2* cDNA with other plant MYB transcription factor at the nucleotide level over the entire sequence using NCBI blastx.

Species	Description	Max score	Total score	Query cover	E value	Identity	Accession
<i>Morus alba</i> var. <i>multicaulis</i>	MYB transcription factor	256	256	52%	6.00E-81	90%	AEE81751.1
<i>Populus tomentosa</i>	MYB domain transcription factor	255	255	52%	1.00E-80	90%	AKE81094.1
<i>Chrysanthemum x morifolium</i>	MYB1	256	256	53%	9.00E-81	89%	AEO27497.1
<i>Populus trichocarpa</i>	myb family transcription factor family protein	254	254	52%	2.00E-80	89%	XP_002306180.1
<i>Gossypium arboreum</i>	Myb-related protein	254	254	53%	2.00E-81	88%	KHF99993.1
<i>Gossypium arboreum</i>	Myb-related protein	251	251	53%	2.00E-80	87%	KHG10399.1
<i>Leucaena leucocephala</i>	MYB1	253	253	53%	2.00E-80	87%	ADY38393.2

Table 4.20: The percentage identity of putative BrMYB2 protein sequence (ORF sequence) with other plant MYB transcription factor at the protein level over the entire sequence using NCBI blastp.

Species	Description	Max score	Total score	Query cover	E value	Identity	Accession
<i>Vitis vinifera</i>	flavonoid-related R2R3 MYB4a repressor transcription factor	273	273	86%	2.00E-88	69%	NP_001268129.1
<i>Brassica oleracea</i> var. <i>oleracea</i>	PREDICTED: transcription repressor MYB4	270	270	83%	2.00E-86	67%	XP_013629672.1
<i>Brassica rapa</i>	PREDICTED: transcription repressor MYB4	271	271	83%	5.00E-87	65%	XP_009109540.1
<i>Camelina sativa</i>	PREDICTED: transcription repressor MYB4 isoform X1	271	271	83%	3.00E-87	66%	XP_010436959.1
<i>Brassica napus</i>	PREDICTED: transcription repressor MYB4-like	270	270	83%	2.00E-86	67%	XP_013652821.1
<i>Camelina sativa</i>	PREDICTED: transcription repressor MYB4-like	271	271	83%	5.00E-87	65%	XP_010446371.1
<i>Salvia miltiorrhiza</i>	R2R3 MYB transcription factor 39	268	268	85%	1.00E-86	71%	AGS48990.1
<i>Triticum aestivum</i>	R2R3 MYB transcriptional factor	267	267	87%	2.00E-86	69%	AEG64799.1
<i>Panicum virgatum</i>	R2R3-MYB transcriptional factor PvMYB4a	267	267	94%	6.00E-86	62%	AEM17348.1
<i>Panicum virgatum</i>	R2R3-MYB transcriptional factor PvMYB4d	266	266	94%	2.00E-85	62%	AEM17351.1
<i>Arabidopsis thaliana</i>	transcription repressor MYB4	266	266	83%	2.00E-85	66%	NP_195574.1

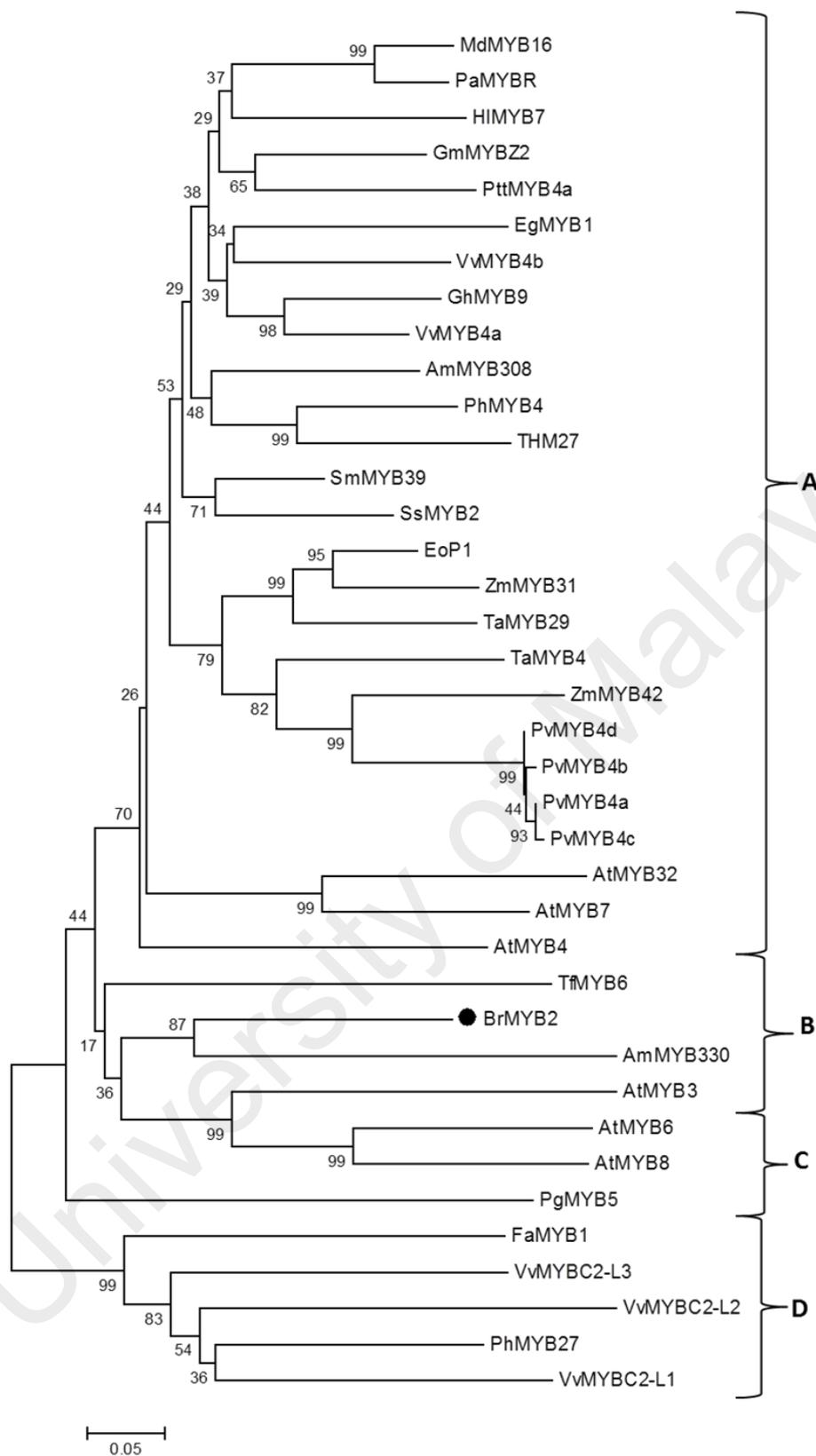


Figure 4.32: Phylogenetic tree analysis of putative BrMYB2 protein with other C2 repressor R2R3 MYB proteins in plants.

The tree was performed using the neighbour-joining method by the MEGA 5.10 version. The numbers next to the nodes are bootstrap values from 1000 replicates. The accession numbers of the R2R3 MYB proteins from other plants to construct the phylogenetic tree are shown in the Appendix J.

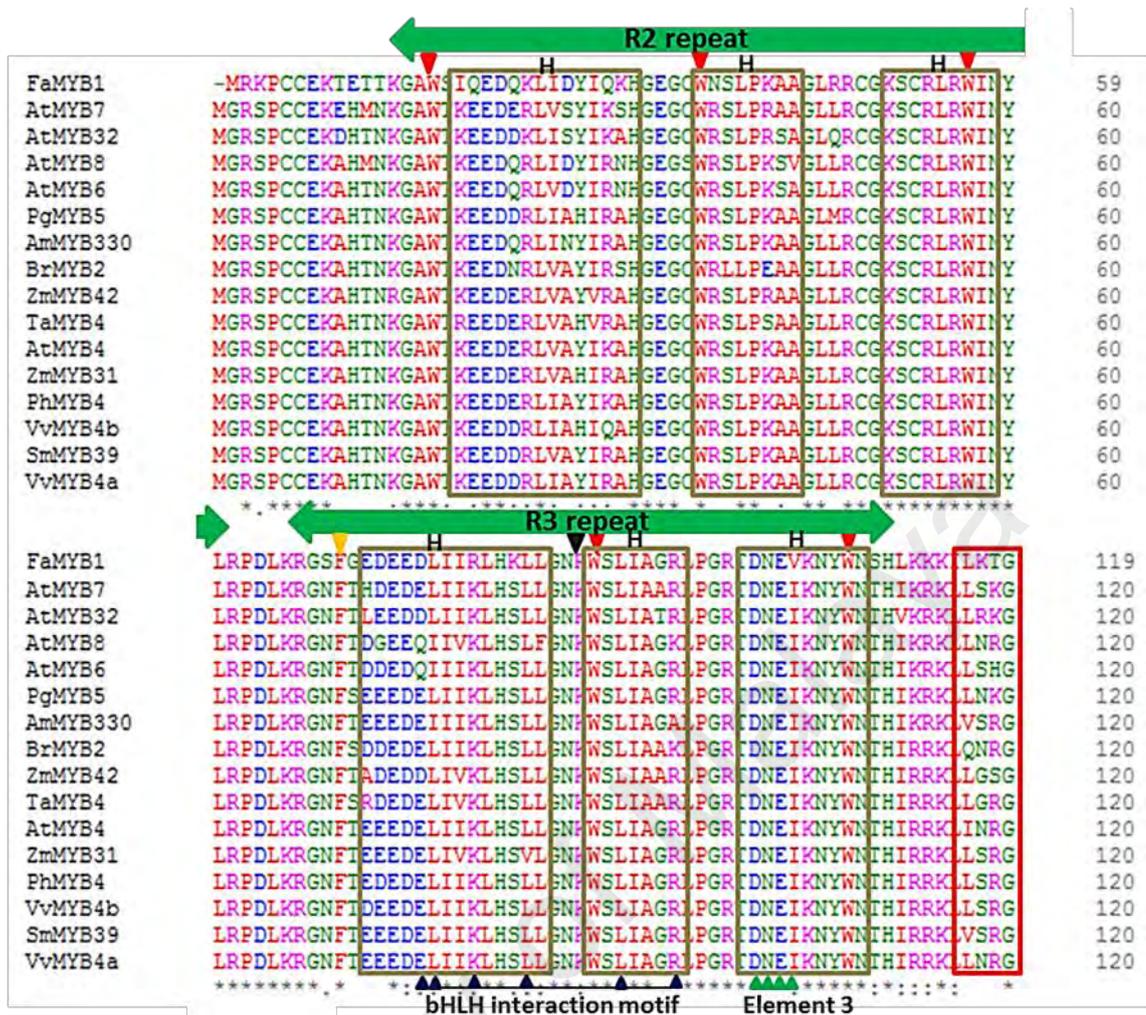


Figure 4.33: Putative BrMYB2 protein multiple sequence alignment with other C2 repressor R2R3 proteins from other plant species.

The black triangle denotes the intron location in the BrMYB2 protein sequence. The R2 and R3 repeat MYB DNA binding region are indicated by a green arrow, while C1, C2 and C4 motif are indicated by the red boxes. Zf motif is indicated by an orange box with the cysteine residue (C) denoted by an orange triangle. The bHLH interaction motif is shown by the blue triangle while the complete interaction signature was underlined. The element 3 is highlighted with a green triangle. The helices (H) are indicated with brown boxes. Conserved tryptophan amino acids in the MYB domain are labeled with a red triangle, while a yellow triangle indicates the tryptophan residue that has been replaced by phenylalanine residue in R3 repeat. The accession numbers of the R2R3 MYB proteins from other plants to construct the multiple sequence alignment are shown in the Appendix J.

C1 motif		
FaMYB1	TTLRPNKPHENHAPNNKL-----VKLFNKMDDE-----	148
AtMYB7	IDPATHRGINEAKIS---DLKKT-KDQIVKDVSVTKFEETDKSG-----DQK	164
AtMYB32	IDPATHRPINEIKTSQDSSDSSKT-EDPLVKILSFGPQLEKIANFG-----DER	168
AtMYB8	IDPKTHGSIIEPKTISFHP-----RNEDLKSTFP-GSVK-----LKMETSCE	161
AtMYB6	IDPQTHRQINEGKTVSSQV---VVPIQNDAVEYSFSNLAVK-----PKTENSDD	166
PgMYB5	LDPQSHRPLGQVHSSNTTCSSLPAPE---HEILAFQSPRTPEI-----	160
AmMYB330	IDPQTHRSLNSATTATATPTVNNNS---CLDFRTSPSNSKN-----ICMPTT-DNNN	168
BrMYB2	VDFVTHRPFTINANPSSSP--SNSEPI-----ISSFRRKNS-----	153
ZmMYB42	IDFVTHRRVAGSAATTIS-----FQPSPTAVAA-----	149
TaMYB4	IDFVTHRPLDAAIVSFV-----H-PA-----	141
AtMYB4	IDFTSHRPIQEGSASQDSKPTQLEPVISNTINISFTSAPKVEIF-----	164
ZmMYB31	IDFVTHRPVTEHNASNI-----TISFETEVAARDDKKGAVFRLKEEER	166
PhMYB4	IDFTTHRMNEPSTQKVT-----TISFAAGNEDI-----KD	151
VvMYB4b	IDPLTHRPINEPAVDVA-----TVSFGGVLLKKE-----EV	150
SmMYB39	IDFTTHRPINEAEAPAT-----TISFNSSNK-----	147
VvMYB4a	IDPSTHRRINEPSPDV-T-----TISFAAAVKEE-----EK	150
C2 motif		
FaMYB1	-----VVDEVSSADSAAAGCLPELNLDLTLISIKTSTGMADPQVA-----	187
AtMYB7	QNKYIRNGLVC-----KEERVVVEEKIQPDNLNLELRISPPNQNR-----	204
AtMYB32	IQKRVE-----YSVVEERGLDLNLELRISPPNQDKLHDERNLR	206
AtMYB8	NCASTS-GTTTD-----EDLRLSVDCDYRYDHLKELNLDLTLGYS--PT--RFVGVG	209
AtMYB6	NGASTS-GTTTD-----EDLRQNGECYSDNSGHIKLNLDLTLGFGSWSG--RIVGVG	216
PgMYB5	-----ADFFQY---ERSESSPIEPAAS--KDEEYPDNLNLELCISLFPVHSAPAASRASS	208
AmMYB330	NSSSSITDDTKCNSSTIEESQSLITPPPK--EEKSTPLVDLELSLGLFSSQCNKNSVLN	226
BrMYB2	-----SSSSSEEEPPSSAW--RLTKQPDNLNLELSIGPPLPQQLEANPMLQ	195
ZmMYB42	-----AAETAQAQAPIKAETAAV--KAPRQPDNLNLDLCISPPCQHEDDGESEEE	196
TaMYB4	-----EATKQQAETEER--KPPRQPDNLNLDLCISLFPQQEERPPARA	181
AtMYB4	---HESISFPGKSEKISMLTFKEEKDECP--VQEKTPDNLNLELRISLPPDDVD--	211
ZmMYB31	NKATMVVGRDRQSQSQSHSHPAGEWQGGK--RPLKQPDNLNLDLCISPPCQEEEMEEAA-	223
PhMYB4	QKTSIKAEFEQI-----KDDEIISKP--IKEQCPDNLNLELRISPPVQGHSDRALQ--	199
VvMYB4b	---MSSGFGCK-----S---RNP--DEENQPDNLNLELRISPPHHHPSPFVLD--	189
SmMYB39	-----LL-----G---KEER--RSPKQPDNLNLDLRISPPVQQ--EPFK--	178
VvMYB4a	INISSTGGFGCK-----I---EKNP--VIEKQPDNLNLELRISPPVQQAETPLK--	194
Zf motif		
FaMYB1	-----EISTC---TASRFYEMNDM-ECSSSETVKQQTENSSSISY	187
AtMYB7	-----FGRVKYFC---SACRFYFGNGK-ECSCNNVKQQTEDSSSSSY	239
AtMYB32	-----Y-----	244
AtMYB8	SC-----Y-----	212
AtMYB6	SS-----ADSKPWCDPVMEARLSLL-----	236
PgMYB5	VD-----RTVDSKFNSSGSEIC---CPMGLQVNY-GAQCNRYSEE-----	244
AmMYB330	SSSSGF---Y---DLFRPPAKVAQRMCV--CKWTLGLQKGEQFCNCQSFNGF-----	270
BrMYB2	VNSRA-----EEFKHARRIP---WFQNIELN-----	218
ZmMYB42	ELDLIKPAVVKREALQAGHGHHGHC---LGCGLGGQKGAAGCSCSN-----	240
TaMYB4	CA-----KPVKMEQLQQGGIC---FRCSILRVGAA-TECSC-----	214
AtMYB4	-----RLQGHGKSTTPFC---FKCSLGMINGM-ECRCGRMRCDVVGSS--	251
ZmMYB31	-----MRVRPAVKREAGIC---FGCSLGLPRTA-DCKCSS-----	254
PhMYB4	-----QSTITSGGASTIC---FTCSLGLKNNK-GCSCSRNRS--	232
VvMYB4b	-----SRERRMIC---FYCSLGLEKSK-ECSCSSGSG-----	217
SmMYB39	-----T---GTASSSTIIC---FACSLGIQNSK-DCSCNT-----	207
VvMYB4a	-----T---GGRSSSTIIC---FACSLGIPNSE-ECSCSIGIS-----	225
C4 motif		
FaMYB1	-----SSIDISSNVGYDFLGLK---TRILDFRSLEMK-----	187
AtMYB7	SSIDISS-SIGYDFLGLN---TRVLDIFSILEMK-----	269
AtMYB32	SSIDISS-SIGYDFLGLN---TRVLDIFSILEMK-----	274
AtMYB8	-----	212
AtMYB6	-----	236
PgMYB5	---NASGFSSHYRLVL-----	257
AmMYB330	-----YRVC-----	274
BrMYB2	-----	218
ZmMYB42	---GHRFLGLR---TSVLDFRGLEMK-----	260
TaMYB4	---GSNFLGLR---AGMLDFRGLEMK-----	234
AtMYB4	---KGSMDMSNGFDLGLAKKETTSLGFRSLEMK-----	282
ZmMYB31	---SSFLGLR---TAMLDFRSLEMK-----	273
PhMYB4	---M---NVAGYDFLGLK---TNGLDYRTLETRK-----	258
VvMYB4b	---N-GS---SDFLGMN---NGVLDYRNLEMKD-----	242
SmMYB39	---TNSGFDLGLK---SGVLDYRRLEMK-----	230
VvMYB4a	---S-GSSSSGYDFLGLT---SGVLDYRGLEMK-----	251

Figure 4.33, continued

4.8.5.3 Putative *BrMYB3* sequence features

The complete putative *BrMYB3* sequence from DNA and cDNA sequences of the respective clones were obtained. Putative *BrMYB3* that was amplified from DNA and cDNA template had shown clear chromatogram peaks indicating good sequencing results (Appendix K.3). The length of putative *BrMYB3* generated from DNA and cDNA were 887 and 807 bp, respectively. Putative *BrMYB3* contains an ORF of 690 bp which encodes 229 amino acids. The multiple sequence alignment between transcriptomic sequence of Unigene19149_All, putative *BrMYB3* from DNA template and putative *BrMYB3* from cDNA template was carried out using Clustal Omega. (Figure 4.34). From the results obtained, there was an intron sequence present in the putative *BrMYB3* amplified from DNA, which indicated that putative *BrMYB3* consist of 2 exons and one intron (Figure 4.34). The nucleotide sequences from both DNA and cDNA putative *BrMYB3* showed 100% nucleotide identity to the reference transcriptomic Unigene19149_All sequence.

The Blastx homology alignment search results showed that putative *BrMYB3* cDNA has two domain regions (Appendix L.3). It has the highest score to MYB3-like from *Musa acuminata* subsp. *malaccensis* with 55% identity and 73% coverage (Table 4.21). Using Blastp homology search, putative *BrMYB3* was annotated as MYB3-like from *Musa acuminata* subsp. *malaccensis* with 55 % identity and 86 % coverage (Table 4.22). Putative *BrMYB3* also showed homology to WER-like transcription factor from *Elaeis guineensis*, *Phoenix dactylifera* and *Musa acuminata* subsp. *malaccensis*; R2R3 MYB transcription factor from *Prunus avium* and *Musa* AB Group; anthocyanin regulatory C1 protein from *Elaeis guineensis*, *Morus notabilis* and *Eucalyptus grandis*; tannin-related transcription factor from *Medicago truncatula*; and transcription repressor from *Gossypium raimondii*, *Prunus mume* and *Eucalyptus grandis*. Through analysis using Blastx and Blastp homology search, it can be concluded that putative *BrMYB3* can be classified under R2R3 MYB transcription factor.

Based on previous phylogenetic tree analysis (Figure 4.22), Unigene19149_All was clustered under subgroup 6 R2R3 MYB transcription factor. However, through the Blastx and Blastp homology search, N-terminal of putative BrMYB3 showed high percentage identity to R2R3 MYB proteins that harboured the bHLH interaction motif such as WER-like, C1-like, tannin-related transcription factors and transcription repressor proteins. Hence, a phylogenetic tree was constructed using bHLH containing subgroup 4, 5, 6, 15 and 27 of R2R3 MYB proteins from other plant species to further classify the putative BrMYB3 protein (Figure 4.35). Based on the phylogenetic tree analysis, putative BrMYB3 was clustered with EgWER-like, PdWER-like, MaWER-like, PdMYB23-like, EgMYB308-like, MaMYB3-like and EgrC1-like proteins.

Subsequently, to analyse the conserved motifs present in BrMYB3, protein sequence alignment was done using bHLH containing R2R3 MYB proteins from subgroup 4, 5, 6, 15 and 27 from other plant species (Figure 4.36). Similar to blastx and blastp results, the N-terminal region of putative BrMYB3 protein sequence showed highly conserved sequence with other bHLH containing R2R3 MYB proteins. Accordingly, the bHLH interaction motif was located at the first and second α -helix of R3 domain. Putative BrMYB3 has a perfect match with the bHLH interaction signature residues (Figure 4.36). Subsequently, the element 3 motif was found in the α -helix 3 of R3 domain. Similar to putative BrMYB2, putative BrMYB3 also exhibited DNEI conserved element 3 motif (Figure 4.36).

From the protein sequence alignment, there were 8 conserved motifs that define subgroup 4, 5, 6, 15 and 27 that were identified at the C-terminal region. The motifs included C1 (LlSrGIDPx[T/S]HRx[I/L]), C2 (pdLNL[D/E]Lxi[G/S]), C3 (DDxF[S/P]SFL[N/D]SLIN[E/D]), C4 (FLGLX₄₋₇[V/L]L[D/G][F/Y][R/S]X₁LEMK), Zf (CX₁₋₂CX₇₋₁₂CX₂C), subgroup 5-specific (DExWLRxxT), subgroup 6-specific

(KPRPR[S/T]F) and subgroup 15-specific (WVxxDxFELSxL) motifs. Based on protein sequences, putative BrMYB3 does not exhibit any of the motifs that defined subgroup 4, 5, 6, 15 and 27 (Figure 4.36). This result was in accordance with the phylogenetic tree analysis as putative BrMYB3 does not fall into the same clade with subgroup 4, 5, 6, 15 and 27 (Figure 4.36).

University of Malaya

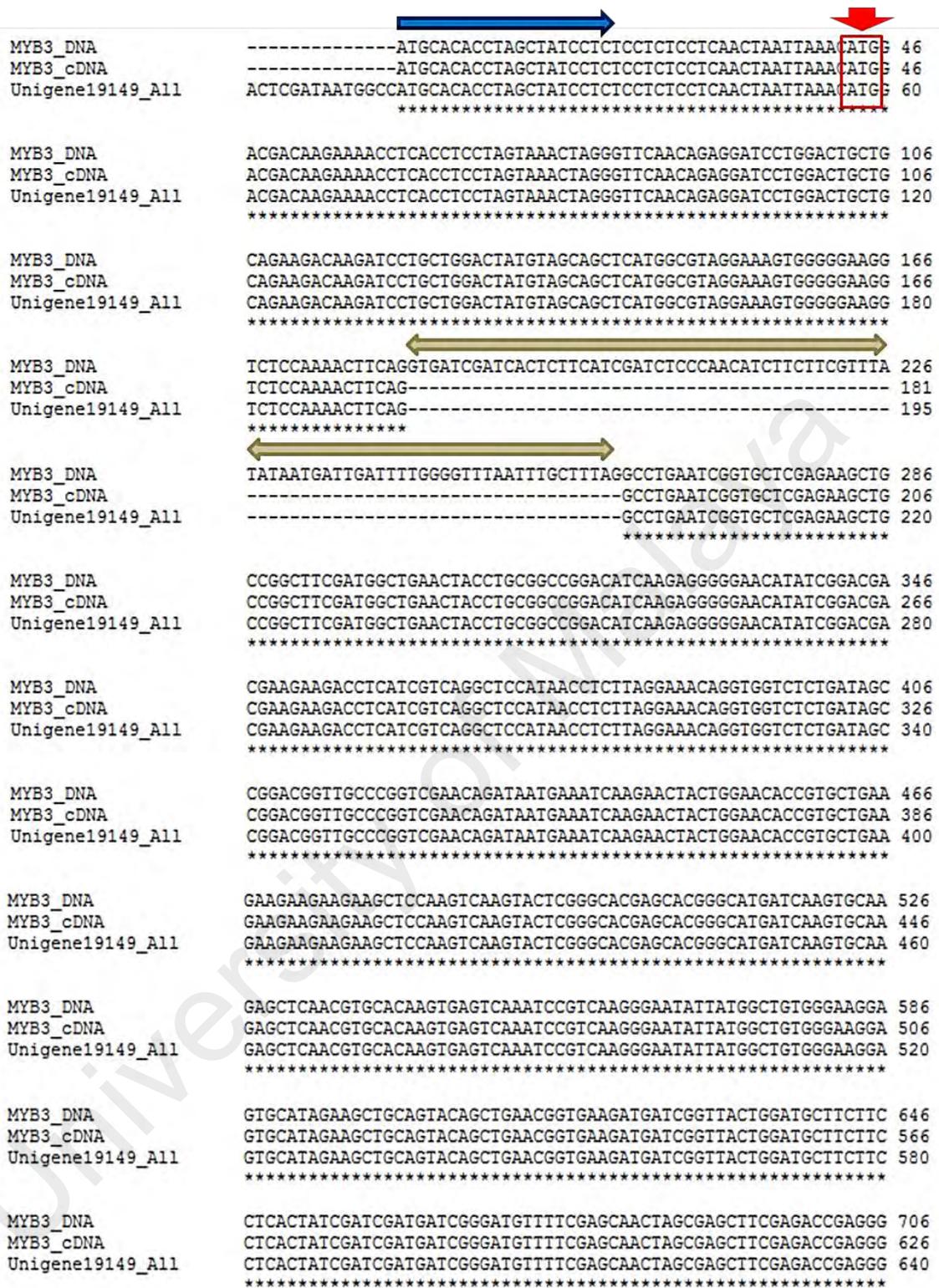


Figure 4.34: Nucleotide sequencing alignment results of transcriptomic generated Unigene19149_All sequence, putative *BrMYB3* sequence from DNA template and putative *BrMYB3* sequence from cDNA template. Forward and reverse primers are indicated by blue arrows. For nucleotide sequence, the start and stop codons are indicated by a red arrow; with ATG as start codon and TAG as stop codon, while in the protein sequence, the start and stop codons are indicated by a red arrow (“M” as start codon and “X” as stop codon). In the nucleotide sequence, the intron is indicated by a brown arrow. In protein sequence alignment, the intron location is marked by a bold triangle.

MYB3_DNA	ACTGCAGCCGAGGAGCATTTTCGACTTGGAAATTCCTCGAACAAAGATGCCGGAATTTGGGG	766
MYB3_cDNA	ACTGCAGCCGAGGAGCATTTTCGACTTGGAAATTCCTCGAACAAAGATGCCGGAATTTGGGG	686
Unigene19149_All	ACTGCAGCCGAGGAGCATTTTCGACTTGGAAATTCCTCGAACAAAGATGCCGGAATTTGGGG	700

MYB3_DNA	AGGTGCTGATTCTCATGTGTATTTTGGTGGCGATTGTTTCATCTAGCTAGCGATCGAGTC	826
MYB3_cDNA	AGGTGCTGATTCTCATGTGTATTTTGGTGGCGATTGTTTCATCTAGCTAGCGATCGAGTC	746
Unigene19149_All	AGGTGCTGATTCTCATGTGTATTTTGGTGGCGATTGTTTCATCTAGCTAGCGATCGAGTC	760

MYB3_DNA	ATGTGCAATTGCATCCTCAAACGCATTAATAATTCCTCCTTGCCATCTGAACAAGATGCT	886
MYB3_cDNA	ATGTGCAATTGCATCCTCAAACGCATTAATAATTCCTCCTTGCCATCTGAACAAGATGCT	806
Unigene19149_All	ATGTGCAATTGCATCCTCAAACGCATTAATAATTCCTCCTTGCCATCTGAACAAGATGCT	820

MYB3_DNA	G-----	887
MYB3_cDNA	G-----	807
Unigene19149_All	GTTCAGATGTATTACCATAAGGCTATTATTATTATAAGTCTTTCACCAAAATAAAATGT	880
	■	
MYB3_DNA	----	
MYB3_cDNA	----	
Unigene19149_All	TGAG 884	

Figure 4.34, continued

University of Malaya

Table 4.21: The percentage identity of putative *BrMYB3* cDNA with other plant MYB transcription factor at the nucleotide level over the entire sequence using NCBI blastx.

<i>Species</i>	Description	Max score	Total score	Query cover	E value	Identity	Accession
<i>Musa acuminata subsp. malaccensis</i>	PREDICTED: transcription factor MYB3-like	188	188	73%	7.00E-55	55%	XP_009410062.1
<i>Musa acuminata subsp. malaccensis</i>	PREDICTED: transcription factor WER-like	187	187	37%	1.00E-53	83%	XP_009403151.1
<i>Elaeis guineensis</i>	PREDICTED: myb-related protein 308-like	183	183	75%	2.00E-52	48%	XP_010917231.1
<i>Phoenix dactylifera</i>	PREDICTED: transcription factor MYB23-like	183	183	38%	4.00E-52	80%	XP_008784295.1
<i>Elaeis guineensis</i>	PREDICTED: myb-related protein Myb4-like	182	182	38%	6.00E-52	81%	XP_010936941.1
<i>Elaeis guineensis</i>	PREDICTED: transcription factor WER-like	182	182	38%	7.00E-52	80%	XP_010932429.1
<i>Jatropha curcas</i>	PREDICTED: transcription factor MYB82-like	181	181	40%	1.00E-51	77%	XP_012076322.1

Table 4.22: The percentage identity of putative BrMYB3 protein sequence (ORF sequence) with other plant MYB transcription factor at the protein level over the entire sequence using NCBI blastp.

Species	Description	Max score	Total score	Query cover	E value	Identity	Accession
<i>Musa acuminata subsp. malaccensis</i>	PREDICTED: transcription factor MYB3-like	213	213	86%	4.00E-65	55%	XP_009410062.1
<i>Elaeis guineensis</i>	PREDICTED: transcription factor WER-like	202	202	86%	4.00E-60	53%	XP_010932429.1
<i>Elaeis guineensis</i>	PREDICTED: myb-related protein 308-like	201	201	86%	9.00E-60	53%	XP_010917231.1
<i>Phoenix dactylifera</i>	PREDICTED: transcription factor MYB23-like	200	200	86%	3.00E-59	51%	XP_008784295.1
<i>Phoenix dactylifera</i>	PREDICTED: transcription factor WER-like	199	199	86%	8.00E-59	52%	XP_008786612.1
<i>Musa acuminata subsp. malaccensis</i>	PREDICTED: transcription factor WER-like	197	197	90%	5.00E-58	48%	XP_009403151.1
<i>Prunus avium</i>	R2R3 MYB transcription factor	182	182	83%	3.00E-52	49%	ALH21142.1
<i>Musa AB Group</i>	R2R3 MYB transcription factor TT2-1	181	181	59%	5.00E-52	57%	AKB92818.1
<i>Elaeis guineensis</i>	PREDICTED: anthocyanin regulatory C1 protein-like	184	184	56%	7.00E-53	67%	XP_010936942.1
<i>Morus notabilis</i>	Anthocyanin regulatory C1 protein	184	184	60%	2.00E-52	62%	XP_010090128.1
<i>Eucalyptus grandis</i>	PREDICTED: anthocyanin regulatory C1 protein-like	181	181	54%	8.00E-52	66%	XP_010047390.1
<i>Medicago truncatula</i>	tannin-related R2R3 MYB transcription factor	185	185	52%	2.00E-53	74%	AFJ53058.1
<i>Gossypium raimondii</i>	PREDICTED: transcription repressor MYB6-like	183	183	49%	5.00E-53	74%	XP_012488054.1
<i>Prunus mume</i>	PREDICTED: transcription repressor MYB6-like	183	183	53%	8.00E-53	67%	XP_008220564.1
<i>Eucalyptus grandis</i>	PREDICTED: transcription repressor MYB5-like	181	181	52%	5.00E-52	69%	XP_010047391.1

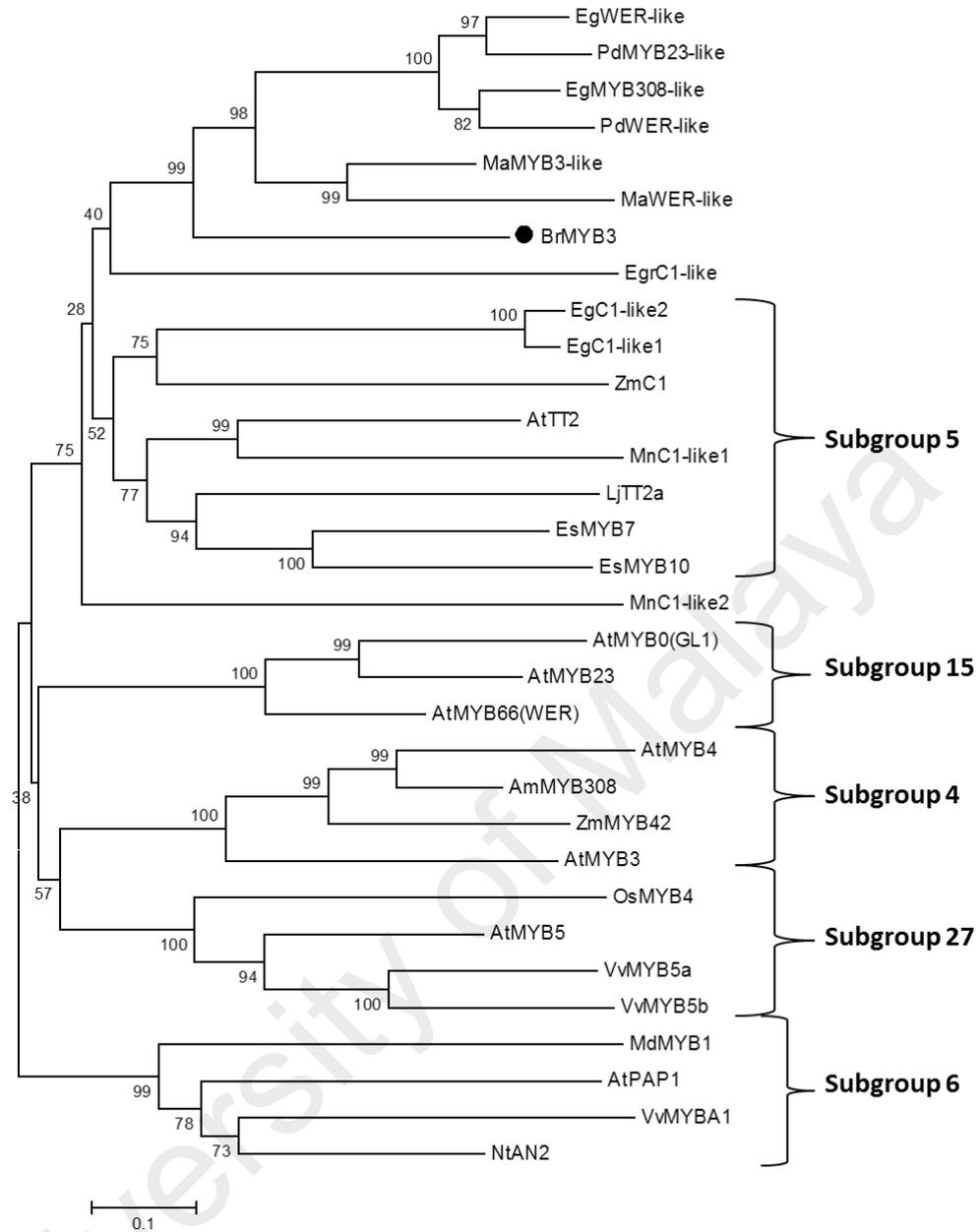


Figure 4.35: Phylogenetic tree analysis of putative BrMYB3 protein with other R2R3 MYB proteins in plants that harbouring bHLH interaction motif which include proteins from subgroup 4, 5, 6 15 and 27.

The tree was performed using the neighbour-joining method by the MEGA 5.10 version. The numbers next to the nodes are bootstrap values from 1000 replicates. The accession numbers of the R2R3 MYB proteins from other plants to construct the phylogenetic tree are shown in the Appendix J.

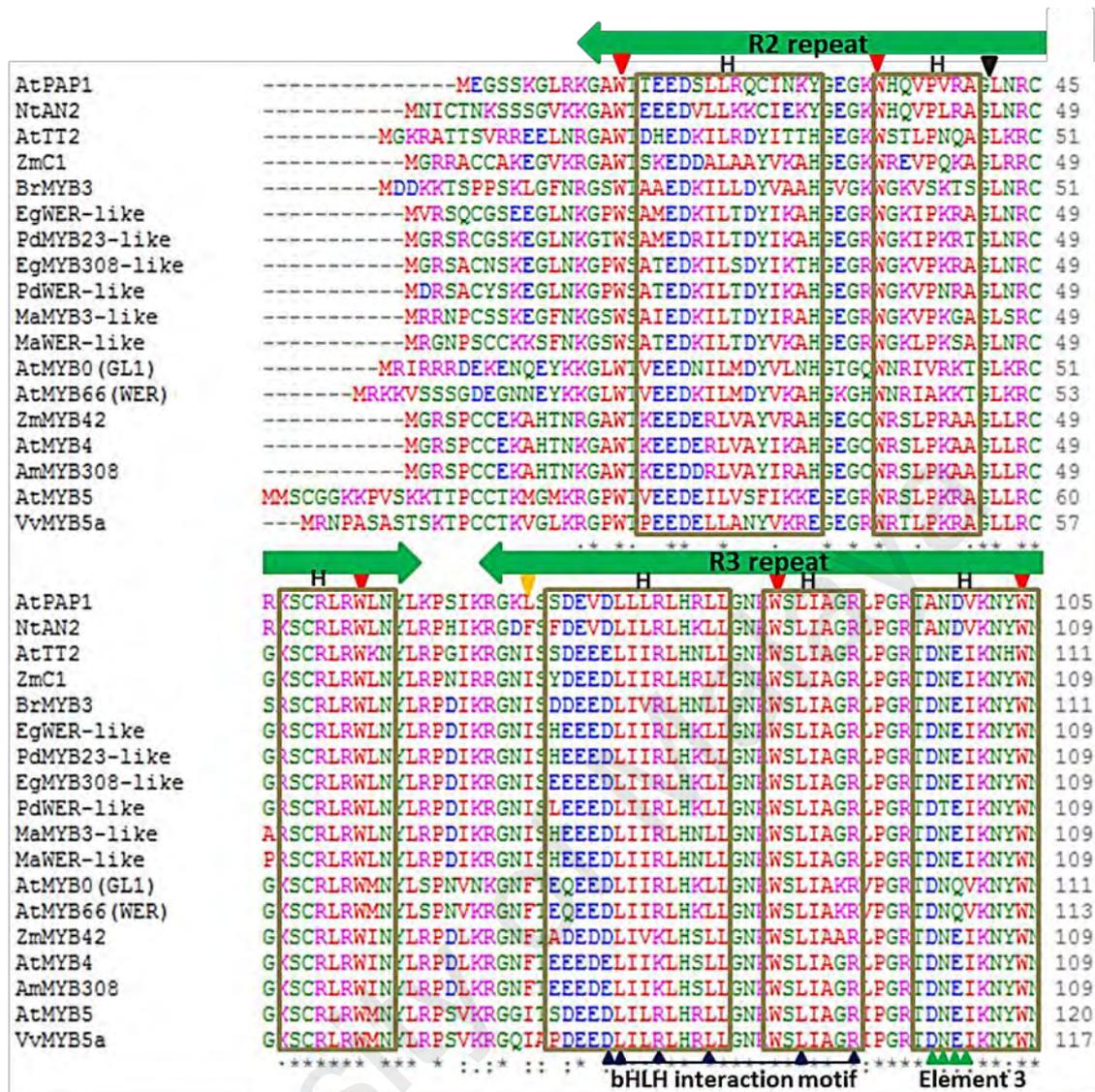


Figure 4.36: Putative BrMYB3 protein multiple sequence alignment with other R2R3 proteins from other plant species that harbouring bHLH interaction motif which include proteins from subgroup 4, 5, 6 15 and 27.

The black triangle denotes the intron location in the BrMYB3 protein sequence. The R2 and R3 repeat MYB DNA binding region is indicated by a green arrow, while C1, C2, C3 and C4 motif are indicated by the red boxes. Zf motif was indicated by orange box with cysteine residue (C) is denoted by orange triangle. Subgroup 5, 6 and 15 specific motifs were indicated by light blue box, green box and dark blue box respectively. The bHLH interaction motif was shown by a blue triangle while the complete interaction signature was underlined. The element 3 was highlighted with green triangle. The helices (H) are indicated with brown boxes. Conserved tryptophan amino acids in the MYB domain were labeled with red triangle, while yellow triangle indicates the tryptophan residue that replaced by phenylalanine, leucine and isoleucine residue in R3 repeat. The accession numbers of the R2R3 MYB proteins from other plants to construct the multiple sequence alignment are shown in the Appendix J.

AtPAP1	THLSKKHEPCCIKMKKRD-----ITPIPTPALKNNV	KPRP-----R--SFVNNDC	152
NtAN2	SHLRKKLIAPHDQKE-----SKQKAKKITH	RPRP-----R--TFSKTNTC	148
AtTI2	SNLRKRLPKIQTKQP---KRIKHSTN-----NENN	SG6 motif VCVRITKAIRC	149
ZmC1	STLGRRAGAGAGAGG---SWVVVAPDTGS---HAT	-----PAATSGAC	146
BrMYB3	TVLKKKKQLQVKYSGTSTGMICKSSSTC-----TSESNP	-----SREYYGCGKEC	156
EgWER-like	TVLKKKVQARSVPLI---NLKCLNHDEKERKSKS	SKTEV-----QAPPFENPARR	155
PdMYB23-like	TVLKKAQVRSVPLI---NLMCSNQDE-ERKSKSKTRV	-----RAPPFENPARC	154
EgMYB308-like	AVLKKKQAGSVPLI---KLVCLNKDE-EKSKNKTSV	-----QAPF-ENPGA-	152
PdWER-like	TVLKKKVQARSVPLI---NLVCLNQDE-ERKSKNKTSV	-----QAPSFENPTQS	154
MaMYB3-like	TVLKKKVQGSV-----DMHSGEDE-HMNGKSQETT	-----RCPKFGNSTGR	150
MaWER-like	TVLKKKVKACSVGSK---KIMHSDKDE-DRSSKSQEKV	-----ICPEFENSTSC	154
AtMYB0 (GL1)	THLSKKLVGDYSSAVKTTG-----EDDDSPPS	-----LFITAATPSSC	149
AtMYB66 (WER)	THLSKKLGIKDOKTKOSNG-----D-----	IVYQINLPNPT	144
ZmMYB42	THIRRLGSGIDPVTHRR-----VLGGAATTI	-----SFQSPSP-N	144
AtMYB4	THIRRLINRGIDPITHRP-----IYESSASQDSKPTQLEPVT	SNTINISFTSAP---K	160
AmMYB308	THIRRLLSRGIDPITHRS-----INDGTASQDQVTT	-----ISFSNAN---S	149
AtMYB5	THLRKLLRQGDIPQTHKP-----LDANNIHKPEEEVS	---GG---QKYPLEP	162
VvMYB5a	THLSKLLISQGDIPRTHKP-----LIPKPNPSPDVNAP	---VS---KSIPNANPNPS	163
:	:	:	:
	C1 motif		

AtPAP1	NHLNAPPKVD-----VNPPCLGLNINNVCDNSI	180	
NtAN2	VKSNITNTVDK-----DIEG---SSEI	166	
AtTI2	SKILLFSDLSQLKKSSTSP-LPLKEQEMDQGGSSIMGDL	EFDFRIHSEF---HF	200
ZmC1	ETGQNSAAHRADP-----DSAG-TITTS-A-----	AAVWAPKAV-----RCT	181
BrMYB3	IEAAVQLNGE-----DDRLLD-ASSSLSIDDRDVF	EQQLASFETEGLPQRSI	201
EgWER-like	SEAEISSNHYGNQNVHKNE-KAIMIDDLAP-ESSSLAP	QDYDLWNLLAGFNTDSL---CL	210
PdMYB23-like	SEAEISLDRHGNQTVHKNELKALVMDDMAP-ESSLLAP	QDHDLLNLVGTGNTDGLFRPCL	213
EgMYB308-like	-----NQEEGNQNVHKTGLESLVIEDLAS-DGSL	SPQDYDLLNAWTGLDTEFLSKPCL	205
PdWER-like	SEAEVSLNQNGNNAHKNGLSLVIEDLAP-DSFLLAP	QDYDLLNLRGTGFIEGLSQPSL	213
MaMYB3-like	NDPPSFPQHGKN---KFGESLVDGDFFS-ETSF	LAPEDCDFNPLMDSNHEELCPRH	204
MaWER-like	NDASVSSQHDKNICKKESKKS SVVSGEHL-DIY	FLGPEDCD---PVMASVIEESCTKHP	210
AtMYB0 (GL1)	HHQQENI---YENIAKSFNG---VVSASY	-----EDK	175
AtMYB66 (WER)	ETSEETK---ISNIVDNNN---ILGDE	-----	165
ZmMYB42	SAAAA---AAET---AAQAPI	-----KAEETAAVKA---PRC	173
AtMYB4	VEIFHES---ISFPGKSEK---ISMLIF	-----KEEKDECFVQ---EKF	195
AmMYB308	KE-----EDTKH---KVAVDI	-----MIKEENSPVQ---ERC	175
AtMYB5	-----IS---SS---HTD	-----	169
VvMYB5a	SSRVGE---IG---SN---HEV	-----KEIESN-ENH---KEP	188

AtPAP1	IYNKDKKKDQLVNNLIDGDNMWLEKFLEESQEV-DILV	PEATTTEKGDTLAFDVDQLW--	237
NtAN2	IRFNDNL-KPTTEELTDDGIQWADLLANNYNNGIEEADN	---SSPTLLHEEMPLL--	219
AtTI2	PDLMDFDGLD-----CGNVTSLVSS	220	
ZmC1	GGLFFFHRDITPAH-----AGETATPMAG	205	
BrMYB3	FDLEFFEQDA-----G	212	
EgWER-like	LDFEFVELLG-----DNIQQDI	227	
PdMYB23-like	LDFELFQLVG-----DNIQQDI	230	
EgMYB308-like	LDFEFFQLG-----GNIQQDM	222	
PdWER-like	PDSQFFQLG-----GNIQQDM	230	
MaMYB3-like	LDLNFFEIFE-----SNLPKEH	221	
MaWER-like	LDIDFSEIFG-----DNIQKDM	227	
AtMYB0 (GL1)	PKQELAQKDVLMATTNDPSHYG-----NNALVH	DDD-FELSS-----LT	215
AtMYB66 (WER)	-----IQEDHQGSN-Y-----LSSLVH	EDE-FELST-----LT-N	193
ZmMYB42	PDLNLDLCISPPCQ-H-----EDDGE	EEDEELDLKPAFVKRE-ALQ-A	213
AtMYB4	PDLNLELRISLPDD-V-----DR	-----LQ-G	215
AmMYB308	PDLNLDLKISPPCQ-Q-----QINYHQE	-----N-LKT-G	202
AtMYB5	-----DI-----TVNGGDG	-----DSKNSINV	187
VvMYB5a	P--NLDQYHSPLADS-----NENQ	SADG-----LVTGLQSTH-G	221

Figure 4.36, continued

AtPAP1	-----S-----LFDGETVKFD-----	248
NtAN2	-----S-----	220
AtTI2	NEI-----LG--ELVPAQGNLDLNRPF--TSCHHRGI DEDWLRDFT -----	258
ZmC1	GGGGGGGEAGSSDDCSSAASVSLRVGSHDEPCFSGDGDGDWMD-----DVRALAS	255
BrMYB3	-----IWGGA--DS-----HVFYFGDLFM-----	229
EgWER-like	-----EDCNN-G--KLCAFFDENIFCSRGMFENGIGTDQIQSYGSLDLGSLAT	272
PdMYB23-like	-----GDRNN-G--KLCVFPDENVISSGGMFENWIGTDHIQPYGSLDLGSLAT	275
EgMYB308-like	-----RDCDN-N--KLSVSSDDNVS-CGRVFENWVGTDHIQPYGSLDLGSLAT	266
PdWER-like	-----RDCDN-G--KLSASFDENVFSSRQVFENWVGTDHIQPYGSLDLGSLAA	275
MaMYB3-like	-----ETLWNV-----	227
MaWER-like	-----R-IWGIV--DAPEPPDDIMLFSNGISDSWFNGDDIQLYGGINLESFTT	272
AtMYB0 (GL1)	MM--NFASGDVEYCL-----	228
AtMYB66 (WER)	MM--DFIDGHCF-----	203
ZmMYB42	GH--GHGHGICLGCGLGG-----QKGAAGCSC-----S	239
AtMYB4	HG--KSTTPHCFKCSLGM-----INGM-ECRCGRMRCVV-----GGSSKGSMDMS	257
AmMYB308	GR--NGSSTILCFVCRIGI-----ONSK-DCSCSDGVGN-----	232
AtMYB5	GGEHGYEDFGFCYDDKFSS-----FLN--SLINDVSDPF-----GNIIPISQPL	229
VvMYB5a	TSNDEDDDIGFCNDDTFPS-----FLN--SLINED--VF-----GNHNHHQQQQ	261
	C3 motif	
AtPAP1	-----	248
NtAN2	-----	220
AtTI2	----- SG5 motif -----	258
ZmC1	-----FLES DEDWLR QTAGQLA-----	273
BrMYB3	-----	229
EgWER-like	-----FFESSES-----	278
PdMYB23-like	-----FFESSES-----	281
EgMYB308-like	-----FFESEPP-----	272
PdWER-like	-----FFESSES-----	281
MaMYB3-like	-----	227
MaWER-like	-----LFPSESEFLGL-----	283
AtMYB0 (GL1)	-----	228
AtMYB66 (WER)	-----	203
ZmMYB42	NG-----HHFLGLR--TSV-LDFRG-----LEMR	260
AtMYB4	NG-----FFELGLAKKETISL-LGFRS-----LEMR	282
AmMYB308	-----	232
AtMYB5	QMDDCK--DGIVG-----ASSSSLGHD-----	249
VvMYB5a	QQQQLQQVQQPSNVIAPLPHPAISVQATFSSSPRIWPEPAALTSTSAPLVHDQKHSMSPP	320
	C4 motif	

Figure 4.36, continued

Characterization of putative *BrMYB2* and *BrMYB3* is summarized in Figure 4.37. The splicing site of the intron is located within the R3 domain region for *BrMYB2*. In contrast, the splicing site of the intron for *BrMYB3* is located within the R2 domain region.

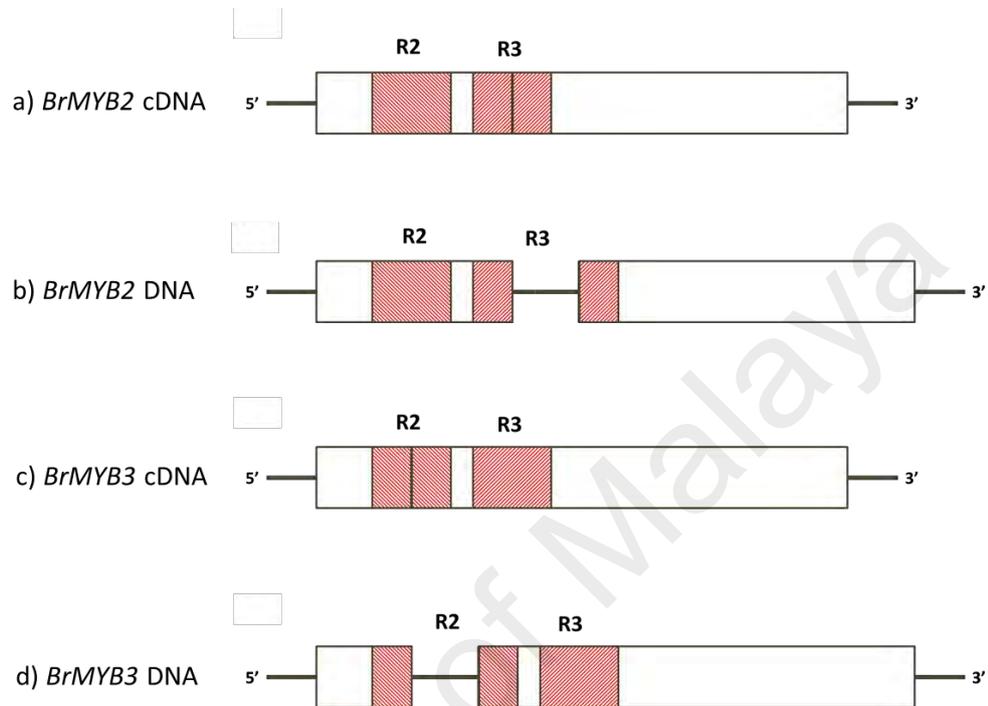


Figure 4.37: Schematic diagram of putative *BrMYB2* and *BrMYB3* genes in *B. rotunda*.

a) *BrMYB2* with an intact two R2R3 repeat domains. b) *BrMYB2* with an intron that is located within R3 domain. c) *BrMYB3* with an intact two R2R3 repeat domains. d) *BrMYB3* with an intron that is located within R2 domain. Exon region is indicated by solid box, while introns are indicated by line. The R2 and R3 repeats are indicated by red diagonal boxes.

4.9 Gene expression study of *BrMYB2*

Among the putative *BrMYB1*, *BrMYB2* and *BrMYB3*, *BrMYB2* were chosen to further analyse their expression in response to phenylalanine treatment in *B. rotunda* callus.

4.9.1 Total RNA extraction

Total RNA from cell suspension culture after 1, 7 14 days phenylalanine treatment were successfully extracted. Figure 4.38 shows the intact 26S and 18S rRNA bands, indicating good quality RNA extraction. The concentration and purity of the RNA are shown in Table 4.23.

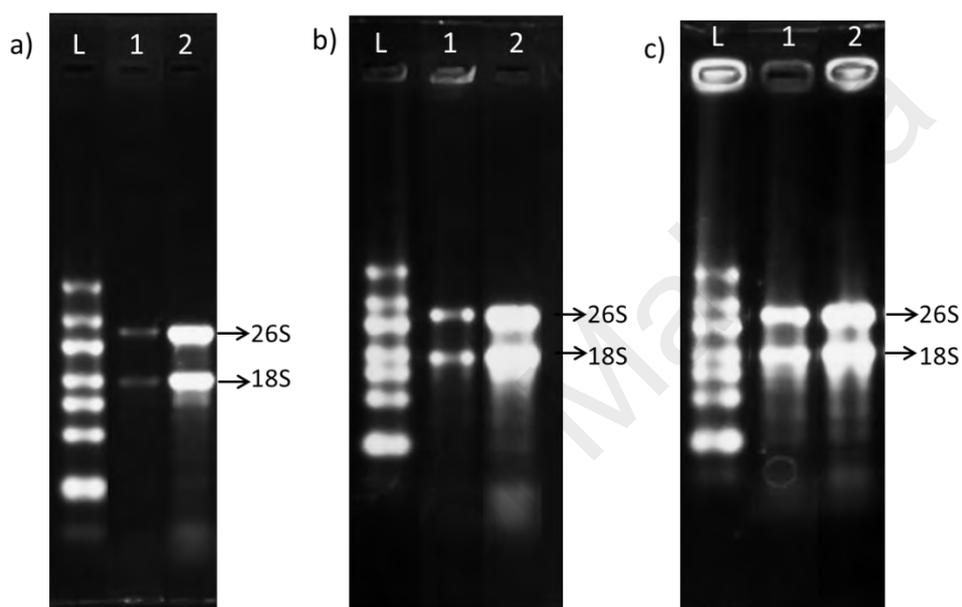


Figure 4.38: Total RNA extraction of control and phenylalanine treated *B. rotunda* callus after a) 1 day; b) 7 days and c) 14 days treatment. Lane L: RiboRuler High Range RNA ladder (Thermo Scientific, USA); lane 1: total RNA extracted from control callus and lane 2: total RNA extracted from phenylalanine treated callus.

Table 4.23: Concentration and purity of total RNA from control and phenylalanine treated samples after 1, 7 and 14 days post-phenylalanine treatment. Abbreviations: CTR; control and TR; phenylalanine treated sample.

Sample	Concentration, ng/ μ l	$A_{260/A280}$
1 CTR	330.4	2.00
1 TR	528.1	2.01
7 CTR	279.6	1.97
7 TR	646.2	2.06
14 CTR	349.3	2.03
14 TR	511.3	2.05

4.9.2 qPCR analysis

Gene regulation of putative *BrMYB2* in response to phenylalanine treatment was further analysed using qPCR. Three pairs of primers (elongation factor 1- α , β -tubulin and *BrMYB2*) that designed for the relative quantification were examined with dissociation program for their specificity. The results are shown in Figure A.15, Appendix M.2. From the dissociation curves, these three pair of primers showed only one single peak, which indicated that a very specific amplification of elongation factor 1- α , β -tubulin and *BrMYB2* for qPCR analysis. Endogenous control profile of elongation factor 1- α and β -tubulin in control and phenylalanine treated samples after 1, 7 and 14 days were shown in Figure A.16 in Appendix M.3. Both of the reference genes show a similar expression profile.

The generated data from the relative quantification was analysed with Quant Studio software. The results were summarized in Table A.32 in Appendix M.1. The expression patterns of putative *BrMYB2* transcription factor in cell suspension culture was monitored after 1, 7 and 14 days phenylalanine treatment. Based on bar chart in Figure 4.39, the mRNA level expression fold change of *BrMYB2* in 1, 7 and 14 days post phenylalanine treated samples were 1.478, 2.298 and 2.599, respectively. Whereas, the mRNA level expression fold change of *BrMYB2* in 1, 7 and 14 days control samples were 1, 1.858 and 2.013 respectively. These results correlates with the transcriptome data. Putative *BrMYB2* expression level in *B. rotunda* callus was increased in response to phenylalanine treatment as compared to the control sample.

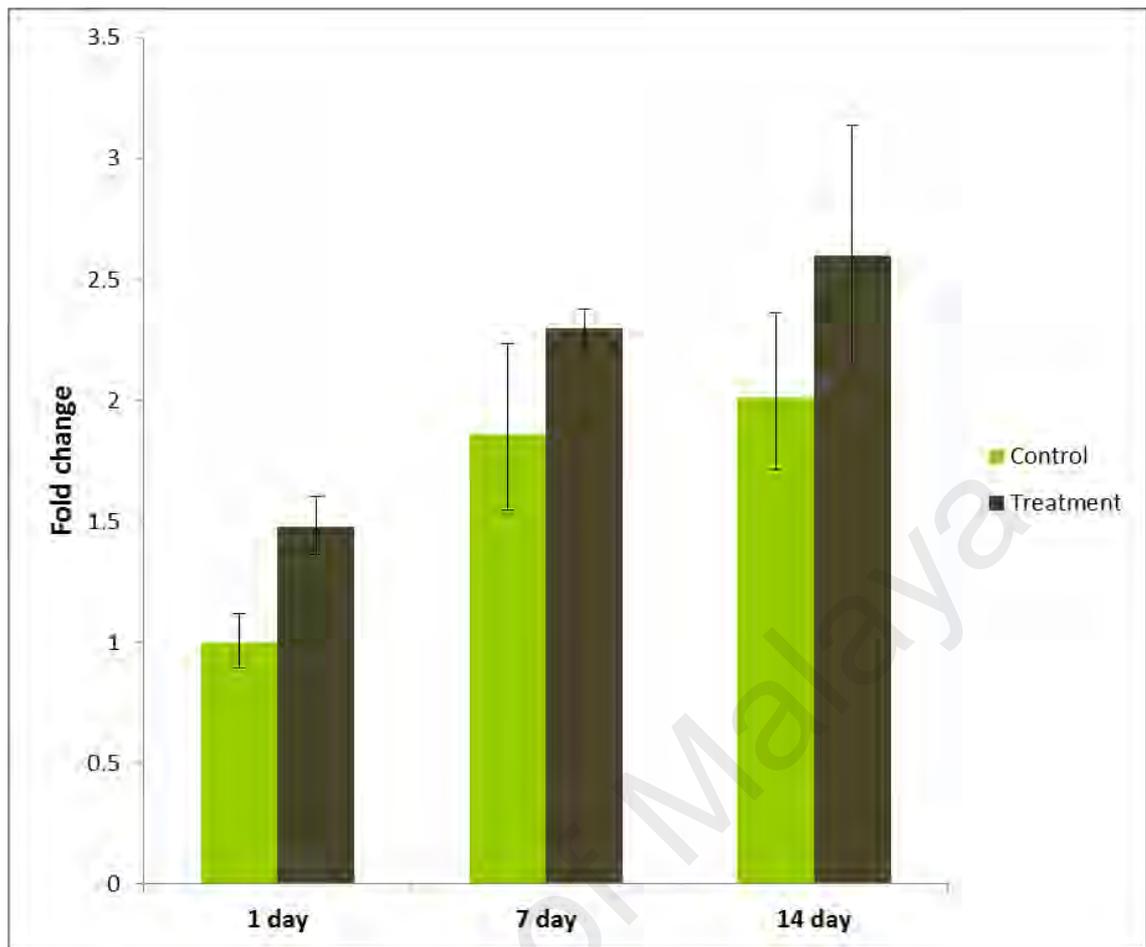


Figure 4.39: Expression pattern of putative *BrMYB2* in *B. rotunda* cell suspension culture after 1, 7 and 14 days post treated with phenylalanine. The fold changes were determined by qPCR using $2^{-\Delta\Delta CT}$ method. Results represent mean standard deviations (\pm SD) of four experimental replicates.

CHAPTER 5: DISCUSSION

5.1 General transcriptome analysis

In general, flavonoid, terpenoid and alkaloid pathways are the main pathways producing secondary metabolites in plants. RNA-Seq together with DEGs data has been used to identify genes that are directly or indirectly involved in the secondary metabolite biosynthetic pathways of target bioactive compounds especially in medicinal plants. For instance, cytochrome P450 (CYP450) had been identified in most reported analysis of late terpenoid pathways. The combination of RNA-Seq technology and methyl jasmonate induction experiments successfully identified one CYP450 and four glycosyltransferases as key enzymes in the ginsenoside biosynthesis in *Panax quinquefolius* (Sun et al., 2010). Subsequently, by combining RNA-Seq technology and phylogenetic tree analysis based on the previously identified CYP450 and glycosyltransferase in *Panax quinquefolius*, two CYP450 and one UDP glycosyltransferase were also elucidated as candidates for ginsenoside biosynthesis in *Panax notoginseng* (Luo et al., 2011; Sun et al., 2010). Additionally, seven CYP450 and five glycosyltransferases were identified in mogrosides biosynthesis in *Siraitia grosvenorii*; and six CYP450 and one glucosidase identified in camptothecin biosynthesis in *Camptotheca acuminata* (Sun et al., 2011; Tang et al., 2011). RNA-Seq analysis from different rhizomes of cultivated *Curcuma longa* cultivars in India also described transcripts potentially related to anticancer and antimalarial terpenoids (Annadurai et al., 2013).

The same approach has been adapted in this study aiming to decipher phenylpropanoid and flavonoid pathways that are involved in panduratin A biosynthesis. Prior to RNA-Seq, the cell suspension cultures of *B. rotunda* were treated with phenylalanine based on a previous report by Tan et al. (2012) where they found that the production of panduratin A increased after the cell suspension was treated with phenylalanine. Phenylalanine is a key substrate in the phenylpropanoid pathway. The aromatic amino acid that is produced

from the shikimate pathway provides the essential 6-carbon ring and 3-carbon side chain to all phenylpropanoid and flavonoids in plants (Herrmann & Weaver, 1999). Other than enhancing the production of panduratin A, phenylalanine has also been reported to stimulate taxol production in *Taxus cupidata* cell suspension culture (Fett-Neto et al., 1993; Fett-Neto et al., 1994).

To ensure the reliability of the transcriptome data the RNA-Seq result i.e the transcriptome data was validated through qPCR analysis. Validation was done using a random selection of unigenes and included some of the unigenes that were annotated from the flavonoid pathway. This included PAL; Unigen10327_All, C4H; Unigene67845_All, 4CL; Unigene41852_All, CHS; Unigene1735_All and F3H; Unigene49558_All (Figure 4.19). As expected the qPCR results correlated with the transcriptome data.

Out of 101,043 assembled unigenes, only 50.41% of the unigenes were successfully annotated in public protein database (Table 4.3). The limited numbers of identified plant genes that were deposited in the database may be the reason for the unannotated 49.59% of the transcriptome unigenes. These unannotated unigenes should be further identified to enrich public plant databases.

Subsequently, differential expression patterns of the transcriptome profiles between control and phenylalanine treated *B. rotunda* cell suspension culture showed 14,644 significantly up-regulated and 14,379 significantly down-regulated unigenes (Table 4.4). The key factor of differential gene expression in *B. rotunda* cell suspension culture between control and phenylalanine treated is proposed to be related to the differential expression of transcription factors and transcription regulators. Transcription factors are regulatory proteins that control the expression of specific groups of genes through sequence-specific DNA binding and protein-protein interactions. They act either as activators or repressors of gene expression, mediating either an increase or decrease in

the accumulation of mRNA depending on tissue type or in response to internal or external signals (Broun, 2004; Vom Endt et al., 2002).

A total of 139 unigenes from *B. rotunda* transcriptome data were mapped to rice transcription factor database with 8 unigenes showed up-regulation and 26 showed down-regulation (Table 4.12). Whereas 46 unigenes mapped to the rice transcription regulator database with 4 and 9 unigenes showing up- and down-regulation, respectively (Table 4.13). Further discussion on transcription factors can be found in section 5.2.

The most abundant up-regulated unigenes in the phenylpropanoid pathway were peroxidases, comprising 40 out of 90 unigenes (Table 4.7). Class III plant peroxidases catalyze plant-specific oxidoreduction between hydrogen peroxide (H_2O_2) and various reductants (Hiraga et al., 2001). Differential expression profile of the peroxidase as isoenzymes in *B. rotunda* suggests that they might be involved in catalyzing different substrate and may be involved in different physiological processes. Peroxidase class III is involved in lignification in higher plants by radical coupling of monolignols (Marjamaa et al., 2009). This oxidoreduction reaction utilizes hydrogen peroxide (H_2O_2) for oxidative power to produce monolignol radicals for lignin polymerization (Marjamaa et al., 2009). Different monolignols in this reaction produce different types of lignin and thus provide different resistance barriers for plants. Lignin provides mechanical strength and resistance against pathogens in plants. Lignification is a normal process for plant growth and development and also occurs in response to environmental stresses (Kim et al., 2008). Additionally, some peroxidase isoenzymes are regulated upon environmental stimuli or prior attack by pathogens, which renders the plant with a self-defense mechanism against physical, chemical and biological stresses (Hiraga et al., 2000).

As for the primary metabolic processes such as carbohydrate metabolism, energy metabolism and amino acid metabolism, most have more down-regulated unigenes

compared to up-regulated unigenes in the same pathway (Table 4.6). Primary metabolism is essential for plant growth, plant development and plant reproduction. In cell suspension cultures, primary metabolism is essential for plant cells to propagate in liquid media. Down-regulation of unigenes in the primary metabolic pathways after 14 days of propagation might be due to depleted nutrients in the liquid media. It can be suggested that by depleting plant nutrients, cell suspension cultures are stressed and eventually induce secondary metabolites. Similar observations have been reported by Lattanzio et al. (2009) who showed that under limited nutrient conditions, increased phenolic compounds were observed with a decrease in biomass production.

There are several hypotheses that relate the carbon limiting step in primary metabolism to secondary metabolite production in plants as a trade-off between growth and the production of carbon-based secondary metabolites such as phenolic compounds (Herms & Mattson, 1992; Lerdau & Coley, 2002). One of the hypotheses is the carbon nutrient hypothesis which suggests that plants modify the allocation of carbon skeletons between primary and secondary metabolism, where in a nutrient depletion situation, the plant restricts growth and the carbon skeleton is allocated to produce phenolic secondary metabolite compounds (Lerdau & Coley, 2002). In addition, the protein competition model of phenolic allocation by Jones and Hartley (1999) suggested that protein and phenolic synthesis were competing for the use of phenylalanine as a precursor (Jones & Hartley, 1999). The theories correlates with the previous findings. Panduratin A production was increased after 14 days of treatment with phenylalanine in *B. rotunda* cell suspension cultures (Tan, 2005). It can be suggested that during nutrient depletion, the exogenous phenylalanine supplied were metabolized and synthesized bioactive compounds in plants.

5.2 Transcription factors

Genes encoding regulatory proteins are important in controlling the yield of the bioactive compounds in *B. rotunda*. Through homology search of TRs and TFs using rice database, transcription regulators and transcription factors present in *B. rotunda* were successfully identified. As mentioned earlier, there were in total 139 transcription factors and 46 transcription regulators found in *B. rotunda* in this study. The most abundant transcription factor found was classified under the C3H family. The second most abundant transcription factor found was MYB followed by NAC, WRKY, bZIP and AP2-EREBP family (Table 4.12). All of these transcription factor families except for C3H showed significant differential expression of their members in the treated sample, suggesting that they played an important role in the induction or repression of the panduratin A biosynthesis pathway.

Other up-regulated transcription factors such as AP2-EREBP, WRKY, bZIP, GRAS and NAC have been reported to modulate the genes for plant growth and plant response to biotic or abiotic stresses (Hirsch & Oldroyd, 2009; Jakoby et al., 2002; Mizoi et al., 2012; Olsen et al., 2005; Ülker & Somssich, 2004). This information would be useful for future analysis on genes that are regulated by these transcription factors especially in relation to phenylpropanoid and flavonoid pathways. The genes encoding transcription factors such as MYB, bHLH and WDR were reported to play important roles in regulating the complex flavonoid pathway in plants (Yang et al., 2012).

In the *B. rotunda* transcriptome data, there were 16 unigenes that were mapped to MYB transcription factor in the rice TFs database (Table 4.12). From the results, one unigene was up-regulation, while four unigenes were down-regulation in response to phenylalanine treatment (Table 4.12). The up-regulated unigene was identified as Unigene14544_All. These results only show the unigenes that mapped to the rice TFs

database. Therefore, a new approach was performed in order to recover more MYB transcription factors in *B. rotunda* RNA-Seq data.

It was reported that plant MYB transcription factor family has four subfamilies (Dubos et al., 2010). One of them is R2R3 MYB transcription factor. R2R3 MYB transcription factor is the largest MYB family in plants that are involved in secondary metabolic pathways. The R2R3 MYB family plays a major role in regulating sets of genes that are responsible for secondary metabolite biosynthetic pathways in plants especially for synthesizing flavonoids in the phenylpropanoid pathway (Vom Endt et al., 2002). Due to these reasons, R2R3 MYB transcriptome-wide analysis in *B. rotunda* was performed.

The structure of R2R3-MYB transcription factors consists of two parts: a highly conserved MYB domain at the N terminus and a conserved amino acid sequence motif-containing domain at the C-terminus (Li et al., 2015). In contrast to these conserved regions, the other regions of R2R3-MYBs are highly variable. The MYB domain, R2 and R3 repeats play key roles in DNA binding and mediating response specificity. It was reported that R2 and R3 repeats contain two helix-turn-helix motifs responsible for binding to target genes (Du et al., 2009).

In R2R3 MYB activators, the specificity to mediate either anthocyanin or non-anthocyanin pathway are determined by a conserved region named element 3 that is located within the third α -helix in the R3 repeat binding domain (Heppel et al., 2013). There were two signature patterns of element 3 that determine the specificity of regulation either in the anthocyanin or proanthocyanidin pathway. R2R3 MYB activator proteins that possess the signature of ANDV or SNDV pattern mediate the anthocyanin pathway, while the proanthocyanidin pathway is mediated by activators harbouring the DNEI motif (Heppel et al., 2013).

In plants, R2R3 MYB transcription factors of subgroups 4, 5, 6, 7 and 27 are found to be the major subgroups in controlling phenylpropanoid metabolism which are involved in phenolic acid, lignin, flavonol, proanthocyanidin (tannins), anthocyanin and flavonoid biosynthesis. Interestingly, with the exception of flavonol regulating subgroup 7 R2R3 MYB family, the other R2R3 MYB subgroups regulating the phenylpropanoid metabolism exhibit a bHLH interaction motif [D/E]Lx2[R/K]x3Lx6Lx3R in their R3 repeat region (Grotewold et al., 1994; Zimmermann et al., 2004). Additionally, subgroup 15 R2R3 MYB family that regulates trichome and root hair development also harbours the same signature motif. The presence of the [D/E]Lx2[R/K]x3Lx6Lx3R motif indicates the involvement of a bHLH transcription factor in the MYB transcriptional regulation either through a MYB-bHLH dimer or a MYB-bHLH-WD40 (MBW) ternary complex formation (Hichri et al., 2011). In addition to the bHLH interaction motif, these R2R3 MYB subgroup families were classified based on the additional subgroup-specific motifs that are located at the C-terminal region (Kranz et al., 1998; Stracke et al., 2001).

Based on R2R3 transcriptome-wide analysis, there were 145 unigenes annotated as R2R3 MYB transcription factors in the *B. rotunda* transcriptome data (Table 4.14). However, there were only 13 unigenes that exhibited both R2 and R3 repeat domains with a complete ORF (Table 4.14). The phylogenetic tree analysis of these 13 R2R3 MYBs using ORF reads resulted in the classification of the MYBs into seven R2R3 MYB subgroups which include abiotic/biotic stress, transcriptional repressor, anthocyanin biosynthesis, lignin biosynthesis, plant development, cell wall thickening and an unknown clade (Figure 4.22). Among these, R2R3 MYB proteins from subgroups 4 and 6 were reported to be involved in regulating the phenylpropanoid metabolism that leads to lignin, anthocyanin, proanthocyanidin, flavonol and flavonoid biosynthesis.

Out of the five unigenes classified in subgroup 4 R2R3 MYB, only the expression of Unigene51701_All was found to be up-regulated upon phenylalanine treatment. Since the unigene might be involved in flavonoid biosynthesis in *B. rotunda*, further molecular cloning and characterization was carried out. The same analysis was also conducted for Unigene14544_All. As mentioned earlier, Unigene14544_All was the only up-regulated unigene that mapped to MYB in rice TFs database. It was the reason of including the Unigene14544_All for further analysis.

Since there is no genome sequence data for *B. rotunda* available, the cloning were done using both DNA and cDNA templates in order to study the location of intron and exon. The cDNA transcripts from Unigene14544_All and Unigene51701_All were named as *BrMYB1* and *BrMYB2*, respectively. Unlike *BrMYB2*, *BrMYB1* was classified in 1R-MYB subfamily. Instead of having two MYB repeat domains, *BrMYB1* only consisted of one MYB domain (Figure 4.30).

There were five major subgroups in the 1R-MYB subfamily including CCA1-like/R-R, CPC-like, TRF-like, TBP-like and I-box-like (Du et al., 2013). The MYB domains of angiosperm CPC-like proteins are also known as R3-type MYB proteins because it contained similar single R3 repeats of 2R-MYB proteins (Du et al., 2013). Based on the protein sequence, *BrMYB1* can be further classified as an R3-type MYB protein. In general, R3-MYBs can be recognized by their short sequence (<120 amino acids length) and the bHLH interaction conserved motif [D/E]Lx2[R/K]x3Lx6Lx3R in R3 repeats (Wang & Chen, 2014; Zimmermann et al., 2004). R3 MYBs also contains a sequence motif WxM that is important for its cell-to-cell movement (Kurata et al., 2005).

In *Arabidopsis* and tomato, R3 MYB proteins are involved in the trichome and root hair development (Schellmann et al., 2007; Schellmann et al., 2002; Tominaga-Wada et al., 2013a; Wang & Chen, 2014; Wang et al., 2008). R3 MYB are also reported to be

involved in the regulation of flower development and stomatal formation (Serna, 2008; Tominaga-Wada et al., 2013b; Tominaga et al., 2008). In addition, R3-type MYB proteins are involved in anthocyanin accumulation. For instance, CPC and SITRY were reported to regulate the anthocyanin pathway in Arabidopsis and tomato, respectively (Tominaga-Wada et al., 2013c; Zhu et al., 2009); while MYBL2 was found to regulate the flavonoid biosynthetic pathway in Arabidopsis (Dubos et al., 2008).

As BrMYB1 is classified as a R3-type MYB protein, further characterization was done using phylogenetic tree using known single repeat R3 MYB proteins and other MYB-related proteins such as AtCPC, AtMYBL2, AtTRY, OsMYBAS1-1 and MaMYBAS1 (Figure 4.29). BrMYB1 was clustered together with MYBAS proteins from other plant species such as *Elaeis guineensis*, *Musa acuminata* subsp. *malaccensis*, *Zea mays*, *Oryza sativa* and *Medicago truncatula*. MYBAS is an alternatively spliced MYB transcription factor. Unlike other well characterized R3 MYB proteins such as AtCPC, AtMYBL2 and AtTRY, there was limited knowledge on the function of MYBAS proteins in plants. In addition, only a few *MYBSA* genes were reported and molecularly characterized. The MYBAS protein may be present as a R2R3 MYB protein, consisting of intact R2 and R3 MYB domain or single repeat R3 MYB protein, resulting from different alternative splicing events in plants. For instance, SoMYBAS1 variant 3, a stress related MYB transcription factor from sugarcane (*Saccharum officinarum*) was demonstrated to be regulated in response to water deficit and salt stress (Geethalakshmi et al., 2015; Prabu et al., 2011). *SoMYBAS1* variant 1 (Gen Bank: HM136780) and 3 (Gen Bank: HM136782) are present as R2R3 MYB proteins, while *SoMYBAS1* variant 2 (Gen Bank: HM136781) is present as a single repeat R3 MYB protein. At the moment, only SoMYBAS1 from variant 3 has been functionally characterized, but not so for the alternatively spliced variant 2 SoMYBAS1.

Similarly, two homologous genes in rice, *OsMYBAS1* (Gen Bank: AK111626) and *OsMYBAS2* (Gen Bank: AK107214) were found to have three distinctively spliced transcripts (type 1, 2 and 3 transcripts), resulting from three alternative splicing events (Li et al., 2006). Through alternative splicing, *OsMYBAS1* and *OsMYBAS2* were able to encode MYB proteins with one or two MYB repeats. The type 3 transcripts produce R2R3 MYB proteins (*OsMYBAS1-3* and *OsMYBAS2-3*), while type 1 and type 2 transcripts produce R3 MYB proteins (*OsMYBAS1-1*, *OsMYBAS1-2*, *OsMYBAS2-1* and *OsMYBAS2-2*). However, no functional studies have been carried out for all of these proteins. Interestingly, unlike *SoMYBAS1*, *OsMYBAS1* and *OsMYBAS2*, *BrMYB1* was not produced from alternative splicing event. Based on the *BrMYB1* gene that was sequenced from the DNA template which was then compared to the *BrMYB1* from the cDNA template, no intron was detected (Figure 4.28).

Protein multiple sequence alignment of *BrMYB1* (Figure 4.30) showed conserved R3 repeats across the MYBAS proteins, except for the first alpha-helix in *BrMYB1*. Compared to the first alpha-helix, the second and third helices in MYB domain appear to be crucial to form helix-turn-helix structure (HTH) for target gene reorganization and binding. The second helix or recognition helix in HTH structure is essential to recognize and bind to specific DNA sequences by forming hydrogen bonds with bases located in the major groove of the DNA double helix, while the first helix stabilizes the overall configuration through hydrogen interactions with the recognition helix (Becker et al., 2003; Ogata et al., 1992). Three regularly spaced tryptophan residues located in the MYB repeats, forming a tryptophan cluster in three-dimensional HTH structure (Ogata et al., 1992). In plants, the first tryptophan of R3 is replaced by phenylalanine or isoleucine (Ambawat et al., 2013). However, in MYBAS proteins, the first tryptophan residue is replaced by methionine (Figure 4.30), which also happens to be the start codon, while in

BrMYB1 tryptophan was replaced by aspartate with a start codon towards the end upstream (4 amino acids away from other plants MYBAS proteins).

In MYB proteins, nuclear localization signals (NLSs) in the R3 repeat was essential for nuclear localization. There were two basic amino acid regions of NLSs in the R3 repeat MYBs; KRGK and RKKAQEKKR (Li et al., 2006). In Arabidopsis, *AtMYB59* transcript produces four types of proteins through alternative splicing event (*AtMYB59-1*, *AtMYB59-2*, *AtMYB59-3* and *AtMYB59-4*). *AtMYB59-2* (R3 MYB) and *AtMYB59-3* (R2R3 MYB) proteins harbouring both NLS1 and NLS2 regions showing a clear localization in the nucleus. In contrast, partial nucleus localization was demonstrated in *AtMYB59-1* (R3 MYB) and *AtMYB59-4* (R2 MYB) which harbour only NLS2 or NLS1, respectively. Based on the protein multiple sequence alignment, BrMYB1 consists of only NLS2 and lack of NLS1. Therefore it can be speculated that BrMYB1 protein would be at least partially localized in the nucleus. However, further experiments will need to be conducted for confirmation. Additionally there are two motif regions; Motif 1 and Motif 2 detected in the BrMYB1, which are also present in other plant MYBAS proteins (Figure 4.30). However, the functions of these motifs in plants are still unknown. In the near future functional studies of BrMYB1 in *B. rotunda* will be carried out.

Unlike BrMYB1, BrMYB2 was classified in 2R-MYB subfamily. Through blastx homology search, the conserved domain of BrMYB2 shows two MYB domains (Appendix L.2). Hence, further characterization was done by constructing a phylogenetic tree and protein multiple sequence alignment using C2 transcriptional repressors R2R3 MYB from other plant species. Phylogenetic tree analysis was performed to identify the most closely related C2 repressor proteins within the subgroup 4 R2R3 MYB family (Figure 4.32), while protein sequence alignment was done to analyse the presence of repressor motifs in BrMYB2 (Figure 4.33).

R2R3 MYB subgroup 4 transcription factors act as transcriptional repressor in the monolignol or lignin biosynthetic pathway. These proteins exhibit a transcriptional repressor motif; pdLNL[D/E]Lxi[G/S] at their C-terminal region. EgMYB1 from eucalyptus represses the lignin biosynthetic pathway (Legay et al., 2007). Furthermore, the ectopic expression of EgMYB1 in Arabidopsis and poplar negatively regulates secondary cell wall formation (Legay et al., 2010). Subgroup 4 R2R3 MYB repressors have also been demonstrated to repress the phenylpropanoid biosynthetic pathway. For instance, AtMYB4 represses sinapate ester biosynthesis in a UV-dependent manner (Jin et al., 2000). In *Salvia miltiorrhiza*, SmMYB39 represses rosmarinic acid biosynthetic pathway by down-regulating the expression of *C4H* and *TAT* (tyrosine aminotransferase) genes (Zhang et al., 2013). Additionally, ectopic expression of snapdragon AmMYB308 and AmMYB330 in tobacco inhibits hydroxycinnamic acid and monolignol accumulation (Tamagnone et al., 1998).

In addition to lignin and phenylpropanoid pathway, subgroup 4 MYB repressors were also reported to regulate the anthocyanin and proanthocyanidin biosynthetic pathway. For instance, TaMYB14 from *Trifolium arvense* activates proanthocyanidin biosynthesis in the legumes *Trifolium repens* and *Medicago sativa* (Hancock et al., 2012). On the other hand, strawberry FaMYB1 suppresses anthocyanin and quercetin accumulation in tobacco (Aharoni et al., 2001). Since the lignin and flavonoid pathways generally use the same common precursors especially 4-coumaroyl CoA, the repression of monolignol pathway can affect the flavonoid content in plants. For example, overexpression of wheat TaMYB4 in tobacco caused changes in metabolic flux from the lignin pathway to the flavonoid pathway in the (Ma et al., 2011). Similarly, overexpression of maize ZmMYB31 in Arabidopsis redirects the phenylpropanoid metabolic flux towards anthocyanin biosynthesis (Fornalé et al., 2010). Reduction of lignin biosynthesis is likely to provide more substrates to produce flavonoids. In contrast, heterologous expression of

chrysanthemum CmMYB1 in Arabidopsis alters lignin composition and represses flavonoid synthesis (Zhu et al., 2013).

The protein sequence alignment of BrMYB2 with other transcriptional repressor R2R3 MYB proteins revealed that BrMYB2 has a highly conserved R2 and R3 domain at the N-terminal region, while more divergent in sequence and length at the C-terminal region (Figure 4.33). In addition, the sequence alignment also revealed the presence of four protein motifs at the C-terminal region of the R2R3 MYB subgroup 4 such as C1 (LlSrGIDP_x[T/S]HR_x[I/L]) (Kranz et al., 1998), C2 (pdLNL[D/E]Lxi[G/S]) (Kranz et al., 1998), Zinc finger or Zf (CX₁₋₂CX₇₋₁₂CX₂C) and C4 (FLGLX₄₋₇[V/L]L[D/G][F/Y][R/S]X₁LEMK) (Shen et al., 2012) (Figure 4.33). These motifs were unique to subgroup 4 R2R3 MYB transcription factors. Additionally, the bHLH interaction motif, [D/E]LX₂[R/K]X₃LX₆LX₃R located within the R3 repeat was also identified.

Originally, subgroup 4 of the R2R3 MYB subfamily was defined by the presence of C1, C2 and Zf motifs (Stracke et al., 2001). Later, the C4 repressor motif identified by Shen et al. (2012) was included as a new motif to define subgroup 4 R2R3 MYB subfamily. However, according to Cavallini et al. (2015), some of the R2R3 MYB C2 repressors lacked either the Zf or the C4 motifs. Therefore, a more accurate MYB C2 repressor classification was established by including all MYB proteins that harbour a C2 repressor motif (Cavallini et al., 2015). The C2 repressor R2R3 MYB proteins can be divided into four groups (A, B, C and D groups) based on the presence of C1, C2, Zf and C4 motifs. In addition, group D can be further divided based on the presence of the C5 (TLLLFR) motif. Group A consists of C1, C2, Zf and C4 motifs; while group B lacks the C4 motif. Subsequently, group C consists of only C1 and C2 motifs. Group D was further

divided into three subgroups, in which D1 consists only C2 motif, D2 consists of C1, C2 and C5 motifs; and D3 consists of C1 and C2 motifs.

As demonstrated by Shen et al. (2012), the C1 motif which is also known as GIDP motif, functions as an activation domain whereas, C2 and C4 motifs function as repression domains. In contrast, the function of the Zf motif in the R2R3 MYB subgroup 4 proteins remains unknown. Based on protein sequence alignment of BrMYB2 with other R2R3 MYB repressor proteins, BrMYB2 consists of only C1 and C2 motifs; and lacked the Zf and C4 motifs (Figure 4.33). Therefore, BrMYB2 can be classified under group C of the C2 repressor proteins. However, BrMYB2 was not clustered together with group C repressor motifs such as AtMYB6 and AtMYB8 in the phylogenetic tree, but clustered with AmMYB330 from *Antirrhinum majus* by sharing the same clade (Figure 4.32). Unlike AtMYB6 and AtMYB8, AmMYB330 was classified under group B, which consists of C1, C2 and Zf motifs. Similarly, although AmMYB308 lacks the C4 motif, the protein was closely related to group A members and clustered into group A in the phylogenetic tree analysis (Figure 4.32). Both AmMYB308 and AmMYB330 act as transcriptional repressor of monolignol biosynthetic genes by repressing the general phenylpropanoid metabolism (4CL and C4H) and lignin branch (cinnamyl alcohol dehydrogenase, CAD) genes (Tamagnone et al., 1998).

The C2 motif (pdLNL[D/E]Lxi[G/S]) is also known as an ethylene response factor-associated amphiphilic repression (EAR) motif and is involved in the repression of transcription (Jin et al., 2000). There were two types of EAR domain motifs that are present within the core site of the C2 motif region, which have either the DLNxxP or LxLxL patterns (Kagale, Links, & Rozwadowski, 2010). From the protein sequence alignment, BrMYB2 has the exact LxLxL pattern, similar to AtMYB4, AtMYB7 and

AtMYB32 from Arabidopsis (Figure 4.33). In contrast, AmMYB330 replaced the first leucine residue of LxLxL pattern with valine.

From the protein alignment, BrMYB2 exhibits the DNEI sequence motif, similar to other reported C2 repressors from groups A, B and C (Figure 4.33) (Cavallini et al., 2015). Although the functions of element 3 have been elucidated for MYB activators, the presence of element 3 has not been functionally characterized in C2 repressors.

Subsequently, a conserved amino acid signature of the bHLH interaction motif ([D/E]LX₂[R/K]X₃LX₆LX₃R) that is located between the α -helix 1 and α -helix 2 of R3 domain region has been shown to be functionally important for the interaction between MYB and bHLH proteins (Zimmermann et al., 2004). Active and passive C2 repressor MYB proteins that exhibit bHLH conserved interaction motif have the ability to inactivate the MYB-bHLH-WDR (MBW) activation complex (Albert et al., 2014). For instance, petunia PhMYBx, a single repeat R3 MYB protein asserts passive repression mechanism. It contains the signature bHLH interaction motif but lacks the active repressor domains and does not bind DNA directly. PhMYBx asserts the repressive function by competing with the R2R3 MYB for the bHLH binding site and results in the inhibition of the MBW complex formation that is required for the anthocyanin biosynthetic pathway regulation (Albert et al., 2014).

In contrast, PhMYB27 a subgroup 4 R2R3 MYB protein from petunia harbouring both EAR and bHLH interaction motifs exhibit active repression mechanism by directly repressing the structural late anthocyanin biosynthetic genes (*F3H*, *F3'5'H*, *DFR*, *ANS*, *3RT*, *5GT* and *GST*) and regulatory *ANI* gene, encoding the key bHLH factor of the MBW activation complex (Albert et al., 2014). PhMYB27 can also act as a co-repressor that binds to the MBW activation complex, eventually converting it into a repressive complex and represses the same target genes as those targeted by the anthocyanin MBW activation

complex (Albert et al., 2014). The bHLH interaction motif was determined in the protein sequence alignment of BrMYB2 (Figure 4.33). BrMYB2 has a perfect match with the signature in the first α -helix, but substituted the arginine residue with lysine in the second α -helix. However, both residues are positively charged amino acid. Maize ZmMYB31 repressor also shows amino acid substitution in the bHLH interaction motif by replacing the second leucine residue with valine at the first α -helix of R3 domain (Figure 4.33). It was demonstrated that ZmMYB31 represses maize lignin biosynthetic gene expression by directly binding to *ZmCOMT* and *ZmF5H* gene promoters (Fornalé et al., 2010).

In contrast to BrMYB2, BrMYB3 exhibits a perfect amino acid signature of the bHLH interaction motif. However, it does not contain any of the motifs that define subgroups 4, 5, 6, 15 or 27 (Figure 4.36). The members of subgroup 5 R2R3 MYB proteins can be identified based on the presence of DExWRLxxT motif (Stracke et al., 2001). However, some of the subgroup 5 members do not fit this classification perfectly. Manual inspection on AtTT2 and ZmC1 revealed that the protein sequences possess DExWLRxxT motif instead of the DExWRLxxT motif (Figure 4.36). Subsequently, anthocyanin regulating subgroup 6 R2R3 MYB family can be identified by the presence of KPRPR[S/T]F motif at their C-terminal region (Stracke et al., 2001). Subgroup 27 of R2R3 MYB family was determined by the presence of C1 (LlsrGIDPx[T/S]HRx[I/L]) and C3 (DDxF[S/P]SFL[N/D]SLIN[E/D]) motifs at their C-terminal region. Members of subgroup 27 involved in broader functions in plant. Subgroup 15 R2R3 MYB family, which can be identified by the presence of WVxxDxFELSxL motif at C-terminal region.

This result was in accordance with the phylogenetic tree analysis as BrMYB3 formed a small clade with EgWER-like, PdWER-like, MaWER-like, PdMYB23-like, EgMYB308-like, MaMYB3-like and EgrC1-like proteins (Figure 4.35). Based on blastp homology search, these proteins were annotated as predicted proteins and named based

on the similarity with other bHLH dependent R2R3 MYB subgroups (Table 4.22). For instance, EgWER-like, PdWER-like, MaWER-like and PdMYB23-like protein are probably named after WER and MYB23 proteins from subgroup 15. Similarly EgC1-like and EgMYB308-like proteins were probably named after ZmC1 protein from subgroup 5 and AmMYB308 from subgroup 4, respectively. However, none of these predicted proteins have the additional subgroup-specific motifs (Figure 4.36). The only similarity found was the bHLH interaction motif at N-terminal region. Furthermore, the entire predicted proteins do not clustered into their predicted subgroup in the phylogenetic tree analysis but formed a new cluster, which include BrMYB3 protein (Figure 4.35).

Accordingly, subgroup 6 R2R3 MYB that regulate anthocyanin pathway possess the signature pattern of ANDV, while element 3 motif for subgroup 15 R2R3 MYB that regulate trichome and hair root development was DNQV (Figure 4.36). In contrast, BrMYB3 exhibits element 3 motif with the signature pattern of DNEI, and which is similar to other non-anthocyanin regulated MYBs from subgroups 4, 5 and 27 (Figure 4.36). Therefore, it can be suggested that BrMYB3 might be involved in a non-anthocyanin biosynthetic pathway in the phenylpropanoid metabolism, and can be classified as a new R2R3 MYB subgroup.

Functional studies of *BrMYB1*, *BrMYB2* and *BrMYB3* will be done in the near future to further understand their roles in plants. The studies will include; identification of target genes *in vitro* through identification of any consensus DNA-binding site at the target genes promoter region; yeast two- and three- hybrid assays to study the interaction of MYB proteins with other transcription factors such as bHLH and WDR protein; and overexpression of *BrMYB1*, *BrMYB2* and *BrMYB3* genes to study the transcriptional regulation *in vivo* and their relationship with flavonoid synthesis, especially panduratin

A. Additionally, a complete genome-wide MYB transcription factor analysis can be accomplished once the *B. rotunda* genome database is available.

In this study, the presence of the introns, if any, in *BrMYB1*, *BrMYB2* and *BrMYB3* genes were identified through a conventional sequencing approach and not through mapping of the target sequences to *B. rotunda* genome database. The sequencing results for *BrMYB1*, *BrMYB2* and *BrMYB3* genes were presented in Figure 4.28, Figure 4.31 and Figure 4.34, respectively. From the sequencing results, no intron was found in *BrMYB1*, while one intron was found in both *BrMYB2* and *BrMYB3* DNA sequences. The location of the introns in *BrMYB2* and *BrMYB3* are represented in Figure 4.37. The *BrMYB2* gene had an 85 nucleotide long intron located within the R3 domain, whereas the *BrMYB3* gene had an 80 nucleotide long intron located within the R2 domain (Figure 4.37). Interestingly, both introns were classified as canonical intron splicing sites by starting with the dinucleotide GT, at the 5' end, donor site and ended with AG at the 3' end, acceptor site (Figure 4.31 and Figure 4.34). In plant system, canonical introns are removed by U2-type spliceosome (Dubrovina et al., 2012).

Subsequently, the expression of *BrMYB2* in response to phenylalanine in cell suspension culture was analyzed using qPCR (Figure 4.39). The results revealed that *BrMYB2* gene expression was up-regulated after 1, 7 and 14 days upon phenylalanine treatment. As expected the up-regulation fold changes was higher than the control samples. This is as previously reported by Tan et al. (2012). Hence, it can be suggested that regulation of *BrMYB2* might repress the lignin pathway.

5.3 Enzymes that are involved in the panduratin A pathway

In addition to genes encoding regulatory proteins, genes which encode enzymes are also important in controlling the yield of the bioactive compounds in *B. rotunda*. Figure 4.13 shows the proposed panduratin A biosynthetic pathway, which is derived from chalcones in the flavonoid pathway. Through RNA-Seq and differentially expressed genes analysis, genes that are potentially involved in panduratin A synthesis were identified (Table 4.7 and Table 4.8). From the results, it can be inferred that the isoflavanoids biosynthetic pathway may not be present in *B. rotunda* as unigene mapped to isoflavonoid pathway in KEGG. Additionally, there was no unigene mapped to flavone synthase (FS), suggesting that flavones might not be present in the *B. rotunda* cell suspension culture. However, in contrast, flavones were successfully isolated and identified in black rhizome of *Boesenbergia pandurata* (Herunsalee et al., 1987; Jaipetch et al., 1983). From the transcriptome data, it could be suggested that different rhizome varieties may have different flavonoid biosynthesis pathways as the source of cell suspension culture in this study originated from yellow rhizomes and this would merit further investigation. The other enzyme that had no unigene mapped to it was anthocyanin reductase (ANR), which converts anthocyanidins to flavan-3-ols, which eventually polymerizes to form proanthocyanidins.

The other enzymes in the flavonoid pathway consist of both up- and down-regulated unigenes except for chalcone isomerase (CHI), dihydroflavonol-4-reductase (DFR) and leucoanthocyanidin reductase (LAR) (Figure 4.13). There were no significant gene regulation patterns in CHI, whereas down-regulated unigenes were identified for both DFR and LAR (Table 4.8). Most of the unigenes that were mapped to the remaining flavonoid enzymes such as phenylalanine ammonia-lyase (PAL), cinnamate-4-hydroxylase (C4H), 4-coumaroyl:coenzyme A ligase (4CL), chalcone synthase (CHS), flavanone-3-hydroxylase (F3H), flavonol synthase (FLS) and anthocyanin synthase

(ANS) were down-regulated. It is suggested that down-regulation of enzyme isomers in the flavonoid pathway causes switch-off of competitive pathways and eventually diversion of the metabolic flux to the production of the desired secondary metabolites.

There were three enzymes known to be directly involved in panduratin A production, PAL, 4CL and CHS (Figure 4.13) (Bowsher et al., 2008). These enzymes are known to be encoded by a multi-gene family. Phenylalanine ammonia-lyase (PAL) catalyzes the first step in phenylpropanoid biosynthetic pathway. In many plant species, several copies of the PAL gene have been found and characterized. Between 2 to 4 PAL genes have been identified in *Arabidopsis*, tobacco, bean and parsley (Cramer et al., 1989; Fukasawa-Akada et al., 1996; Lois et al., 1989; Wanner et al., 1995). More than 40 PAL genes were identified in potato (Joos & Hahlbrock, 2005). Although more than one PAL gene is present in each plant species, the regulation of each PAL gene depends on different response of stimuli (Lois et al., 1989). In this study, there were 14 unigenes that were mapped as PAL. However, only 2 unigenes were up-regulated in response to the addition of phenylalanine (Table 4.7).

The second enzyme 4CL, which is one of the main enzyme involved in the pathway showed a gene regulation pattern directly involved in the panduratin A production. There were in total 44 unigenes mapped to 4CL in *B. rotunda*. However, only 3 unigenes were up-regulated and 15 unigenes were down-regulated after 14 days post treatment with phenylalanine (Table 4.7). 4CLs can be divided into two types in *Arabidopsis thaliana*; type I is responsible for lignin formation and type II leads to branching of the flavonoid pathways to produce flower pigments and defence mechanisms (Ehltling et al., 2002). However, in rice, other than type I 4CLs cluster, none were clustered in type II, but instead clustered separately in type III (Gui et al., 2011). Although type I 4CLs in dicots and type III 4CLs in monocots are suggested to lead to lignin formation, they differ in sequences

and substrate preference (Gui et al., 2011). Similarly for type II 4CLs, which also have differences in substrate preference and eventually cause branching in flavonoid biosynthetic pathway (Ehlting et al., 2002). Hence it could be suggested that the remaining non-regulated 4CLs in *B. rotunda* might also be involved in lignin formation or possess different substrate preference.

As for chalcone synthase (CHS), it is categorized under the type III polyketide synthase superfamily (Austin & Noel, 2003). It catalyzes the formation of chalcones by condensing one p-coumaroylCoA and three units of malonyl-CoA (Ferrer et al., 1999). Different combination of thioesters and three units of malonyl-CoA were catalysed by CHS and eventually produce different chalcones for instance, a condensation reaction of p-coumaroyl-CoA gave rise to naringenin chalcone while condensation of cinnamoyl-CoA gave rise to pinocembrin chalcone (Austin & Noel, 2003). It was reported that each CHS has a different substrate preference by *in vitro* determining CHS relative activity percentage (Yamazaki et al., 2001). Although more than one CHS gene was isolated from one species, some CHS isoenzymes were constitutively expressed throughout the plant development with varying expression levels but some were expressed upon induction by environmental stresses including wounding, UV light and pathogen infections (Wingender et al., 1989). In the transcriptome data, 25 CHSs were mapped in the KEGG database, with only Unigene1735_All shown up-regulation while other seven unigenes shown down-regulation (Table 4.8 and Table 4.9).

As presented in Figure 4.13, panduratin A was proposed to be derived from chalcone. Other than panduratin A, bioactive compounds such as cardamonin, alpinetin, pinocembrin, pinostrobin and 4-hydroxypanduratin A were also proposed to be derived from chalcone (Figure 4.14). These compounds were successfully extracted from *B. rotunda* rhizome (Kiat et al., 2006). However, the biosynthetic pathway in *B. rotunda* has

not been established in the plant KEGG pathway. Hence, the proposed enzymes of the biosynthetic pathway was solely based on the compound structures. In this thesis, other than enzymes mentioned earlier, there were two more important enzymes that were proposed to be involved in the bioactive compounds production in *B. rotunda*. As presented in Figure 4.14, the enzymes are prenyltransferase (PT) and flavonoid O-methyltransferase (ChOMT1, ChOMT2, 5-OMT and 7-OMT).

A diverse complex of flavonoid derivatives can be produced through flavonoid modification such as prenylation and O-methylation. In this thesis, it was proposed that prenyltransferase and flavonoid O-methyltransferase are the key enzymes in synthesizing important bioactive compounds in *B. rotunda* such as cardamonin, alpinetin, pinocembrin, pinostrobin, 4-hydroxypanduratin A and panduratin A (Figure 4.14). Based on the compound structures and the reported compounds that were extracted in *B. rotunda*, it could be suggested that alpinetin, pinostrobin, 4-hydroxypanduratin A and panduratin A are the end products of the pathway (Tan, 2005).

5.4 Prenyltransferase

In this thesis, flavonoid pathway leading towards cardamonin, alpinetin, pinocembrin, pinostrobin, 4-hydroxypanduratin A and panduratin A production was proposed (Figure 4.14). The first key enzyme in the proposed pathway is flavonoid prenyltransferase, which is involved in the synthesis of important prenylated flavonoid in *B. rotunda* (Figure 4.14). Flavonoid prenyltransferase catalyzes the process of transferring the prenyl moiety to flavonoid compound (Brandt et al., 2009). It is proposed that PT2 transfers the prenyl moiety to pinocembrin chalcone to produce 4-hydroxypanduratin A whereas, prenylation on pinostrobin chalcone by PT1 produces panduratin A (Figure 4.14).

In the *B. rotunda* transcriptome database, there was only one unigene (Unigene31983_All) annotated as flavonoid prenyltransferase. There were two conserved motifs found in the predicted Unigene31983_All ORF which included the prenyl moiety binding motif (NQxxDxxxD) and flavonoid or homogentisate binding motif (KDxxDx[E/D]GD). These motifs were highly conserved across plant flavonoid prenyltransferase proteins (Figure 4.17). In order to predict the function of Unigene31983_All, a phylogenetic tree was constructed using other plant aromatic and homogentisate prenyltransferases (Figure 4.18). According to the phylogenetic tree analysis by Wang et al. (2014), flavonoid prenyltransferases evolved from HG prenyltransferases (Figure 4.18).

Interestingly, leguminous flavonoid prenyltransferases and moraceous flavonoid prenyltransferases were branched into two different clusters. The leguminous flavonoid prenyltransferases were evolved from the HG prenyltransferases that synthesizes vitamin E, whereas moraceous flavonoid prenyltransferases were evolved from the HG prenyltransferases that synthesizes plastoquinone. The results implies that flavonoid prenyltransferases evolved independently in the two plant lineages (Wang et al., 2014). Although the candidate flavonoid prenyltransferase was clustered under homogentisate prenyltransferase and not the flavonoid prenyltransferase, further functional characterization of the Unigene31983_All such as substrate-specific study should be done. This is because Unigene31983_All has the possibility to function as flavonoid prenyltransferase as demonstrated by leguminous flavonoid prenyltransferases, which evolved from HG prenyltransferase. Moreover, no flavonoid prenyltransferase has been characterized from the Zingiberaceae family.

5.5 Flavonoid O-methyltransferase

In plants, flavonoid O-methylation was catalyzed by flavonoid O-methyltransferase (OMT), which uses S-adenosylmethionine (SAM) as a methyl group donor and flavonoid as methyl group acceptor (Kim et al., 2010). Flavonoid O-methyltransferase can be classified into two groups depending on the substrates used; either flavonoids (Group I) or isoflavonoids (Group II) (Kim et al., 2010).

In this thesis, there are four flavonoid O-methyltransferases proposed to be involved in the panduratin A biosynthetic pathway in *B. rotunda* which includes two chalcone O-methyltransferases and two flavanone O-methyltransferases (Figure 4.14). Based on the chemical structures, it is suggested that chalcone O-methyltransferases: pinocembrin chalcone 4'-O-methyltransferase (ChOMT1) and pinocembrin chalcone 2'-O-methyltransferase (ChOMT2) transfer a methyl group to catalyze O-methylation on pinocembrin chalcone. Both ChOMT1 and ChOMT2 compete for the same substrate, but different products are produced due to O-methylation on different hydroxyl group locations. It is proposed that ChOMT1 transfers the methyl group to 4'-hydroxyl group producing pinostrobin chalcone, while ChOMT2 transfers the methyl group to 2'-hydroxyl group producing cardamonin (Figure 4.14).

Subsequently, pinocembrin 5-O-methyltransferase (5-OMT) and pinocembrin 7-O-methyltransferase (7-OMT) are flavanone O-methyltransferases that are proposed to catalyze O-methylation on 5- and 7-hydroxyl group of pinocembrin, producing alpinetin and pinostrobin, respectively (Figure 4.14). However, it was reported that methylation on 5'-hydroxyl group was hindered due to the tendency of 5'-hydroxyl group to form hydrogen bond with carbonyl group located at C4 (Kim et al., 2010). Therefore, it can be suggested that alpinetin in *B. rotunda* might be produced through isomerization of

cardamonin, not by O-methylation of pinocembrin. Hence, 5-OMT might not be present in *B. rotunda*.

In the *B. rotunda* transcriptome database, there was only one unigene (Unigene891_All) annotated as flavonoid O-methyltransferase that has a complete ORF (Appendix I.2). While another unigene, (Unigene25107_All) that was also annotated as flavonoid O-methyltransferase, had an incomplete ORF (Appendix I.2).

Multiple sequence alignment of Unigene891_All with other plant flavonoid O-methyltransferases revealed that Unigene 891_All consists of a conserved Histidine residue at amino acid 268 which serves as a base for the O-methylation reaction in FOMTs (Figure 4.15). From the alignment, it showed that the ORF of Unigene21507_All was incomplete towards the end of C-terminus. A dimerization domain, involved in the formation of dimer interface with the substrate binding site was located at the N-terminus, whereas the catalytic O-methyltransferase domain was located at the C-terminus (Zubieta et al., 2001).

Subsequently, the phylogenetic tree analysis showed that Unigene891_All does not cluster into chalcone O-methyltransferase (MsCHMT) or flavanone 7-O-methyltransferase (OsNMT). Instead, Unigene891_All formed a small clade with flavone O-methyltransferase1-like (MaFl-OMT1-like_1 and MaFl-OMT1-like_3) and tricetin 3',4',5'-O-trimethyltransferase-like proteins (MaFl-OMT1-like_2) from *Musa acuminata* subsp. *malaccensis* (Figure 4.16). These proteins were annotated as predicted proteins and named based on similarity with flavone O-methyltransferase. It was reported that flavone 7-O-methyltransferase, Fl-OMT from barley (*Hordeum vulgare* subsp. *vulgare*) functions as a flavone 7-O-methyltransferase (Christensen et al., 1998). Although MaFl-OMT1-like_1, MaFl-OMT1-like_2 and MaFl-OMT1-like_3 might be named based on similarity with Fl-OMT, they were not clustered together in the phylogenetic tree.

Interestingly, Unigene891_All together with MaFl-OMT1-like_1, MaFl-OMT1-like_2 and MaFl-OMT1-like_3 are also not clustered into other flavone O-methyltransferase that use flavone such as HvOMT, TaOMT2, ZmOMT1 and OsOMT1 (Figure 4.16). From the results, it can be speculated that Unigene891_All might not catalyse O-methylation on flavone substrate. These results correlate with previous analysis that suggests that no flavones are produced in *B. rotunda* as no unigene mapped to flavone synthase (Figure 4.13). Therefore, it can be predicted that Unigene891_All is a gene encoding O-methyltransferase enzyme. Further functional studies should be carried out in order to determine the substrate specificity and catalytic mechanism.

On the other hand, Unigene25107_All shares the same clade with Fl-OMT. Therefore, it can be suggested that Unigene25107_All might also catalyze 7-O-methyltransferase activity. Further functional characterization studies can be suggested in order to determine Unigene 25107_All substrate specificity. Other than Fl-OMT, it was reported that O-methyltransferase from *Oryza sativa*, OsNOMT also catalyzes 7-methyltransferase activity. However, unlike Fl-OMT that transfers a methyl group to flavone, OsNMT transfers a methyl group to naringenin, a flavanone compound. Although both Fl-OMT and OsNOMT catalyzes the same activity, they were not clustered together. According to Kim et al. (2010), 7-OMTs were substrate exclusive and flavonoid 7-OMTs do not show high sequence similarity to one another, which explains the different clusters for Fl-OMT and OsNOMT in this phylogenetic tree (Figure 4.16).

From the phylogenetic analysis, none of the candidate unigenes clustered with chalcone O-methyltransferase, MsCHMT. It was reported that MsCHMT, a chalcone O-methyltransferase from *Medicago sativa* catalyzes O-methylation on the 2'-hydroxygroup of a chalcone, isoliquiritigenin producing 4,4'-dihydroxy-2'-methoxy chalcone. According to Zubieta et al. (2001), the substrate binding site of MsCHMT

appeared to be pre-arranged upon SAM binding. Accordingly, this methylation reaction prevents cyclization of isoliquiritigenin to flavanone liquiritigenin (7,4'-dihydroxyflavanone) by CHI (Zubieta et al., 2001). Hence, producing less flavanone liquiritigenin.

5.6 Bioactive compounds proposed pathway

Panduratin A is a unique bioactive compound that is exclusively synthesized in *B. rotunda*. As mentioned previously, panduratin A exhibits a number of medicinal properties (Table 2.4), resulting in increasing demands of the compound in medicinal and health care sectors. However, the production of panduratin A in nature was limited (Tewtrakul et al., 2009). One way to increase the production yield of panduratin A is through metabolic engineer its biosynthetic pathway. Nevertheless, panduratin A biosynthetic pathway has not been established.

Panduratin A shares a similar carbon structure with other bioactive compounds extracted in *B. rotunda* such as pinostrobin, pinocembrin, alpinetin, cardamonin 4-hydroxypanduratin A (Figure 2.10). Hence, in order to see the biosynthetic pathway relationship between these compounds, a new flavonoid pathway leading towards naringenin, pinostrobin, pinocembrin, alpinetin, cardamonin, panduratin A and 4-hydroxypanduratin A is proposed in this thesis (Figure 4.14)

As shown in Figure 4.14, through cinnamic acid, the pathway is diverged into two main pathways. The first pathway is the pathway leading towards the production of pinostrobin, pinocembrin, alpinetin, cardamonin, panduratin A and 4-hydroxypanduratin A through pinocembrin chalcone and pinostrobin chalcone. The second pathway is the pathway leading towards the production of naringenin through naringenin chalcone.

The extraction of bioactive compounds from *B. rotunda* rhizome yielded 1,136.3 µg/g of pinostrobin followed by 4,918.2 µg/g of pinocembrin, 3,738.0 µg/g of alpinetin, 791.4 µg/g of 4-hydroxyanduratin A, 428.6 µg/g of panduratin A, 146.2 µg/g of cardamonin and 95.9 µg/g of pinostrobin chalcone (Tan, 2005). According to Tan (2005), there was no naringenin produced in the *B. rotunda* rhizome. However, naringenin was successfully extracted from other plants such as soybean (Porter et al., 1986) and *Mimosa hostilis* (Ohsaki et al., 2006). Therefore, it can be speculated that naringenin biosynthetic pathway might not be present in *B. rotunda*.

Based on the amount of compounds produced in *B. rotunda* rhizome reported by Tan (2005), hypothesis on metabolic flux in the proposed flavonoid pathway leading towards pinostrobin, pinocembrin, alpinetin, cardamonin, panduratin A and 4-hydroxyanduratin A production is suggested in this thesis (Figure 5.1). It can be speculated that there are four pathways leading to the production of bioactive compounds. As illustrated in Figure 5.1, the first pathway is pinostrobin and pinocembrin production pathway, followed by alpinetin and cardamonin production pathway, 4-hydroxyanduratin A production pathway and lastly panduratin A production pathway.

Subsequently, as shown in Figure 5.1, pinocembrin chalcone is the central substrate producing all of the bioactive compounds in this proposed pathway. It was reported that the highest amount of bioactive compounds extracted from *B. rotunda* rhizome was pinostrobin (Tan, 2005). Therefore it can be speculated that most of the pinocembrin chalcone is converted to pinostrobin and pinocembrin. This conversion might occur through spontaneous non-enzymatic isomerization rather than enzymatic CHI reaction. Spontaneous isomerization of pinocembrin chalcone to pinocembrin might occur without any involvement of CHI due to relatively unstable chalcone compounds in nature (Bowsher et al., 2008). Supporting the hypothesis, unigenes annotated as CHI in *B.*

rotunda transcriptome data also shown non-significantly regulated in response to phenylalanine treatment (Table 4.8). Therefore it can be speculated that the biggest metabolic flux flow in the proposed flavonoid pathway is leading towards pinostrobin production. Pinostrobin is the end product of this pathway. Hence the yield was higher compared to pinocembrin.

In addition, pinocembrin chalcone also serves as primary substrate to produce cardamonin, 4-hydroxyanduratin A and pinostrobin chalcone (Figure 5.1). It is suggested that enzyme competition occur between ChOMT2, ChOMT1 and PT2 to catalyse pinocembrin chalcone in the flavonoid proposed pathway (Figure 5.1). Unfortunately, flavonoid O-methyltransferase and prenyltransferase found in *B. rotunda* transcriptome data does not show correlations with results reported by Tan (2005). The unigenes analysed were non-significantly regulated. Further analysis have to be carried out to elucidate the genes encoding both flavonoid O-methyltransferase and prenyltransferase that involved in this pathway. Based on the report of bioactive compounds extracted from *B. rotunda* rhizome, alpinetin production was higher compared to 4-hydroxyanduratin A and anduratin A (Tan, 2005). Thus, it can be speculated that the second high metabolic flux in the proposed flavonoid pathway is towards alpinetin production followed by 4-hydroxyanduratin A production.

Hence, anduratin A biosynthetic pathway has the least metabolic flux in the proposed flavonoid pathway (Figure 5.1). From the flavonoid proposed pathway, it can be suggested that most of the pinostrobin chalcone, the substrate for anduratin A biosynthesis might be isomerized spontaneously to form pinostrobin. The metabolic flux of this isomerization is greater compared to prenylation of pinostrobin chalcone, which explain the low abundance of anduratin A in *B. rotunda* rhizome as reported by Tan (2005). To validate and establish this proposed pathway, molecular characterization and

functional study through enzyme assay analysis of the cDNA encoding enzymes that are involved in the proposed pathway will be conducted in the future.

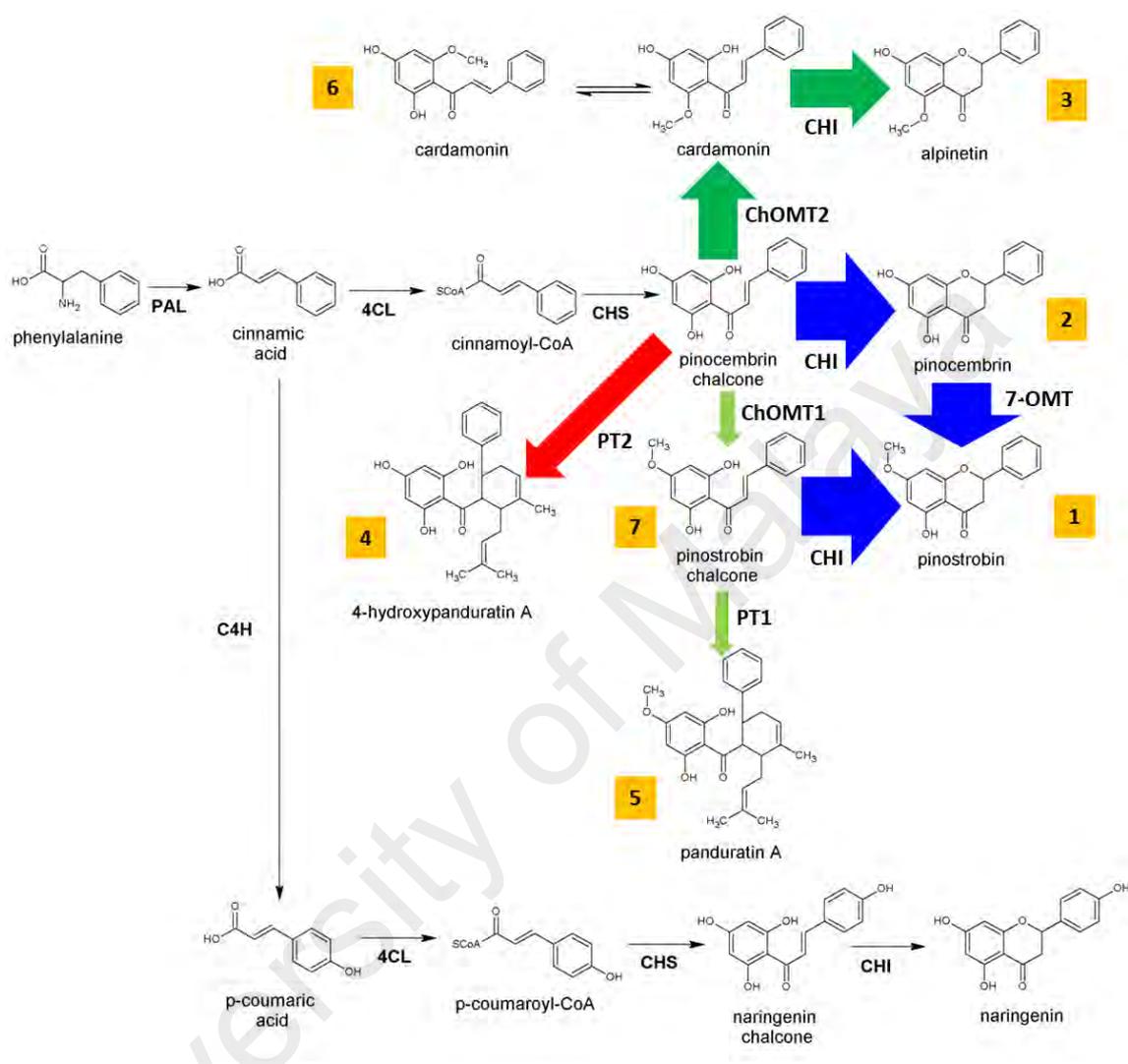


Figure 5.1: Hypothesis on metabolic flux in the proposed flavonoid pathway leading to cardamonin, alpinetin, pinocembrin, pinostrobin, 4-hydroxypanduratin A and panduratin A production. Established flavonoid pathway was adapted from Bowsher et al., 2008.

Abbreviations: PAL, phenylalanine ammonia lyase; C4H, cinnamate-4-hydroxylase; 4CL, 4-coumaroyl:coenzyme A ligase; CHS, chalcone synthase; CHI, chalcone isomerase; PT1, prenyltransferase1; PT2, prenyltransferase2; 7-OMT, pinocembrin 7-O-methyltransferase; ChOMT1, pinocembrin chalcone 4'-O-methyltransferase and ChOMT2, pinocembrin chalcone 2'-O-methyltransferase. Blue arrows indicate the metabolic flux towards production of pinocembrin and pinostrobin. Dark green arrows indicate the metabolic flux towards production of cardamonin and alpinetin. Red arrow indicates the metabolic flux towards production of 4-hydroxypanduratin A. Light green arrows indicate the metabolic flux towards production of panduratin A. The thickness of the arrows indicate the metabolic flux flow tendency. The relative amount of bioactive compound yield in *B. rotunda* rhizome as reported by Tan (2005) was denoted by number 1 – 7, with number 1 indicates the highest, while number 7 indicates the lowest.

5.7 Proposed prenylation mechanism

As mentioned previously, proposed panduratin A biosynthetic pathway in *B. rotunda* has been suggested in Figure 5.1. As a cyclohexenyl chalcone compound, cyclization after prenylation of panduratin A is a crucial step. Thus, in this thesis, the prenylation mechanism steps are proposed based on the chemical structure backbones of panduratin A and other prenylated flavonoids that found in *B. rotunda* (Table 5.1 and Figure 5.2). In nature, there have been reports of more than 10 prenylated flavonoids that has been extracted from *B. rotunda* (Morikawa et al., 2008; Win et al., 2007) (Figure 2.16). Table 5.1 shows the proposed aromatic compound, type of prenyl moiety and C-prenylation location that involved in the biosynthesis of prenylated flavonoids in *B. rotunda*.

Based on the structure backbone, all of the prenylated flavonoids are proposed to have geranyl side chain prenylated on either flavanone or chalcone compounds (Figure 2.16 and Table 5.1). Thus, it can be suggested that prenyltransferase involved in prenylated flavonoids biosynthesis in *B. rotunda* are specific to geranyl diphosphate, transferring geranyl side chain to flavanone and chalcone. Next, some C-prenylation are suggested to take place on ring A at C-6 or C-8 of flavanone and C-3' of chalcone producing prenylated flavanone and prenylated chalcone, respectively (Figure 2.16 and Table 5.1). Prenyated flavonoids such as 6-geranylpinostrobin and 6-geranylpinocembrin do not undergo further modification after prenylation.

However, some of the prenylated flavonoids such as panduratin A and 4-hydroxypanduratin A are proposed to undergo further cyclization step after prenylation (Table 5.1). In this thesis, the proposed prenylation mechanism of chalcones to produce prenylated chalcones or cyclohexenyl chalcones is illustrated in Figure 5.2. The proposed prenylation mechanism involves C-prenylation of of chalcones at C- α or C- β , followed by isomerization by cis-trans isomerase, dehydrogenation by dehydrogenase (NAD⁺) and

cyclization by nucleophile attacks to the carbocation intermediate producing cyclohexenyl chalcones (Figure 5.2). It is suggested that isomerization by cis-trans isomerase will results in changing the arrangement of the geranyl moiety structure that located at either C- α or C- β of the chalcones. Thus the prenylated chalcones have the same molecular formula but a different physical structure (Figure 5.2). Next, dehydrogenation by dehydrogenase (NAD⁺) will results in carbocation intermediate formation. This carbocation intermediate will cause nucleophile attacks located either at C- α or C- β (prenylation at C- β or C- α , respectively) and eventually resulting in hexene ring formation (Figure 5.2).

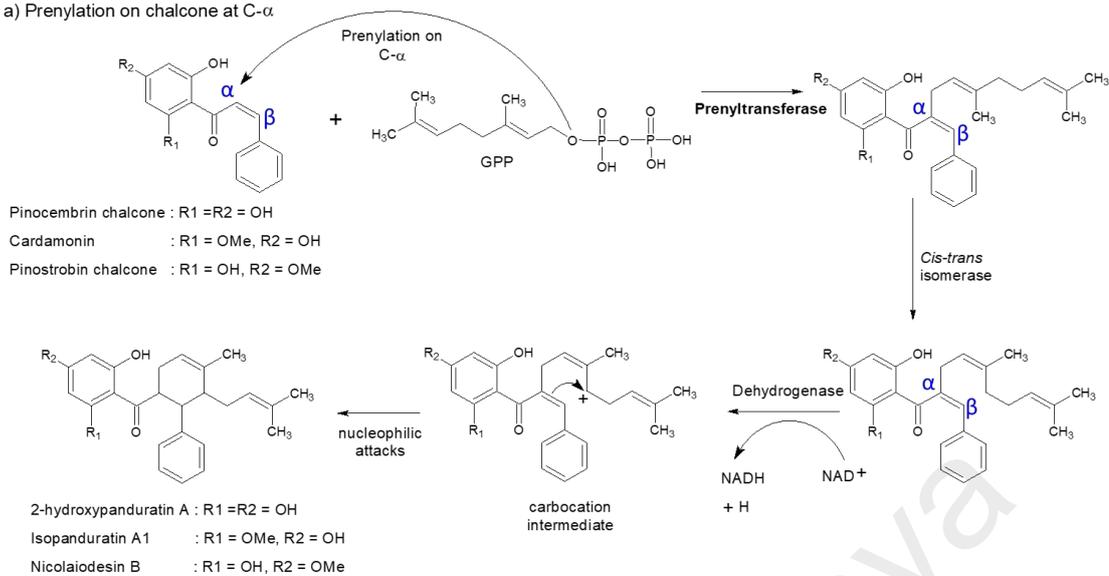
Different C-prenylation location on the same chalcone producing different cyclohexenyl chalcones. Prenylation on pinocembrin chalcone, cardamonin and pinostrobin chalcone at C- α result in biosynthesis of 2-hydroxypanduratin A, isopanduratin A1 and nicolaidisin B, respectively. On the other hand, prenylation of the same chalcone compounds at C- β produce 4-hydroxypanduratin A, isopanduratin A2 and panduratin A, respectively.

It is proposed that at least another two enzymes involve in the cyclization after prenylation of the pinostrobin chalcone in panduratin A biosynthesis. However, in this study, the unigenes that annotated as cis-trans isomerase and dehydrogenase (NAD⁺) in *B. rotunda* transcriptome data were not analysed. Further identification and characterization of these unigenes will be done in the future to understand more about panduratin A biosynthesis.

Table 5.1: Proposed prenylation of prenylated flavonoids found in *B. rotunda*, which include the type of prenyl moiety, type of aromatic compound, C-prenylation location and involvement of further modification after prenylation.

Bioactive compounds	Prenyl moiety	Aromatic compound	C-prenylation location	Further modification (Cyclization)
3'-geranylcardamonin	Geranyl	chalcone	ring A, C-3'	No
Rotundaflavone I	Geranyl	flavanone	ring A, C-8	No
Rotundaflavone II	Geranyl	flavanone	ring A, C-8	No
6-geranylpinostrobin	Geranyl	flavanone	ring A; C-6	No
6-geranylpinocembrin	Geranyl	flavanone	ring A; C-6	No
8-geranylpinostrobin	Geranyl	flavanone	ring A; C-8	No
8-geranylpinocembrin	Geranyl	flavanone	ring A; C-8	No
7,8-dihydro-5-hydroxy-2-methyl-2-(4'-methyl-3'-pentenyl)-8-phenyl-2H,6H-benzo[2,1- <i>b</i> :5,4- <i>b'</i>]dipyran-6-one	Geranyl	flavanone	ring A; C-6	Yes
Panduratin A	Geranyl	chalcone	C- β	Yes
Panduratin C	Geranyl	chalcone	C- β	Yes
4-hydroxypanduratin A	Geranyl	chalcone	C- β	Yes
Krachaizin A	Geranyl	chalcone	C- β	Yes
Krachaizin B	Geranyl	chalcone	C- β	Yes
2-hydroxyisopanduratin A	Geranyl	chalcone	C- α	Yes
Isopanduratin A1	Geranyl	chalcone	C- α	Yes
Isopanduratin A2	Geranyl	chalcone	C- β	Yes
6-methoxypanduratin A	Geranyl	chalcone	C- β	Yes
Nicolaoidesin B	Geranyl	chalcone	C- α	Yes

a) Prenylation on chalcone at C- α



b) Prenylation on chalcone at C- β

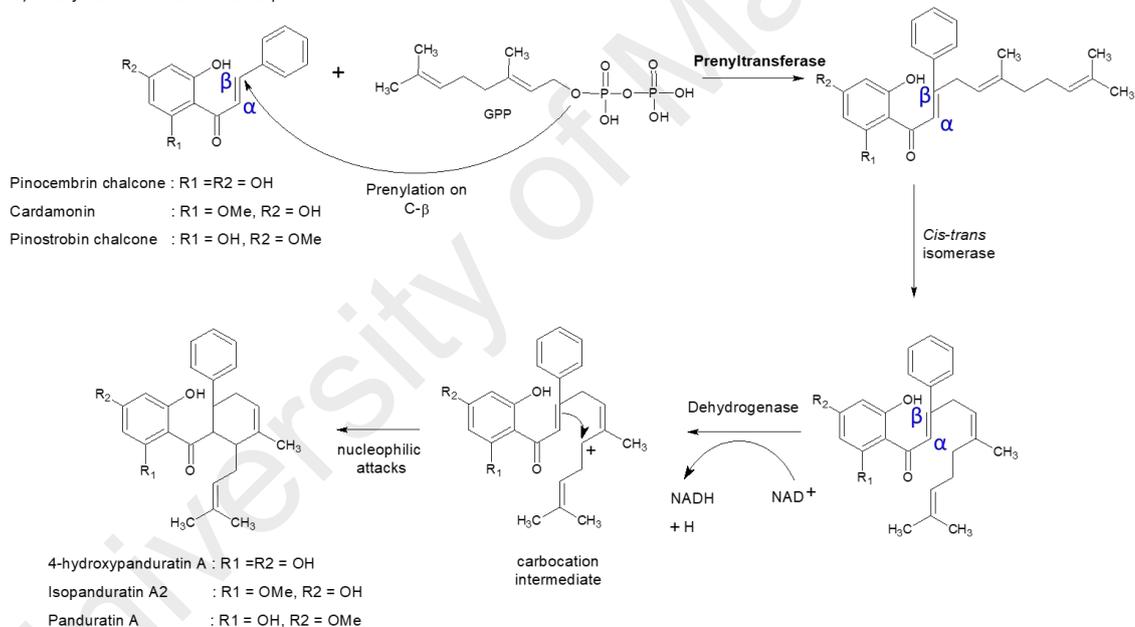


Figure 5.2: Proposed mechanism of prenylated chalcone by prenyltransferase through C- α (a) and C- β (b) prenylation. Isomerization, dehydrogenation and cyclization reactions were adapted from Bailey & Bailey (2000).

5.8 Approaches to increase panduratin A production in the future

Flavonoid pathways in *B. rotunda* can potentially be manipulated through metabolic engineering using the information generated in this study. The metabolic engineering approach can be achieved through overexpression and/or silencing of the genes encoding enzymes in the proposed panduratin A biosynthetic pathway (Figure 5.1). Prior to metabolically engineering the panduratin A production pathway in *B. rotunda* plant, it is crucial to establish the proposed pathway first. Although the genes that might be encoded as the enzymes in the proposed pathway can be elucidated through *B. rotunda* transcriptome data, the first step to establish the pathway is to perform the functional characterization of the genes. This approach is also important in order to select the best genes encoding the relevant enzymes.

Functional characterization can be achieved through mutant yeast complementation experiments. This complementary experimental approach is based on employing the specific mutant yeast that cannot grow on certain media composition. However, mutant yeast harbouring the gene of interest could grow on the selectable media due to their ability to restore the physical deficiency of the mutated yeast (Avelange-Macherel & Joyard, 1998; Bolognese & McGraw, 2000; Campbell et al., 1998; Frémont et al., 2013), which consequently indirectly verifies their gene function. The functional complementary approach using mutant yeast has successfully elucidated genes in *Arabidopsis thaliana* such as for isopentenyl diphosphate isomerase encoded by the *IPP* gene, methyltransferase encoded by the *COQ3* gene, phosphor-ethanilamine N-methyltransferase and acetylornithine aminotransferase encoded by the *TUP5* gene (Avelange-Macherel & Joyard, 1998; Bolognese & McGraw, 2000; Campbell et al., 1998; Frémont et al., 2013). The same approach has also elucidated for the iron transporter gene function (*MxIRT1*) in *Malus xiaojinesis* (Zhang et al., 2013).

Besides, another approach to determine the function of the putative genes is through substrate-specificity study and enzyme assay. It has been reported that through these approaches, functional characterization of prenyltransferases from *Lithium erythrorizon* (LePGT-1) and *Sophora flavescens* (SfN8DT-1, SfG6DT and SfiLDT) were elucidated (Yazaki et al., 2002; Sasaki et al., 2008; Sasaki et al., 2011). The genes were transformed into yeast w303-1A- Δ coq2 strain which had disrupted *COQ2* gene. The gene encodes for 4-hydroxybenzoate prenyltransferase that utilized prenyl diphosphate for ubiquinone synthesis in yeast. According to Sasaki et al. (2009), the disruption of *COQ2* gene was important to provide a prenyl donor efficiently for prenylation. Eventually, substrate-specificity study and enzyme assay were performed using microsomal fraction of the transformed yeast.

Since panduratin A biosynthetic pathway is the least favourable pathway in this proposed flavonoid pathway, it is a challenging task to increase panduratin A production in *B. rotunda* (Figure 5.1). First, at the regulatory gene level, it can be suggested that the lignin pathway can be repressed through overexpression of the subgroup 4 R2R3 MYB transcription factor repressor. Hence, the metabolic flux will be diverted to the flavonoid pathway, providing more pinocembrin chalcone substrate in the proposed flavonoid pathway (Figure 5.1). From *B. rotunda* transcriptome data, *BrMYB2* was annotated as subgroup 4 R2R3 MYB transcription factor. Up-regulation of *BrMYB2* upon phenylalanine treatment in *B. rotunda* cell suspension culture makes it as a good candidate for overexpression study in *B. rotunda*. However, prior to overexpression of *BrMYB2* in *B. rotunda*, functional study of *BrMYB2* should be performed first to ensure it plays roles as lignin pathway repressor.

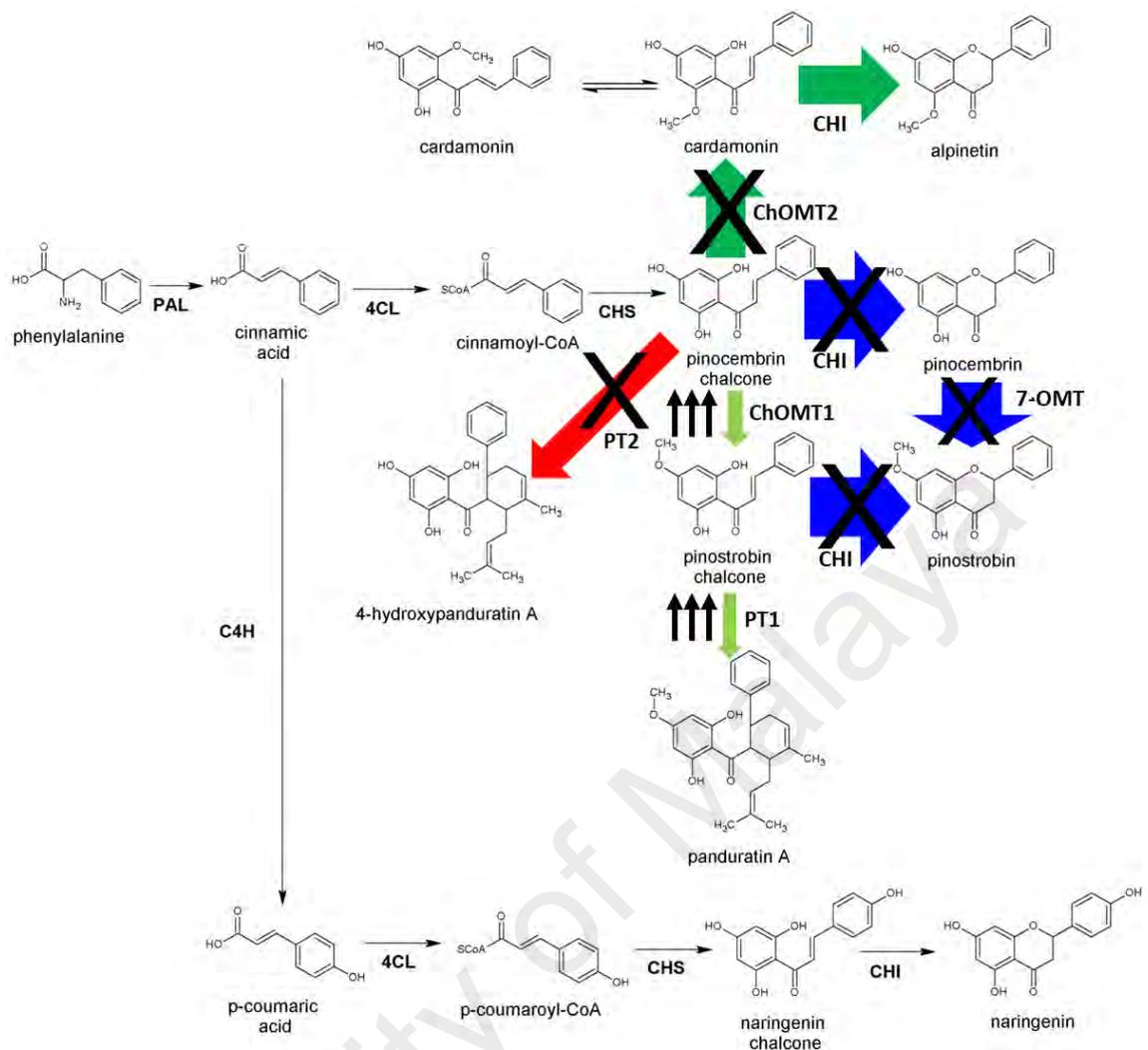


Figure 5.3: Suggestions on changing the metabolic flux in the proposed flavonoid pathway leading to increase panduratin A production. Established flavonoid pathway was adapted from Bowsher et al., 2008.

Abbreviations: PAL, phenylalanine ammonia lyase; C4H, cinnamate-4-hydroxylase; 4CL, 4-coumaroyl:coenzyme A ligase; CHS, chalcone synthase; CHI, chalcone isomerase; PT1, prenyltransferase1; PT2, prenyltransferase2; 5-OMT, pinocembrin 5-O-methyltransferase, 7-OMT, pinocembrin 7-O-methyltransferase, ChOMT1, pinocembrin chalcone 4'-O-methyltransferase and ChOMT2, pinocembrin chalcone 2'-O-methyltransferase. Blue arrows indicate the metabolic flux towards production of pinocembrin and pinostrobin. Dark green arrows indicate the metabolic flux towards production of cardamomin and alpinetin. Red arrow indicates the metabolic flux towards production of 4-hydroxypanduratin A. Light green arrows indicate the metabolic flux towards production of and panduratin A. The thickness of the arrows indicate the metabolic flux flow tendency. For metabolic engineer the pathway, black X indicate silencing genes strategy, while tripple black arrows indicate overexpression genes strategy.

Additionally, based on the proposed flavonoid pathway, there are a few suggestions to increase panduratin A production. In this thesis, it is suggested that in order to divert the metabolic flux towards panduratin A production, pinocembrin chalcone must serve as the only substrate to produce pinostrobin chalcone by silencing *ChOMT2*, *PT2* and *CHI* (Figure 5.3). As mentioned previously, since isomerization of pinocembrin chalcone can occur spontaneously without the involvement of *CHI*, it can be suggested to silence *7-OMT* as well in order to prevent the depletion of pinocembrin chalcone as the substrate in panduratin A biosynthetic pathway.

Although silencing of *CHI* and *7-OMT* might increase the accumulation of pinocembrin chalcone, the spontaneous isomerization of pinocembrin chalcone to pinocembrin might hinder the effort to increase the panduratin A production by the accumulation of pinocembrin rather than panduratin A. Therefore, it is suggested to overexpress *ChOMT1* and *PT1* together with the silencing approaches in order to make the metabolic flux towards panduratin A production greater than metabolix flux towards pinocembrin and pinostrobin production, respectively (Figure 5.3).

In addition with the pathway information available, the production of panduratin A can also be achieved through biotransformation of microorganisms by transferring the genes encoding the related enzymes into the system. As mentioned by Kim et al. (2010), the rate of success of engineering of plant pathways in microorganisms is critically dependent on the selection and copy number of expressed genes, substrate supply and optimal fermentation conditions. Since panduratin A is not produced in yeast system in nature, it can be a good alternative host system to produce panduratin A.

It has been reported that the entire β -bitter acid pathway was successfully reconstructed in a yeast system (Li et al., 2015). There were five enzymes co-expressed in mutant DD104 yeast strain which include HICCL2, HICCL4, HIVPS, HIPT1-L and HIPT2. The

mutant DD104 yeast strain was originally designed for monoterpene production by down-regulating the endogenous farnesyl diphosphate which provides large pool of dimethylallyltransferase (DMAPP) for prenylations in β -bitter acid biosynthetic pathway (Fischer et al., 2011). It was reported that up to $2.18 \mu\text{mol L}^{-1}$ of total β -bitter acids were produced in this engineered yeast system (Li et al., 2015). It is certainly hoped that the production of panduratin A can also be successfully increased in the future by employing the information and strategies discussed above.

University of Malaya

CHAPTER 6: CONCLUSION

In conclusion, all objectives have been achieved. Total RNA in *B. rotunda* in response to exogenous phenylalanine treatment was successfully sequenced through RNA-Seq technology. This is the first report of *B. rotunda* transcriptome data to elucidate gene regulation pathways in response to exogenous phenylalanine treatment. The transcriptome data will also enrich the plant database as a reference for other Zingiberaceae family members. Subsequently, transcriptome data between the phenylalanine treated and untreated *B. rotunda* cell suspension cultures also achieved in this study. Through RNA-Seq and DEGs analysis, genes encoding flavonoid enzymes and transcription factors that leads to panduratin A biosynthesis in *B. rotunda* were successfully elucidated.

In addition, the first transcriptome-wide R2R3 MYB transcription factor analysis in *B. rotunda* has been successfully identified and characterized. Based on results obtained, *BrMYB1*, *BrMYB2* and *BrMYB3* were successfully cloned and molecularly characterized. *BrMYB2* and *BrMYB3* encode the repressor R2R3 MYB subgroup 4 protein and bHLH-dependent R2R3 MYB protein, respectively while *BrMYB1* encodes for a R3-MYB related protein.

In this thesis, pathways leading to bioactive compounds including panduratin A in *B. rotunda* are proposed. The hypothesis on the metabolic flux in the proposed pathways is also suggested in the thesis. Through the metabolic flux hypothesis, approaches to increase panduratin A yield were deduced. The approaches includes to overexpress and/or silence the genes encoding related enzymes in the proposed pathway in *B. rotunda*. Alternatively, the production of panduratin A can be achieved through biotransformation of microorganisms by transferring the genes encoding related enzymes to microorganism host such as yeast.

However, in order to achieve that, the proposed biosynthetic pathway must be established first. The establishment of the pathway can be done through functional characterization of the candidate genes that were identified from *B. rotunda* transcriptome data. Substrate-specificity study and enzyme assay are important to verify the functions of the enzymes. Thus, the best genes can be chosen for gene overexpression and/or silencing study.

University of Malaya

REFERENCES

- Abdelwahab, S. I., Mohan, S., Abdulla, M. A., Sukari, M. A., Abdul, A. B., Taha, M. M. E., . . . Lee, K. -H. (2011). The methanolic extract of *Boesenbergia rotunda* (L.) Mansf. and its major compound pinostrobin induces anti-ulcerogenic property *in vivo*: possible involvement of indirect antioxidant action. *Journal of Ethnopharmacology*, *137*(2), 963-970.
- Aharoni, A., De Vos, C., Wein, M., Sun, Z., Greco, R., Kroon, A., . . . O'Connell, A. P. (2001). The strawberry FaMYB1 transcription factor suppresses anthocyanin and flavonol accumulation in transgenic tobacco. *The Plant Journal*, *28*(3), 319-332.
- Akagi, T., Ikegami, A., & Yonemori, K. (2010). DkMyb2 wound-induced transcription factor of persimmon (*Diospyros kaki* Thunb.), contributes to proanthocyanidin regulation. *Planta*, *232*(5), 1045-1059.
- Akagi, T., Ikegami, A., Tsujimoto, T., Kobayashi, S., Sato, A., Kono, A., & Yonemori, K. (2009). DkMyb4 is a Myb transcription factor involved in proanthocyanidin biosynthesis in persimmon fruit. *Plant Physiology*, *151*(4), 2028-2045.
- Akashi, T., Sasaki, K., Aoki, T., Ayabe, S. -I., & Yazaki, K. (2009). Molecular cloning and characterization of a cDNA for pterocarpan 4-dimethylallyltransferase catalyzing the key prenylation step in the biosynthesis of glyceollin, a soybean phytoalexin. *Plant Physiology*, *149*(2), 683-693.
- Albert, N. W., Davies, K. M., Lewis, D. H., Zhang, H., Montefiori, M., Brendolise, C., . . . Schwinn, K. E. (2014). A conserved network of transcriptional activators and repressors regulates anthocyanin pigmentation in eudicots. *The Plant Cell*, *26*(3), 962-980.
- Albert, N. W., Lewis, D. H., Zhang, H., Schwinn, K. E., Jameson, P. E., & Davies, K. M. (2011). Members of an R2R3-MYB transcription factor family in *Petunia* are developmentally and environmentally regulated to control complex floral and vegetative pigmentation patterning. *The Plant Journal*, *65*(5), 771-784.
- Alhassan, A. M., Abdullahi, M. I., Uba, A., & Umar, A. (2014). Prenylation of aromatic secondary metabolites: A new frontier for development of novel drugs. *Tropical Journal of Pharmaceutical Research*, *13*(2), 307-314.
- Al-Obaidi, J. R., Mohd-Yusuf, Y., Tay, T., Mhd-Noh, N., & Othman, R. Y. (2010). Identification of a partial oil palm polygalacturonase-inhibiting protein (*EgPGIP*) gene and its expression during basal stem rot infection caused by *Ganoderma boninense*. *African Journal of Biotechnology*, *9*(46), 7788-7797.

- Ambawat, S., Sharma, P., Yadav, N. R., & Yadav, R. C. (2013). MYB transcription factor genes as regulators for plant responses: an overview. *Physiology and Molecular Biology of Plants*, 19(3), 307-321.
- Annadurai, R. S., Neethiraj, R., Jayakumar, V., Damodaran, A. C., Rao, S. N., Katta, M. A., . . . Niranjan, V. (2013). *De novo* transcriptome assembly (NGS) of *Curcuma longa* L. rhizome reveals novel transcripts related to anticancer and antimalarial terpenoids. *PLOS ONE*, 8(2), e56217.
- Audic, S., & Claverie, J. -M. (1997). The significance of digital gene expression profiles. *Genome Research*, 7(10), 986-995.
- Austin, M. B., & Noel, J. P. (2003). The chalcone synthase superfamily of type III polyketide synthases. *Natural Product Reports*, 20(1), 79-110.
- Avelange-Macherel, M. -H., & Joyard, J. (1998). Cloning and functional expression of *AtCOQ3*, the Arabidopsis homologue of the yeast *COQ3* gene, encoding a methyltransferase from plant mitochondria involved in ubiquinone biosynthesis. *The Plant Journal*, 14(2), 203-213.
- Azuma, A., Kobayashi, S., Mitani, N., Shiraishi, M., Yamada, M., Ueno, T., . . . Koshita, Y. (2008). Genomic and genetic analysis of Myb-related genes that regulate anthocyanin biosynthesis in grape berry skin. *Theoretical and Applied Genetics*, 117(6), 1009-1019.
- Babu, P. V. A., Liu, D., & Gilbert, E. R. (2013). Recent advances in understanding the anti-diabetic actions of dietary flavonoids. *The Journal of Nutritional Biochemistry*, 24(11), 1777-1789.
- Bailey, P. S., & Bailey, C. A. (2000). *Organic chemistry: a brief survey of concepts and applications*. United States of America: Pearson Education.
- Bairoch, A., & Apweiler, R. (2000). The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Research*, 28(1), 45-48.
- Ban, Y., Honda, C., Hatsuyama, Y., Igarashi, M., Bessho, H., & Moriguchi, T. (2007). Isolation and functional analysis of a MYB transcription factor gene that is a key regulator for the development of red coloration in apple skin. *Plant and Cell Physiology*, 48(7), 958-970.

- Barg, R., Sobolev, I., Eilon, T., Gur, A., Chmelnitsky, I., Shabtai, S., . . . Salts, Y. (2005). The tomato early fruit specific gene *Lefsm1* defines a novel class of plant-specific SANT/MYB domain proteins. *Planta*, *221*(2), 197-211.
- Barron, D., & Ibrahim, R. K. (1996). Isoprenylated flavonoids—a survey. *Phytochemistry*, *43*(5), 921-982.
- Baumann, K., Perez-Rodriguez, M., Bradley, D., Venail, J., Bailey, P., Jin, H., . . . Martin, C. (2007). Control of cell and petal morphogenesis by R2R3 MYB transcription factors. *Development*, *134*(9), 1691-1701.
- Becker, W. M., Kleinsmith, L. J., & Hardin, J. (2003). *The world of cells* (5th ed). United States of America: Benjamin Cummings.
- Bhamarapavati, S., Juthapruth, S., Mahachai, W., & Mahady, G. (2006). Antibacterial activity of *Boesenbergia rotunda* (L.) Mansf. and *Myristica fragrans* Houtt. against *Helicobacter pylori*. *Songklanakarin Journal of Science and Technology*, *28*(1), 157-163.
- Boeckler, G. A., Gershenzon, J., & Unsicker, S. B. (2011). Phenolic glycosides of the Salicaceae and their role as anti-herbivore defenses. *Phytochemistry*, *72*(13), 1497-1509.
- Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M. -C., Estreicher, A., Gasteiger, E., . . . Schneider, M. (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Research*, *31*(1), 365-370.
- Bogs, J., Jaffé, F. W., Takos, A. M., Walker, A. R., & Robinson, S. P. (2007). The grapevine transcription factor VvMYBPA1 regulates proanthocyanidin synthesis during fruit development. *Plant Physiology*, *143*(3), 1347-1361.
- Bolognese, C. P., & McGraw, P. (2000). The isolation and characterization in yeast of a gene for Arabidopsis S-adenosylmethionine: phospho-ethanolamine N-methyltransferase. *Plant Physiology*, *124*(4), 1800-1813.
- Borovsky, Y., Oren-Shamir, M., Ovadia, R., De Jong, W., & Paran, I. (2004). The A locus that controls anthocyanin accumulation in pepper encodes a MYB transcription factor homologous to Anthocyanin2 of Petunia. *Theoretical and Applied Genetics*, *109*(1), 23-29.

- Bowman, M. J., Park, W., Bauer, P. J., Udall, J. A., Page, J. T., Raney, J., . . . Campbell, B. T. (2013). RNA-Seq transcriptome profiling of upland cotton (*Gossypium hirsutum* L.) root tissue under water-deficit stress. *PLOS ONE*, 8(12), e82634.
- Bowsher, C., Steer, M. W., & Tobin, A. K. (2008). *Plant biochemistry*. New York, United States of America: Garland Science.
- Brandt, W., Bräuer, L., Günnewich, N., Kufka, J., Rausch, F., Schulze, D., Schulze, E., . . . Wessjohann, L. (2009). Molecular and structural basis of metabolic diversity mediated by prenyldiphosphate converting enzymes. *Phytochemistry*, 70(15-16), 1758-1775.
- Broun, P. (2004). Transcription factors as tools for metabolic engineering in plants. *Current Opinion in Plant Biology*, 7(2), 202-209.
- Bryant, J. P., Provenza, F. D., Pastor, J., Reichardt, P. B., Clausen, T. P., & du Toit, J. T. (1991). Interactions between woody plants and browsing mammals mediated by secondary metabolites. *Annual Review of Ecology and Systematics*, 22, 431-446.
- Buckingham, J., & Munasinghe, V. R. N. (2015) *Dictionary of Flavonoids with CD-ROM*. United States of America: Taylor & Francis Inc.
- Burkill, I. H., Birtwistle, W., Foxworthy, F. W., Scrivenor, J. B., & Watson, J. G. (1966). *A dictionary of the economic products of the Malay peninsula* (Vol. 1): Published on behalf of the governments of Malaysia and Singapore by the Ministry of Agriculture and cooperatives, Kuala Lumpur, Malaysia.
- Bustos, R., Castrillo, G., Linhares, F., Puga, M. I., Rubio, V., Pérez-Pérez, J., . . . Paz-Ares, J. (2010). A central regulatory system largely controls transcriptional activation and repression responses to phosphate starvation in *Arabidopsis*. *PLOS Genetics*, 6(9), e1001102.
- Campbell, M., Hahn, F. M., Poulter, C. D., & Leustek, T. (1998). Analysis of the isopentenyl diphosphate isomerase gene family from *Arabidopsis thaliana*. *Plant Molecular Biology*, 36(2), 323-328.
- Cao, H., Nuruzzaman, M., Xiu, H., Huang, J., Wu, K., Chen, X., . . . Luo, Z. (2015). Transcriptome analysis of methyl jasmonate-elicited *Panax ginseng* adventitious roots to discover putative ginsenoside biosynthesis and transport genes. *International Journal of Molecular Sciences*, 16(2), 3035-3057.

- Cao, Z. -H., Zhang, S. -Z., Wang, R. -K., Zhang, R. -F., & Hao, Y. -J. (2013). Genome wide analysis of the apple MYB transcription factor family allows the identification of *MdoMYB121* gene conferring abiotic stress tolerance in plants. *PLOS ONE*, 8(7), e69955.
- Cavallini, E., Matus, J. T., Finezzo, L., Zenoni, S., Loyola, R., Guzzo, F., . . . Tornielli, G. B. (2015). The phenylpropanoid pathway is controlled at different branches by a set of R2R3-MYB C2 repressors in grapevine. *Plant Physiology*, 167(4), 1448-1470.
- Chagné, D., Lin-Wang, K., Espley, R. V., Volz, R. K., How, N. M., Rouse, S., . . . Allan, A. C. (2013). An ancient duplication of apple MYB transcription factors is responsible for novel red fruit-flesh phenotypes. *Plant Physiology*, 161(1), 225-239.
- Cheah, S. -C., Appleton, D. R., Lee, S. -T., Lam, M. -L., Hadi, A. H. A., & Mustafa, M. R. (2011). Panduratin A inhibits the growth of A549 cells through induction of apoptosis and inhibition of NF-KappaB translocation. *Molecules*, 16(3), 2583-2598.
- Cheah, S. -C., Lai, S. -L., Lee, S. -T., Hadi, A. H. A., & Mustafa, M. R. (2013). Panduratin A, a possible inhibitor in metastasized A549 cells through inhibition of NF-kappa B translocation and chemoinvasion. *Molecules*, 18(8), 8764-8778.
- Chee, C. F., Abdullah, I., Buckle, M. J. C., & Rahman, N. A. (2010). An efficient synthesis of (\pm)-panduratin A and (\pm)-isopanduratin A, inhibitors of dengue-2 viral activity. *Tetrahedron Letters*, 51(3), 495-498.
- Cheenpracha, S., Karalai, C., Ponglimanont, C., Subhadhirasakul, S., & Tewtrakul, S. (2006). Anti-HIV-1 protease activity of compounds from *Boesenbergia pandurata*. *Bioorganic & Medical Chemistry*, 14(6), 1710-1714.
- Chen, R., Liu, X., Zou, J., Yin, Y., Ou, B., Li, J., . . . Dai, J. (2013). Regio- and stereospecific prenylation of flavonoids by *Sophora flavescens* prenyltransferase. *Advanced Synthesis & Catalysis*, 355(9), 1817-1828.
- Chen, Y., Yang, X., He, K., Liu, M., Li, J., Gao, Z., . . . Qu, L. -J. (2006). The MYB transcription factor superfamily of Arabidopsis: expression analysis and phylogenetic comparison with the rice MYB family. *Plant Molecular Biology*, 60(1), 107-124.

- Ching, A. Y. L., Tang, S. W., Sukari, M. A., Lian, G. E. C., Rahmani, M., & Khalid, K. (2007). Characterization of flavonoid derivatives from *Boesenbergia rotunda* (L.). *The Malaysian Journal of Analytical Sciences*, 11(1), 154-159.
- Chiou, C. -Y., & Yeh, K. -W. (2008). Differential expression of MYB gene (*OgMYB1*) determines color patterning in floral tissue of *Oncidium* Gower Ramsey. *Plant Molecular Biology*, 66(4), 379-388.
- Chiu, L. -W., Zhou, X., Burke, S., Wu, X., Prior, R. L., & Li, L. (2010). The purple cauliflower arises from activation of a MYB transcription factor. *Plant Physiology*, 154(3), 1470-1480.
- Choi, Y., Kim, M. S., & Hwang, J. -K. (2012). Inhibitory effects of panduratin A on allergy-related mediator production in rat basophilic leukemia mast cells. *Inflammation*, 35(6), 1904-1915.
- Christensen, A. B., Gregersen, P. L., Olsen, C. E., & Collinge, D. B. (1998). A flavonoid 7-O-methyltransferase is expressed in barley leaves in response to pathogen attack. *Plant Molecular Biology*, 36(2), 219-227.
- Coccolone, S. M., & Cone, K. C. (1993). *Pl-Bh*, an anthocyanin regulatory gene of maize that leads to variegated pigmentation. *Genetics*, 135(2), 575-588.
- Cone, K. C., Burr, F. A., & Burr, B. (1986). Molecular analysis of the maize anthocyanin regulatory locus C1. *Proceedings of the National Academy of Sciences*, 83(24), 9631-9635.
- Cone, K. C., Coccolone, S. M., Moehlenkamp, C. A., Weber, T., Drummond, B. J., Tagliani, L. A., . . . Perrot, G. H. (1993). Role of the regulatory gene *pl* in the photocontrol of maize anthocyanin pigmentation. *The Plant Cell*, 5(12), 1807-1816.
- Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., & Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, 21(18), 3674-3676.
- Cortinovis, C., & Caloni, F. (2015). Alkaloid-containing plants poisonous to cattle and horses in Europe. *Toxins*, 7(12), 5301-5307.
- Cramer, C., Edwards, K., Dron, M., Liang, X., Dildine, S., Bolwell, G. P., . . . Schuch, W. (1989). Phenylalanine ammonia-lyase gene organization and structure. *Plant Molecular Biology*, 12(4), 367-383.

- Cui, M. H., Yoo, K. S., Hyoung, S., Nguyen, H. T. K., Kim, Y. Y., Kim, H. J., . . . Shin, J. S. (2013). An Arabidopsis R2R3-MYB transcription factor, AtMYB20, negatively regulates type 2C serine/threonine protein phosphatases to enhance salt tolerance. *FEBS Letters*, 587(12) 1773-1778.
- Czemmel, S., Stracke, R., Weisshaar, B., Cordon, N., Harris, N. N., Walker, A. R., . . . Bogs, J. (2009). The grapevine R2R3-MYB transcription factor VvMYB1 regulates flavonol synthesis in developing grape berries. *Plant Physiology*, 151(3), 1513-1530.
- De Luca, V., & St Pierre, B. (2000). The cell and developmental biology of alkaloid biosynthesis. *Trends in Plant Science*, 5(4), 168-173.
- De Vos, M., Denekamp, M., Dicke, M., Vuylsteke, M., Van Loon, L., Smeeckens, S. C., & Pieterse, C. (2006). The Arabidopsis thaliana transcription factor AtMYB102 functions in defense against the insect herbivore *Pieris rapae*. *Plant Signaling & Behavior*, 1(6), 305-311.
- Delmulle, L., Bellahcène, A., Dhooge, W., Comhaire, F., Roelens, F., Huvaere, K., . . . De Keukeleire, D. (2006). Anti-proliferative properties of prenylated flavonoids from hops (*Humulus lupulus* L.) in human prostate cancer cell lines. *Phytomedicine*, 13(9), 732-734.
- Deluc, L., Barrieu, F., Marchive, C., Lauvergeat, V., Decendit, A., Richard, T., . . . Hamdi, S. (2006). Characterization of a grapevine R2R3-MYB transcription factor that regulates the phenylpropanoid pathway. *Plant Physiology*, 140(2), 499-511.
- Deluc, L., Bogs, J., Walker, A. R., Ferrier, T., Decendit, A., Merillon, J. -M., . . . Barrieu, F. (2008). The transcription factor VvMYB5b contributes to the regulation of anthocyanin and proanthocyanidin biosynthesis in developing grape berries. *Plant Physiology*, 147(4), 2041-2053.
- Ding, Z., Li, S., An, X., Liu, X., Qin, H., & Wang, D. (2009). Transgenic expression of MYB15 confers enhanced sensitivity to abscisic acid and improved drought tolerance in Arabidopsis thaliana. *Journal of Genetics and Genomics*, 36(1), 17-29.
- Doyle, J. (1991). DNA protocols for plants. In *Molecular techniques in taxonomy* (pp. 283-293). Springer Berlin Heidelberg.
- Du, H., Wang, Y. -B., Xie, Y., Liang, Z., Jiang, S. -J., Zhang, S. -S., . . . Tang, Y. -X. (2013). Genome-wide identification and evolutionary and expression analyses of MYB-related genes in land plants. *DNA Research*, 20(5), 437-448.

- Du, H., Yang, S. -S., Liang, Z., Feng, B. -R., Liu, L., Huang, Y. -B., & Tang, Y. -X. (2012). Genome-wide analysis of the MYB transcription factor superfamily in soybean. *BMC Plant Biology*, *12*(1), 106.
- Du, H., Zhang, L., Liu, L., Tang, X. F., Yang, W. J., Wu, Y. M., . . . Tang, Y. X. (2009). Biochemical and molecular characterization of plant MYB transcription factor family. *Biochemistry (Moscow)*, *74*(1), 1-11.
- Dubos, C., Le Gourrierec, J., Baudry, A., Huet, G., Lanet, E., Debeaujon, I., . . . Lepiniec, L. (2008). MYBL2 is a new regulator of flavonoid biosynthesis in *Arabidopsis thaliana*. *The Plant Journal*, *55*(6), 940-953.
- Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C., & Lepiniec, L. (2010). MYB transcription factors in Arabidopsis. *Trends in Plant Science*, *15*(10), 573-581.
- Dubrovina, A., Kiselev, K., & Zhuravlev, Y. N. (2012). The role of canonical and noncanonical pre-mRNA splicing in plant stress responses. *BioMed Research International*, *2013*, 1-14.
- Ehrling, J., Büttner, D., Wang, Q., Douglas, C. J., Somssich, I. E., & Kombrink, E. (2002). Three 4-coumarate: coenzyme A ligases in *Arabidopsis thaliana* represent two evolutionarily divergent classes in angiosperms. *The Plant Journal*, *19*(1), 9-20.
- Elomaa, P., Uimari, A., Mehto, M., Albert, V. A., Laitinen, R. A., & Teeri, T. H. (2003). Activation of anthocyanin biosynthesis in *Gerbera hybrida* (Asteraceae) suggests conserved protein-protein and protein-promoter interactions between the anciently diverged monocots and eudicots. *Plant Physiology*, *133*(4), 1831-1842.
- Espley, R. V., Hellens, R. P., Putterill, J., Stevenson, D. E., Kutty-Amma, S., & Allan, A. C. (2007). Red colouration in apple fruit is due to the activity of the MYB transcription factor, MdMYB10. *The Plant Journal*, *49*(3), 414-427.
- Ferrer, J. -L., Austin, M., Stewart, C., & Noel, J. (2008). Structure and function of enzymes involved in the biosynthesis of phenylpropanoids. *Plant Physiology and Biochemistry*, *46*(3), 356-370.
- Ferrer, J. L., Jez, J. M., Bowman, M. E., Dixon, R. A., & Noel, J. P. (1999). Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis. *Nature Structural & Molecular Biology*, *6*(8), 775-784.

- Fett-Neto, A. G., Melanson, S. J., Nicholson, S. A., Pennington, J. J., & DiCosmo, F. (1994). Improved taxol yield by aromatic carboxylic acid and amino acid feeding to cell cultures of *Taxus cuspidata*. *Biotechnology and Bioengineering*, *44*(8), 967-971.
- Fett-Neto, A. G., Melanson, S. J., Sakata, K., & DiCosmo, F. (1993). Improved growth and taxol yield in developing calli of *Taxus cuspidata* by medium composition modification. *Nature Biotechnology*, *11*(6), 731-734.
- Fischer, M. J., Meyer, S., Claudel, P., Bergdoll, M., & Karst, F. (2011). Metabolic engineering of monoterpene synthesis in yeast. *Biotechnology and Bioengineering*, *108*(8), 1883-1892.
- Forkmann, G., & Martens, S. (2001). Metabolic engineering and applications of flavonoids. *Current Opinion in Biotechnology*, *12*(2), 155-160.
- Fornalé, S., Shi, X., Chai, C., Encina, A., Irar, S., Capellades, M., . . . Caparrós-Ruiz, D. (2010). ZmMYB31 directly represses maize lignin genes and redirects the phenylpropanoid metabolic flux. *The Plant Journal*, *64*(4), 633-644.
- Frémont, N., Riefler, M., Stolz, A., & Schmülling, T. (2013). The Arabidopsis TUMOR PRONE5 gene encodes an acetylornithine aminotransferase required for arginine biosynthesis and root meristem maintenance in blue light. *Plant Physiology*, *161*(3), 1127-1140.
- Fukasawa-Akada, T., Kung, S. -D., & Watson, J. (1996). Phenylalanine ammonia-lyase gene structure, expression, and evolution in *Nicotiana*. *Plant Molecular Biology*, *30*(4), 711-722.
- Galbiati, M., Matus, J. T., Francia, P., Rusconi, F., Cañón, P., Medina, C., . . . Arce-Johnson, P. (2011). The grapevine guard cell-related VvMYB60 transcription factor is involved in the regulation of stomatal activity and is differentially expressed in response to ABA and osmotic stress. *BMC Plant Biology*, *11*(1), 142.
- Galway, M. E., Masucci, J. D., Lloyd, A. M., Walbot, V., Davis, R. W., & Schiefelbein, J. W. (1994). The *TTG* gene is required to specify epidermal cell fate and cell patterning in the Arabidopsis root. *Developmental Biology*, *166*(2), 740-754.
- Gao, J. -J., Shen, X. -F., Zhang, Z., Peng, R. -H., Xiong, A. -S., Xu, J., . . . Yao, Q. -H. (2011). The myb transcription factor MdMYB6 suppresses anthocyanin biosynthesis in transgenic Arabidopsis. *Plant Cell, Tissue and Organ Culture (PCTOC)*, *106*(2), 235-242.

- García-Lafuente, A., Guillamón, E., Villares, A., Rostagno, M. A., & Martínez, J. A. (2009). Flavonoids as anti-inflammatory agents: implications in cancer and cardiovascular disease. *Inflammation Research*, 58(9), 537-552.
- Garritano, S., Pinto, B., Giachi, I., Pistelli, L., & Reali, D. (2005). Assessment of estrogenic activity of flavonoids from Mediterranean plants using an *in vitro* short-term test. *Phytomedicine*, 12(1), 143-147.
- Geekiyanage, S., Takase, T., Ogura, Y., & Kiyosue, T. (2007). Anthocyanin production by over-expression of grape transcription factor gene *VlmybA2* in transgenic tobacco and Arabidopsis. *Plant Biotechnology Reports*, 1(1), 11-18.
- Geethalakshmi, S., Barathkumar, S., & Prabu, G. (2015). The MYB transcription factor family genes in sugarcane (*Saccharum* sp.). *Plant Molecular Biology Reporter*, 33(3), 512-531.
- Gene Ontology Consortium. (2004). The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Research*, 32(suppl_1), D258-D261.
- Gill, S. S., & Tuteja, N. (2010). Reactive oxygen species and antioxidant machinery in abiotic stress tolerance in crop plants. *Plant Physiology and Biochemistry*, 48(12), 909-930.
- Gonzalez, A., Zhao, M., Leavitt, J. M., & Lloyd, A. M. (2008). Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in Arabidopsis seedlings. *The Plant Journal*, 53(5), 814-827.
- Großkinsky, D. K., van der Graaff, E., & Roitsch, T. (2012). Phytoalexin transgenics in crop protection—Fairy tale with a happy end? *Plant Science*, 195, 54-70.
- Grotewold, E., Drummond, B. J., Bowen, B., & Peterson, T. (1994). The myb-homologous *P* gene controls phlobaphene pigmentation in maize floral organs by directly activating a flavonoid biosynthetic gene subset. *Cell*, 76(3), 543-553.
- Gui, J., Shen, J., & Li, L. (2011). Functional characterization of evolutionarily divergent 4-coumarate: Coenzyme A ligases in rice. *Plant Physiology*, 157(2), 574-586.
- Guo, Q., Ma, X., Wei, S., Qiu, D., Wilson, I. W., Wu, P., . . . Zu, W. (2014). *De novo* transcriptome sequencing and digital gene expression analysis predict biosynthetic pathway of rhynchophylline and isorhynchophylline from *Uncaria rhynchophylla*, a non-model plant with potent anti-alzheimer's properties. *BMC Genomics*, 15(1), 676.

- Han, R., Takahashi, H., Nakamura, M., Bunsupa, S., Yoshimoto, N., Yamamoto, H., . . . Saito, K. (2015). Transcriptome analysis of nine tissues to discover genes involved in the biosynthesis of active ingredients in *Sophora flavescens*. *Biological and Pharmaceutical Bulletin*, 38(6), 876-883.
- Han, S., Wu, Z., Jin, Y., Yang, W., & Shi, H. (2015). RNA-Seq analysis for transcriptome assembly, gene identification, and SSR mining in ginkgo (*Ginkgo biloba* L.). *Tree Genetics & Genomes*, 11(3), 37.
- Hancock, K. R., Collette, V., Fraser, K., Greig, M., Xue, H., Richardson, K., . . . Rasmussen, S. (2012). Expression of the R2R3-MYB transcription factor TaMYB14 from *Trifolium arvense* activates proanthocyanidin biosynthesis in the legumes *Trifolium repens* and *Medicago sativa*. *Plant Physiology*, 159(3), 1204-1220.
- Hartmann, T. (1996). Diversity and variability of plant secondary metabolism: a mechanistic view. *Entomologia Experimentalis et Applicata*, 80(1), 177-188.
- He, M., Wang, Y., Hua, W., Zhang, Y., & Wang, Z. (2012). *De novo* sequencing of *Hypericum perforatum* transcriptome to identify potential genes involved in the biosynthesis of active metabolites. *PLOS ONE*, 7(7), e42081.
- Heil, M. (2011). Nectar: generation, regulation and ecological functions. *Trends in Plant Science*, 16(4), 191-200.
- Heim, K. E., Tagliaferro, A. R., & Bobilya, D. J. (2002). Flavonoid antioxidants: chemistry, metabolism and structure-activity relationships. *The Journal of Nutritional Biochemistry*, 13(10), 572-584.
- Heppel, S. C., Jaffé, F. W., Takos, A. M., Schellmann, S., Rausch, T., Walker, A. R., & Bogs, J. (2013). Identification of key amino acids for the evolution of promoter target specificity of anthocyanin and proanthocyanidin regulating MYB factors. *Plant Molecular Biology*, 82(4-5), 457-471.
- Herms, D. A., & Mattson, W. J. (1992). The dilemma of plants: to grow or defend. *Quarterly Review of Biology*, 283-335.
- Herrmann, K. M., & Weaver, L. M. (1999). The shikimate pathway. *Annual Review of Plant Biology*, 50(1), 473-503.

- Herunsalee, A., Pancharoen, O., & Tuntiwachwuttikul, P. (1987). Further studies of flavonoids of the black rhizomes *Boesenbergia pandurata*. *Journal of Science Society of Thailand*, 13(2), 119–122.
- Hichri, I., Barrieu, F., Bogs, J., Kappel, C., Delrot, S., & Lauvergeat, V. (2011). Recent advances in the transcriptional regulation of the flavonoid biosynthetic pathway. *Journal of Experimental Botany*, 62(8), 2465-2483.
- Himi, E., Yamashita, Y., Haruyama, N., Yanagisawa, T., Maekawa, M., & Taketa, S. (2012). *Ant28* gene for proanthocyanidin synthesis encoding the R2R3 MYB domain protein (Hvmyb10) highly affects grain dormancy in barley. *Euphytica*, 188(1), 141-151.
- Hiraga, S., Sasaki, K., Ito, H., Ohashi, Y., & Matsui, H. (2001). A large family of class III plant peroxidases. *Plant and Cell Physiology*, 42(5), 462-468.
- Hiraga, S., Yamamoto, K., Ito, H., Sasaki, K., Matsui, H., Honma, M., Nagamura, Y., Sasaki, T., & Ohashi, Y. (2000). Diverse expression profiles of 21 rice peroxidase genes. *FEBS Letters*, 471(2), 245-250.
- Hirsch, S., & Oldroyd, G. E. D. (2009). GRAS-domain transcription factors that regulate plant development. *Plant Signal & Behaviour*, 4(8), 698-700.
- Höll, J., Vannozzi, A., Czemplin, S., D'Onofrio, C., Walker, A. R., Rausch, T., . . . Bogs, J. (2013). The R2R3-MYB transcription factors MYB14 and MYB15 regulate stilbene biosynthesis in *Vitis vinifera*. *The Plant Cell*, 25(10), 4135-4149.
- Hou, X. -J., Li, S. -B., Liu, S. -R., Hu, C. -G., & Zhang, J. -Z. (2014). Genome-wide classification and evolutionary and expression analyses of citrus MYB transcription factor families in sweet orange. *PLOS ONE*, 9(11), e112375.
- Hsu, C. -Y., Jenkins, J. N., Saha, S., & Ma, D. -P. (2005). Transcriptional regulation of the lipid transfer protein gene *LTP3* in cotton fibers by a novel MYB protein. *Plant Science*, 168(1), 167-181.
- Hua, W., Zheng, P., He, Y., Cui, L., Kong, W., & Wang, Z. (2014). An insight into the genes involved in secoiridoid biosynthesis in *Gentiana macrophylla* by RNA-seq. *Molecular Biology Reports*, 41(7), 4817-4825.
- Huang, Q., Liang, L., He, R., Ma, X., Zhan, R., & Chen, W. (2015). The first insight into transcriptome profile of herbaceous plant *Nervilia fordii* based on RNA-seq. *Plant Omics*, 8(6), 493-499.

- Huang, W., Sun, W., Lv, H., Luo, M., Zeng, S., Pattanaik, S., Yuan, L., & Wang, Y. (2013). A R2R3-MYB transcription factor from *Epimedium sagittatum* regulates the flavonoid biosynthetic pathway. *PLOS ONE*, 8(8), e70778.
- Huang, Y. F., Vialet, S., Guiraud, J. L., Torregrosa, L., Bertrand, Y., Cheynier, V., . . . Terrier, N. (2014). A negative MYB regulator of proanthocyanidin accumulation, identified through expression quantitative locus mapping in the grape berry. *New Phytologist*, 201(3), 795-809.
- Huang, Y. -J., Song, S., Allan, A. C., Liu, X. -F., Yin, X. -R., Xu, C. -J., & Chen, K. -S. (2013). Differential activation of anthocyanin biosynthesis in Arabidopsis and tobacco over-expressing an R2R3 MYB from Chinese bayberry. *Plant Cell, Tissue and Organ Culture (PCTOC)*, 113(3), 491-499.
- Hwang, J. K., Chung, J. Y., Baek, N. I., & Park, J. H. (2004). Isopanduratin A from *Kaempferia pandurata* as an active antibacterial agent against cariogenic *Streptococcus mutans*. *International Journal of Antimicrobial Agents*, 23(4), 377-381.
- Hyun, T. K., Lee, S., Rim, Y., Kumar, R., Han, X., Lee, S. Y., . . . Kim, J. -Y. (2014). *De-novo* RNA sequencing and metabolite profiling to identify genes involved in anthocyanin biosynthesis in Korean black raspberry (*Rubus coreanus* Miquel). *PLOS ONE*, 9(2), e88292.
- Ibarra-Laclette, E., Zamudio-Hernández, F., Pérez-Torres, C. A., Albert, V. A., Ramírez-Chávez, E., Molina-Torres, J., . . . Herrera-Estrella, L. (2015). *De novo* sequencing and analysis of *Lophophora williamsii* transcriptome, and searching for putative genes involved in mescaline biosynthesis. *BMC Genomics*, 16(1), 657.
- Iriti, M., & Faoro, F. (2009). Chemical diversity and defence metabolism: how plants cope with pathogens and ozone pollution. *International. Journal of Molecular. Sciences*, 10(8), 3371-3399.
- Isa, N., Abdelwahab, S., Mohan, S., Abdul, A., Sukari, M., Taha, M., . . . Mustafa, M. R. (2012). *In vitro* anti-inflammatory, cytotoxic and antioxidant activities of boesenbergin A, a chalcone isolated from *Boesenbergia rotunda* (L.)(fingerroot). *Brazilian Journal of Medical and Biological Research*, 45(6), 524-530.
- Iseli, C., Jongeneel, C. V., & Bucher, P. (1999). *ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences*. Paper presented at the Intelligent Systems for Molecular Biology.

- Ito, M. (2005). Conservation and diversification of three-repeat Myb transcription factors in plants. *Journal of Plant Research*, 118(1), 61-69.
- Jaipetch, T., Kanghae, S., Pancharoen, O., Patrick, V., Reutrakul, V., & Tuntiwachwuttikul, P., White, A. (1982). Constituents of *Boesenbergia pandurata* (syn. *Kaempferia pandurata*): isolation, crystal structure and synthesis of (±)-Boesenbergin A. *Australian Journal of Chemistry*, 35(2), 351-361.
- Jaipetch, T., Reutrakul, V., Tuntiwachwuttikul, P., & Santisuk, T. (1983). Flavonoids in the black rhizomes of *Boesenbergia pandurata*. *Phytochemistry*, 22(2), 625-626.
- Jakoby, M., Weisshaar, B., Dröge-Laser, W., Vicente-Carbajosa, J., Tiedemann, J., Kroj, T., & Parcy, F. (2002). bZIP transcription factors in Arabidopsis. *Trends in Plant Science*, 7(3), 106-111.
- Jin, H., Cominelli, E., Bailey, P., Parr, A., Mehrtens, F., Jones, J., . . . Martin, C. (2000). Transcriptional repression by AtMYB4 controls production of UV-protecting sunscreens in Arabidopsis. *The EMBO Journal*, 19(22), 6150-6161.
- Jones, C. G., & Hartley, S. E. (1999). A protein competition model of phenolic allocation. *Oikos*, 86(1), 27-44.
- Joos, H. J., & Hahlbrock, K. (2005). Phenylalanine ammonia-lyase in potato (*Solanum tuberosum* L.). *European Journal of Biochemistry*, 204(2), 621-629.
- Kagale, S., Links, M. G., & Rozwadowski, K. (2010). Genome-wide analysis of ethylene-responsive element binding factor-associated amphiphilic repression motif-containing transcriptional regulators in Arabidopsis. *Plant Physiology*, 152(3), 1109-1134.
- Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, 28(1), 27-30.
- Karamat, F., Olry, A., Munakata, R., Koeduka, T., Sugiyama, A., Paris, C., . . . Yazaki, K. (2014). A coumarin-specific prenyltransferase catalyzes the crucial biosynthetic reaction for furanocoumarin formation in parsley. *The Plant Journal*, 77(4), 627-638.
- Katiyar, A., Smita, S., Lenka, S. K., Rajwanshi, R., Chinnusamy, V., & Bansal, K. C. (2012). Genome-wide classification and expression analysis of MYB transcription factor families in rice and Arabidopsis. *BMC Genomics*, 13(1), 544.

- Kawai, M., Hirano, T., Higa, S., Arimitsu, J., Maruta, M., Kuwahara, Y., . . . Tanaka, T. (2007). Flavonoids and related compounds as anti-allergic substances. *Allergology International*, 56(2), 113-123.
- Keiler, A. M., Zierau, O., & Kretzschmar, G. (2013). Hop extracts and hop substances in treatment of menopausal complaints. *Planta Medica*, 79(7), 576-579.
- Kiat, T. S., Phippen, R., Yusof, R., Ibrahim, H., Khalid, N., & Rahman, N. A. (2006). Inhibitory activity of cyclohexenyl chalcone derivatives and flavonoids of fingerroot, *Boesenbergia rotunda* (L.), towards dengue-2 virus NS3 protease. *Bioorganic & Medicinal Chemistry Letters*, 16(12), 3337-3340.
- Kim, B. -G., Sung, S. H., Chong, Y., Lim, Y., & Ahn, J. -H. (2010). Plant flavonoid O-methyltransferases: substrate specificity and application. *Journal of Plant Biology*, 53(5), 321-329.
- Kim, D. -Y., Kim, M. -S., Sa, B. -K., Kim, M. -B., & Hwang, J. -K. (2012). *Boesenbergia pandurata* attenuates diet-induced obesity by activating AMP-activated protein kinase and regulating lipid metabolism. *International Journal of Molecular Sciences*, 13(1), 994-1005.
- Kim, D., Lee, M., Jo, K., Lee, K., & Hwang, J. (2011). Therapeutic potential of panduratin A, LKB1-dependent AMP-activated protein kinase stimulator, with activation of PPAR α/δ for the treatment of obesity. *Diabetes, Obesity and Metabolism*, 13(7), 584-593.
- Kim, J. H., Nguyen, N. H., Jeong, C. Y., Nguyen, N. T., Hong, S. -W., & Lee, H. (2013). Loss of the R2R3 MYB, AtMyb73, causes hyper-induction of the *SOS1* and *SOS3* genes in response to high salinity in Arabidopsis. *Journal of Plant Physiology*, 170(16), 1461-1465.
- Kim, M. -S., Pyun, H. -B., & Hwang, J. -K. (2013). Effects of orally administered fingerroot (*Boesenbergia pandurata*) extract on oxazolone-induced atopic dermatitis-like skin lesions in hairless mice. *Food Science and Biotechnology*, 22(1), 257-264.
- Kim, M. -S., Pyun, H. -B., & Hwang, J. -K. (2014). Panduratin A, an activator of PPAR- α/δ , suppresses the development of oxazolone-induced atopic dermatitis-like symptoms in hairless mice. *Life Sciences*, 100(1), 45-54.
- Kim, Y. H., Kim, C. Y., Song, W. K., Park, D. S., Kwon, S. Y., Lee, H. S., . . . Kwak, S. S. (2008). Overexpression of sweetpotato swpa4 peroxidase results in increased

hydrogen peroxide production and enhances stress tolerance in tobacco. *Planta*, 227(4), 867-881.

Kirana, C., Jones, G. P., Record, I. R., & McIntosh, G. H. (2007). Anticancer properties of panduratin A isolated from *Boesenbergia pandurata* (Zingiberaceae). *Journal of Natural Medicines*, 61(2), 131-137.

Klempnauer, K. -H., Gonda, T. J., & Bishop, J. M. (1982). Nucleotide sequence of the retroviral leukemia gene v-myb and its cellular progenitor c-myb: the architecture of a transduced oncogene. *Cell*, 31(2), 453-463.

Kobayashi, S., Goto Yamamoto, N., & Hirochika, H. (2005). Association of *VvmybA1* gene expression with anthocyanin production in grape (*Vitis vinifera*) skin-color mutants. *Journal of the Japanese Society for Horticultural Science (Japan)*, 74(3), 196-203.

Kobayashi, S., Ishimaru, M., Hiraoka, K., & Honda, C. (2002). Myb-related genes of the Kyoho grape (*Vitis labruscana*) regulate anthocyanin biosynthesis. *Planta*, 215(6), 924-933.

Koo, H. J., McDowell, E. T., Ma, X., Greer, K. A., Kapteyn, J., Xie, Z., . . . Gang, D. R. (2013). Ginger and turmeric expressed sequence tags identify signature genes for rhizome identity and development and the biosynthesis of curcuminoids, gingerols and terpenoids. *BMC Plant Biology*, 13(1), 27.

Kranz, H. D., Denekamp, M., Greco, R., Jin, H., Leyva, A., Meissner, R. C., . . . Weisshaar, B. (1998). Towards functional characterisation of the members of the R2R3-MYB gene family from *Arabidopsis thaliana*. *The Plant Journal*, 16(2), 263-276.

Kumar, S., & Pandey, A. K. (2013). Chemistry and biological activities of flavonoids: an overview. *The Scientific World Journal*, 2013, 1-16.

Kumar, S., Kalra, S., Singh, B., Kumar, A., Kaur, J., & Singh, K. (2016). RNA-Seq mediated root transcriptome analysis of *Chlorophytum borivilianum* for identification of genes involved in saponin biosynthesis. *Functional & Integrative Genomics*, 16(1), 37-55.

Kurata, T., Ishida, T., Kawabata-Awai, C., Noguchi, M., Hattori, S., Sano, R., . . . Wada, T. (2005). Cell-to-cell movement of the CAPRICE protein in Arabidopsis root epidermal cell differentiation. *Development*, 132(24), 5387-5398.

- Lai, S. -L., Cheah, S. -C., Wong, P. -F., Noor, S. M., & Mustafa, M. R. (2012). *In vitro* and *in vivo* anti-angiogenic activities of panduratin A. *PLOS ONE*, 7(5), e38103.
- Lai, S. L., Wong, P. F., Lim, T. K., Lin, Q., & Mustafa, M. R. (2015). Cytotoxic mechanisms of Panduratin A on A375 melanoma cells: a quantitative and temporal proteomics analysis. *Proteomics*, 15(9), 1608-1621.
- Laitinen, R. A., Ainasoja, M., Broholm, S. K., Teeri, T. H., & Elomaa, P. (2008). Identification of target genes for a MYB-type anthocyanin regulator in *Gerbera hybrida*. *Journal of Experimental Botany*, 59(13), 3691-3703.
- Larsen, K. (1996). Preliminary checklist of the Zingiberaceae of Thailand. *Thai Forest Bulletin, Botany*, (24), 35-49.
- Lattanzio, V., Cardinali, A., Ruta, C., Fortunato, I. M., Lattanzio, V. M., Linsalata, V., & Cicco, N. (2009). Relationship of secondary metabolism to growth in oregano (*Origanum vulgare* L.) shoot cultures under nutritional stress. *Environmental and Experimental Botany*, 65(1), 54-62.
- Lattanzio, V., Di Venere, D., Linsalata, V., Lima, G., Ippolito, A., & Salerno, M. (1996). Antifungal activity of 2, 5-dimethoxybenzoic acid on postharvest pathogens of strawberry fruits. *Postharvest Biology and Technology*, 9(3), 325-334.
- Latté, K. P., & Kolodziej, H. (2000). Antifungal effects of hydrolysable tannins and related compounds on dermatophytes, mould fungi and yeasts. *Zeitschrift für Naturforschung C*, 55(5-6), 467-472.
- Lee, C. W., Kim, H. S., Kim, H. K., Kim, J. W., Yoon, J. H., Cho, Y., & Hwang, J. K. (2010). Inhibitory effect of panduratin a isolated from *Kaempferia pandurata* Roxb. on melanin biosynthesis. *Phytotherapy Research*, 24(11), 1600-1604.
- Lee, M. M., & Schiefelbein, J. (1999). WEREWOLF, a MYB-related protein in Arabidopsis, is a position-dependent regulator of epidermal cell patterning. *Cell*, 99(5), 473-483.
- Legay, S., Lacombe, E., Goicoechea, M., Brière, C., Séguin, A., Mackay, J., & Grima-Pettenati, J. (2007). Molecular characterization of *EgMYB1*, a putative transcriptional repressor of the lignin biosynthetic pathway. *Plant Science*, 173(5), 542-549.
- Legay, S., Sivadon, P., Blervacq, A. S., Pavy, N., Baghdady, A., Tremblay, L., . . . Grima-Pettenati, J. (2010). EgMYB1, an R2R3 MYB transcription factor from

eucalyptus negatively regulates secondary cell wall formation in Arabidopsis and poplar. *New Phytologist*, 188(3), 774-786.

Lerdau, M., & Coley, P. D. (2002). Benefits of the carbon-nutrient balance hypothesis. *Oikos*, 98(3), 534-536.

Li, C., Ng, C. K. -Y., & Fan, L. -M. (2015). MYB transcription factors, active players in abiotic stress signaling. *Environmental and Experimental Botany*, 114, 80-91.

Li, H., Ban, Z., Qin, H., Ma, L., King, A. J., & Wang, G. (2015). A heteromeric membrane-bound prenyltransferase complex from hop catalyzes three sequential aromatic prenylations in the bitter acid pathway. *Plant Physiology*, 167(3), 650-659.

Li, H., Dong, Y., Yang, J., Liu, X., Wang, Y., Yao, N., . . . Li, X. (2012). *De novo* transcriptome of safflower and the identification of putative genes for oleosin and the biosynthesis of flavonoids. *PLOS ONE*, 7(2), e30987.

Li, J., Chen, R., Wang, R., Liu, X., Xie, D., Zou, J., & Dai, J. (2014). GuA6DT, a regio-specific prenyltransferase from *Glycyrrhiza uralensis*, catalyzes the 6-prenylation of flavones. *ChemBioChem*, 15(11), 1673-1681.

Li, J., Li, X., Guo, L., Lu, F., Feng, X., He, K., . . . Gu, H. (2006). A subgroup of MYB transcription factor genes undergoes highly conserved alternative splicing in Arabidopsis and rice. *Journal of Experimental Botany*, 57(6), 1263-1273.

Li, R., Yu, C., Li, Y., Lam, T. W., Yiu, S. M., Kristiansen, K., & Wang, J. (2009). SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics*, 25(15), 1966-1967.

Li, R., Zhu, H., Ruan, J., Qian, W., Fang, X., Shi, Z., . . . Wang, J. (2010). *De novo* assembly of human genomes with massively parallel short read sequencing. *Genome Research*, 20(2), 265-272.

Li, X. -W., Li, J. -W., Zhai, Y., Zhao, Y., Zhao, X., Zhang, H. -J., . . . Wang, Q. -Y. (2013). A R2R3-MYB transcription factor, GmMYB12B2, affects the expression levels of flavonoid biosynthesis genes encoding key enzymes in transgenic Arabidopsis plants. *Gene*, 532(1), 72-79.

Li, Z. -H., Wang, Q., Ruan, X., Pan, C. -D., & Jiang, D. -A. (2010). Phenolics and plant allelopathy. *Molecules*, 15(12), 8933-8952.

- Listyawati, S., Sismindari, S., Mubarika, S., & Murti, Y.B. (2015). Protective effect ethanolic extract of *Boesenbergia pandurata* (ROXB.) Schlecht. against UVB-induced DNA damages in Balb/c mice. *Indonesian Journal of Pharmacy*, 26(2), 108-113.
- Liu, A. -L., Wang, H. -D., Lee, S. M., Wang, Y. -T., & Du, G. -H. (2008). Structure–activity relationship of flavonoids as influenza virus neuraminidase inhibitors and their *in vitro* anti-viral activities. *Bioorganic & Medicinal Chemistry*, 16(15), 7141-7147.
- Liu, J., Osbourn, A., & Ma, P. (2015). MYB transcription factors as regulators of phenylpropanoid metabolism in plants. *Molecular Plant*, 8(5), 689-708.
- Lois, R., Dietrich, A., Hahlbrock, K., & Schulz, W. (1989). A phenylalanine ammonia-lyase gene from parsley: structure, regulation and identification of elicitor and light responsive cis-acting elements. *EMBO Journal*, 8(6), 1641-1648.
- Loke, K. -K., Rahnamaie-Tajadod, R., Yeoh, C. -C., Goh, H. -H., Mohamed-Hussein, Z. -A., Noor, N. M., . . . Ismail, I. (2016). RNA-seq analysis for secondary metabolite pathway gene discovery in *Polygonum minus*. *Genomics Data*, 7, 12-13.
- Lu, J., Du, Z. -X., Kong, J., Chen, L. -N., Qiu, Y. -H., Li, G. -F., . . . Zhu, S. -F. (2012). Transcriptome analysis of *Nicotiana tabacum* infected by *Cucumber mosaic virus* during systemic symptom development. *PLOS ONE*, 7(8), e43447.
- Lu, S. X., Knowles, S. M., Andronis, C., Ong, M. S., & Tobin, E. M. (2009). CIRCADIAN CLOCK ASSOCIATED1 and LATE ELONGATED HYPOCOTYL function synergistically in the circadian clock of *Arabidopsis*. *Plant Physiology*, 150(2), 834-843.
- Luo, H., Sun, C., Sun, Y., Wu, Q., Li, Y., Song, J., . . . Chen, S. (2011). Analysis of the transcriptome of *Panax notoginseng* root uncovers putative triterpene saponin-biosynthetic genes and genetic markers. *BMC Genomics*, 12(5), S5.
- Ma, Q. -H., Wang, C., & Zhu, H. -H. (2011). TaMYB4 cloned from wheat regulates lignin biosynthesis through negatively controlling the transcripts of both cinnamyl alcohol dehydrogenase and cinnamoyl-CoA reductase genes. *Biochimie*, 93(7), 1179-1186.
- Mahidol, C., Tuntiwachwuttikul, P., Reutrakul, V., & Taylor, W. (1984). Constituents of *Boesenbergia pandurata* (syn. *Kaempferia pandurata*). III. Isolation and synthesis of (±)-boesenbergin B. *Australian Journal of Chemistry*, 37(8), 1739-1745.

- Mahmood, A., Mariod, A. A., Abdelwahab, S. I., Ismail, S., & Al-Bayaty, F. (2010). Potential activity of ethanolic extract of *Boesenbergia rotunda* (L.) rhizomes extract in accelerating wound healing in rats. *Journal of Medicinal Plants Research*, 4(15), 1570-1576.
- Marjamaa, K., Kukkola, E. M., & Fagerstedt, K. V. (2009). The role of xylem class III peroxidases in lignification. *Journal of Experimental Botany*, 60(2), 367-376.
- Mathews, H., Clendennen, S. K., Caldwell, C. G., Liu, X. L., Connors, K., Matheis, N., . . . Wagner, D. R. (2003). Activation tagging in tomato identifies a transcriptional regulator of anthocyanin biosynthesis, modification, and transport. *The Plant Cell*, 15(8), 1689-1703.
- Matsui, K., Umemura, Y., & Ohme-Takagi, M. (2008). AtMYBL2, a protein with a single MYB domain, acts as a negative regulator of anthocyanin biosynthesis in Arabidopsis. *The Plant Journal*, 55(6), 954-967.
- Matus, J. T., Aquea, F., & Arce-Johnson, P. (2008). Analysis of the grape MYB R2R3 subfamily reveals expanded wine quality-related clades and conserved gene structure organization across Vitis and Arabidopsis genomes. *BMC Plant Biology*, 8(1), 83.
- McCarthy, R. L., Zhong, R., & Ye, Z. -H. (2009). MYB83 is a direct target of SND1 and acts redundantly with MYB46 in the regulation of secondary cell wall biosynthesis in Arabidopsis. *Plant and Cell Physiology*, 50(11), 1950-1964.
- Mehrtens, F., Kranz, H., Bednarek, P., & Weisshaar, B. (2005). The Arabidopsis transcription factor MYB12 is a flavonol-specific regulator of phenylpropanoid biosynthesis. *Plant Physiology*, 138(2), 1083-1096.
- Mellway, R. D., Tran, L. T., Prouse, M. B., Campbell, M. M., & Constabel, C. P. (2009). The wound-, pathogen-, and ultraviolet B-responsive MYB134 gene encodes an R2R3 MYB transcription factor that regulates proanthocyanidin synthesis in poplar. *Plant Physiology*, 150(2), 924-941.
- Middleton Jr, E., Kandaswami, C., & Theoharides, T. (2000). The effects of plant flavonoids on mammalian cells: implications for inflammation, heart disease, and cancer. *Pharmacological Reviews*, 52(4), 673-751.
- Mierziak, J., Kostyn, K., & Kulma, A. (2014). Flavonoids as important molecules of plant interactions with the environment. *Molecules*, 19(10), 16240-16265.

- Mikusanti, Jenie, B. S. L., Priosoeryanto, B. P., Syarief, R., & Rekso, G. T. (2008). Mode of action Temu Kunci (*Kaempferia pandurata*) essential oil on *E. coli* K1. 1 cell determined by leakage of material cell and salt tolerance assays. *HAYATI Journal of Biosciences*, 15(2), 56-60.
- Milligan, S., Kalita, J., Pocock, V., Van De Kauter, V., Stevens, J., Deinzer, M., . . . De Keukeleire, D. (2000). The endocrine activities of 8-prenylnaringenin and related hop (*Humulus lupulus* L.) flavonoids. *The Journal of Clinical Endocrinology & Metabolism*, 85(12), 4912-4915.
- Miranda, C. L., Stevens, J. F., Ivanov, V., McCall, M., Frei, B., Deinzer, M. L., & Buhler, D. R. (2000). Antioxidant and prooxidant actions of prenylated and nonprenyated chalcones and flavanones *in vitro*. *Journal of Agricultural and Food Chemistry*, 48(9), 3876-3884.
- Miranda, C., Stevens, J., Helmrich, A., Henderson, M., Rodriguez, R., Yang, Y. -H., . . . Buhler, D. (1999). Antiproliferative and cytotoxic effects of prenylated flavonoids from hops (*Humulus lupulus*) in human cancer cell lines. *Food and Chemical Toxicology*, 37(4), 271-285.
- Mizoi, J., Shinozaki, K., & Yamaguchi-Shinozaki, K. (2012). AP2/ERF family transcription factors in plant abiotic stress responses. *Biochimica et Biophysica Acta (BBA)-Regulatory Mechanisms*, 1819(2), 86-96.
- Molina, C., Zaman-Allah, M., Khan, F., Fatnassi, N., Horres, R., Rotter, B., . . . Kahl, G. (2011). The salt-responsive transcriptome of chickpea roots and nodules via deepSuperSAGE. *BMC Plant Biology*, 11(1), 31.
- Mongkolsuk, S., & Dean, F. M. (1964). Pinostrobin and alpinetin from *Kaempferia pandurata*. *Journal of Chemical Society*, 4654-4655.
- Morikawa, T., Funakoshi, K., Ninomiya, K., Yasuda, D., Miyagawa, K., Matsuda, H., & Yoshikawa, M. (2008). Medicinal foodstuffs. XXXIV. Structures of new prenylchalcones and prenylflavanones with TNF-alpha and aminopeptidase N inhibitory activities from *Boesenbergia rotunda*. *Chemical and Pharmaceutical Bulletin*, 56(7), 956-962.
- Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., & Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*, 5(7), 621-628.

- Munakata, R., Inoue, T., Koeduka, T., Karamat, F., Olry, A., Sugiyama, A., . . . Yazaki, K. (2014). Molecular cloning and characterization of a geranyl diphosphate-specific aromatic prenyltransferase from lemon. *Plant Physiology*, 166(1), 80-90.
- Murashige, T., & Skoog, F. (1962). A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiologia Plantarum*, 15(3), 473-497.
- Muriira, N. G., Xu, W., Muchugi, A., Xu, J., & Liu, A. (2015). *De novo* sequencing and assembly analysis of transcriptome in the Sodom apple (*Calotropis gigantea*). *BMC Genomics*, 16(1), 723.
- Nagegowda, D. A. (2010). Plant volatile terpenoid metabolism: biosynthetic genes, transcriptional regulation and subcellular compartmentation. *FEBS Letters*, 584(14), 2965-2973.
- Nakatsuka, T., Saito, M., Yamada, E., Fujita, K., Kakizaki, Y., & Nishihara, M. (2012). Isolation and characterization of GtMYBP3 and GtMYBP4, orthologues of R2R3-MYB transcription factors that regulate early flavonoid biosynthesis, in gentian flowers. *Journal of Experimental Botany*, 63(18), 6505-6517.
- Nesi, N., Jond, C., Debeaujon, I., Caboche, M., & Lepiniec, L. (2001). The Arabidopsis TT2 gene encodes an R2R3 MYB domain protein that acts as a key determinant for proanthocyanidin accumulation in developing seed. *The Plant Cell*, 13(9), 2099-2114.
- Newman, L. J., Perazza, D. E., Juda, L., & Campbell, M. M. (2004). Involvement of the R2R3-MYB, AtMYB61, in the ectopic lignification and dark-photomorphogenic components of the det3 mutant phenotype. *The Plant Journal*, 37(2), 239-250.
- Niu, S. -S., Xu, C. -J., Zhang, W. -S., Zhang, B., Li, X., Lin-Wang, K., . . . Chen, K. -S. (2010). Coordinated regulation of anthocyanin biosynthesis in Chinese bayberry (*Myrica rubra*) fruit by a R2R3 MYB transcription factor. *Planta*, 231(4), 887-899.
- Noda, K. -I., Glover, B. J., Linstead, P., & Martin, C. (1994). Flower colour intensity depends on specialized cell shape controlled by a Myb-related transcription factor. *Nature*, 369(6482), 661-664.
- Ogata, K., Hojo, H., Aimoto, S., Nakai, T., Nakamura, H., Sarai, A., . . . Nishimura, Y. (1992). Solution structure of a DNA-binding unit of Myb: a helix-turn-helix-related motif with conserved tryptophans forming a hydrophobic core. *Proceedings of the National Academy of Sciences*, 89(14), 6428-6432.

- Ohara, K., Yamamoto, K., Hamamoto, M., Sasaki, K., & Yazaki, K. (2006). Functional characterization of *OsPPT1*, which encodes p-Hydroxybenzoate polyprenyltransferase involved in ubiquinone biosynthesis in *Oryza sativa*. *Plant and Cell Physiology*, 47(5), 581-590.
- Ohsaki, A., Yokoyama, R., Miyatake, H., & Fukuyama, Y. (2006). Two diterpene rhamnosides, mimosasides B and C, from *Mimosa hostilis*. *Chemical and Pharmaceutical Bulletin*, 54(12), 1728-1729.
- Okada, K., Ohara, K., Yazaki, K., Nozaki, K., Uchida, N., Kawamukai, M., . . . Yamane, H. (2004). The *AtPPT1* gene encoding 4-hydroxybenzoate polyprenyl diphosphate transferase in ubiquinone biosynthesis is required for embryo development in *Arabidopsis thaliana*. *Plant Molecular Biology*, 55(4), 567-577.
- Okada, Y., Sano, Y., Kaneko, T., Abe, I., Noguchi, H., & Ito, K. (2004). Enzymatic reactions by five chalcone synthase homologs from hop (*Humulus lupulus* L.). *Bioscience, Biotechnology, and Biochemistry*, 68(5), 1142-1145.
- Olsen, A. N., Ernst, H. A., Leggio, L. L., & Skriver, K. (2005). NAC transcription factors: structurally distinct, functionally diverse. *Trends in Plant Science*, 10(2), 79-87.
- Omer, S., Kumar, S., & Khan, B. M. (2013). Over-expression of a subgroup 4 R2R3 type MYB transcription factor gene from *Leucaena leucocephala* reduces lignin content in transgenic tobacco. *Plant Cell Reports*, 32(1), 161-171.
- Oppenheimer, D. G., Herman, P. L., Sivakumaran, S., Esch, J., & Marks, M. D. (1991). A myb gene required for leaf trichome differentiation in *Arabidopsis* is expressed in stipules. *Cell*, 67(3), 483-493.
- Pabo, C. O., & Sauer, R. T. (1992). Transcription factors: structural families and principles of DNA recognition. *Annual Review of Biochemistry*, 61(1), 1053-1095.
- Palapol, Y., Ketsa, S., Lin-Wang, K., Ferguson, I. B., & Allan, A. C. (2009). A MYB transcription factor regulates anthocyanin biosynthesis in mangosteen (*Garcinia mangostana* L.) fruit during ripening. *Planta*, 229(6), 1323-1334.
- Pancharoen, O., Picker, K., Reutrakul, V., Taylor, W., & Tuntiwachwuttikul, P. (1987). Constituents of the Zingiberaceae. X: diastereomers of [7-hydroxy-5-methoxy-2-methyl-2-(4'-methylpent-3'-enyl)-2H-chromen-8-yl][3''-methyl-2'-(3'''-methylbut-2'''-enyl)-6''-phenylcyclohex-3''-enyl] methanone (panduratin B), a constituent of the red rhizomes of a variety of *Boesenbergia pandurata*. *Australian Journal of Chemistry*, 40(3), 455-459.

- Park, J. -S., Kim, J. -B., Cho, K. -J., Cheon, C. -I., Sung, M. -K., Choung, M. -G., & Roh, K. -H. (2008). Arabidopsis R2R3-MYB transcription factor AtMYB60 functions as a transcriptional repressor of anthocyanin biosynthesis in lettuce (*Lactuca sativa*). *Plant Cell Reports*, 27(6), 985-994.
- Park, K. -M., Choo, J. -H., Sohn, J. -H., Lee, S. -H., & Hwang, J. -K. (2005). Antibacterial activity of panduratin A isolated from *Kaempferia pandurata* against *Porphyromonas gingivalis*. *Food Science and Biotechnology*, 14(2), 286-289.
- Pattanaik, S., Kong, Q., Zaitlin, D., Werkman, J. R., Xie, C. H., Patra, B., & Yuan, L. (2010). Isolation and functional characterization of a floral tissue-specific R2R3 MYB regulator from tobacco. *Planta*, 231(5), 1061-1076.
- Payne, C. T., Zhang, F., & Lloyd, A. M. (2000). GL3 encodes a bHLH protein that regulates trichome development in Arabidopsis through interaction with GL1 and TTG1. *Genetics*, 156(3), 1349-1362.
- Paz-Ares, J., Ghosal, D., Wienand, U., Peterson, P., & Saedler, H. (1987). The regulatory *cl* locus of *Zea mays* encodes a protein with homology to myb proto-oncogene products and with structural similarities to transcriptional activators. *The EMBO Journal*, 6(12), 3553-3558.
- Perez-Rodriguez, M., Jaffe, F. W., Butelli, E., Glover, B. J., & Martin, C. (2005). Development of three different cell types is associated with the activity of a specific MYB transcription factor in the ventral petal of *Antirrhinum majus* flowers. *Development*, 132(2), 359-370.
- Perez-Rodriguez, P., Riano-Pachon, D. M., Correa, L. G. G., Rensing, S. A., Kersten, B., & Mueller-Roeber, B. (2009). PlnTFDB: updated content and new features of the plant transcription factor database. *Nucleic Acids Research*, 38(suppl_1), D822-D827.
- Perry, L. M., & Metzger, J. (1980). *Medicinal plants of East and Southeast Asia: attributed properties and uses*. London: MIT Press.
- Pertea, G., Huang, X., Liang, F., Antonescu, V., Sultana, R., Karamycheva, S., . . . Quackenbush, J. (2003). TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics*, 19(5), 651-652.
- Porter, P., Banwart, W., & Hassett, J. (1986). Phenolic acids and flavonoids in soybean root and leaf extracts. *Environmental and Experimental Botany*, 26(1), 65-73.

- Prabu, G., Kavar, P. G., Pagariya, M. C., & Prasad, D. T. (2011). Identification of water deficit stress upregulated genes in sugarcane. *Plant Molecular Biology Reporter*, 29(2), 291-304.
- Preston, J., Wheeler, J., Heazlewood, J., Li, S. F., & Parish, R. W. (2004). AtMYB32 is required for normal pollen development in *Arabidopsis thaliana*. *The Plant Journal*, 40(6), 979-995.
- Pruitt, K. D., Tatusova, T., & Maglott, D. R. (2006). NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research*, 35(suppl_1), D61-D65.
- Quattrocchio, F., Wing, J. F., Va, K., Mol, J. N., & Koes, R. (1998). Analysis of bHLH and MYB domain proteins: species-specific regulatory differences are caused by divergent evolution of target anthocyanin genes. *The Plant Journal*, 13(4), 475-488.
- Quattrocchio, F., Wing, J., van der Woude, K., Souer, E., de Vetten, N., Mol, J., & Koes, R. (1999). Molecular analysis of the anthocyanin2 gene of petunia and its role in the evolution of flower color. *The Plant Cell*, 11(8), 1433-1444.
- Ratty, A., & Das, N. (1988). Effects of flavonoids on nonenzymatic lipid peroxidation: structure-activity relationship. *Biochemical Medicine and Metabolic Biology*, 39(1), 69-79.
- Ravaglia, D., Espley, R. V., Henry-Kirk, R. A., Andreotti, C., Ziosi, V., Hellens, R. P., . . . Allan, A. C. (2013). Transcriptional regulation of flavonoid biosynthesis in nectarine (*Prunus persica*) by a set of R2R3 MYB transcription factors. *BMC Plant Biology*, 13(1), 68.
- Rerie, W. G., Feldmann, K. A., & Marks, M. D. (1994). The GLABRA2 gene encodes a homeo domain protein required for normal trichome development in Arabidopsis. *Genes & Development*, 8(12), 1388-1399.
- Rukayadi, Y., Han, S., Yong, D., & Hwang, J. -K. (2010). *In vitro* antibacterial activity of panduratin A against enterococci clinical isolates. *Biological and Pharmaceutical Bulletin*, 33(9), 1489-1493.
- Rukayadi, Y., Lee, K., Han, S., Yong, D., & Hwang, J. K. (2009). *In vitro* activities of panduratin A against clinical *Staphylococcus* strains. *Antimicrobial Agents and Chemotherapy*, 53(10), 4529-4532.

- Salama, S. M., Abdulla, M. A., AlRashdi, A. S., & Hadi, A. H. A. (2013a). Mechanism of hepatoprotective effect of *Boesenbergia rotunda* in thioacetamide-induced liver damage in rats. *Evidence-Based Complementary and Alternative Medicine*, 2013, 1-13.
- Salama, S. M., AlRashdi, A. S., Abdulla, M. A., Hassandarvish, P., & Bilgen, M. (2013b). Protective activity of panduratin A against thioacetamide-induced oxidative damage: demonstration with in vitro experiments using WRL-68 liver cell line. *BMC Complementary and Alternative Medicine*, 13(1), 279.
- Salama, S. M., Bilgen, M., Al Rashdi, A. S., & Abdulla, M. A. (2012). Efficacy of *Boesenbergia rotunda* treatment against thioacetamide-induced liver cirrhosis in a rat model. *Evidence-Based Complementary and Alternative Medicine*, 2012, 1-12.
- Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989). *Molecular cloning* (Vol. 2). New York: Cold Spring Harbor Laboratory Press.
- Saralamp, P., Chuakul, W., Temsiririrkkul, R., & Clayton, T. (1996). *Medicinal plants in Thailand* (Vol. 1). Bangkok: Amarin Printing and Publishing Public Co.
- Sasaki, K., Mito, K., Ohara, K., Yamamoto, H., & Yazaki, K. (2008). Cloning and characterization of naringenin 8-prenyltransferase, a flavonoid-specific prenyltransferase of *Sophora flavescens*. *Plant Physiology*, 146(3), 1075-1084.
- Sasaki, K., Tsurumaru, Y., Yamamoto, H., & Yazaki, K. (2011). Molecular characterization of a membrane-bound prenyltransferase specific for isoflavone from *Sophora flavescens*. *Journal of Biological Chemistry*, 286(27), 24125-24134.
- Schaart, J. G., Dubos, C., Romero De La Fuente, I., Houwelingen, A. M., Vos, R. C., Jonker, H. H., . . . Bovy, A. G. (2013). Identification and characterization of MYB-bHLH-WD40 regulatory complexes controlling proanthocyanidin biosynthesis in strawberry (*Fragaria* × *ananassa*) fruits. *New Phytologist*, 197(2), 454-467.
- Schellmann, S., Hülkamp, M., & Uhrig, J. (2007). Epidermal pattern formation in the root and shoot of *Arabidopsis*. *Biochemical Society Transactions*, 35(1), 146-148.
- Schellmann, S., Schnittger, A., Kirik, V., Wada, T., Okada, K., Beermann, A., . . . Hülkamp, M. (2002). TRIPTYCHON and CAPRICE mediate lateral inhibition during trichome and root hair patterning in *Arabidopsis*. *The EMBO Journal*, 21(19), 5036-5046.

- Schwinn, K., Venail, J., Shang, Y., Mackay, S., Alm, V., Butelli, E., . . . Martin, C. (2006). A small family of MYB-regulatory genes controls floral pigmentation intensity and patterning in the genus *Antirrhinum*. *The Plant Cell*, *18*(4), 831-851.
- Seelinger, G., Merfort, I., Wölfl, U., & Schempp, C. M. (2008). Anti-carcinogenic effects of the flavonoid luteolin. *Molecules*, *13*(10), 2628-2651.
- Seo, P. J., Xiang, F., Qiao, M., Park, J. -Y., Lee, Y. N., Kim, S. -G., . . . Park, C. -M. (2009). The MYB96 transcription factor mediates abscisic acid signaling during drought stress response in *Arabidopsis*. *Plant Physiology*, *151*(1), 275-289.
- Serna, L. (2008). CAPRICE positively regulates stomatal formation in the *Arabidopsis* hypocotyl. *Plant Signaling & Behavior*, *3*(12), 1077-1082.
- Shelton, D., Stranne, M., Mikkelsen, L., Pakseresht, N., Welham, T., Hiraka, H., . . . Bailey, P. (2012). Transcription factors of Lotus: regulation of isoflavonoid biosynthesis requires coordinated changes in transcription factor activity. *Plant Physiology*, *159*(2), 531-547.
- Shen, G., Huhman, D., Lei, Z., Snyder, J., Sumner, L. W., & Dixon, R. A. (2012). Characterization of an isoflavonoid-specific prenyltransferase from *Lupinus albus*. *Plant Physiology*, *159*(1), 70-80.
- Shen, H., He, X., Poovaiah, C. R., Wuddineh, W. A., Ma, J., Mann, D. G., . . . Dixon, R. A. (2012). Functional characterization of the switchgrass (*Panicum virgatum*) R2R3-MYB transcription factor PvMYB4 for improvement of lignocellulosic feedstocks. *New Phytologist*, *193*(1), 121-136.
- Shim, J. -S., Han, Y. -S., & Hwang, J. -K. (2009). The effect of 4-hydroxypanduratin A on the mitogen-activated protein kinase-dependent activation of matrix metalloproteinase-1 expression in human skin fibroblasts. *Journal of Dermatological Science*, *53*(2), 129-134.
- Shim, J. S., Kwon, Y. Y., Han, Y. S., & Hwang, J. K. (2008a). Inhibitory effect of panduratin A on UV-induced activation of mitogen-activated protein kinases (MAPKs) in dermal fibroblast cells. *Planta Medica*, *74*(12), 1446-1450.
- Shim, J. S., Kwon, Y. Y., & Hwang, J. K. (2008b). The effects of panduratin A isolated from *Kaempferia pandurata* on the expression of matrix metalloproteinase-1 and type-1 procollagen in human skin fibroblasts. *Planta Medica*, *74*(3), 239-244.

- Shin, B., Choi, G., Yi, H., Yang, S., Cho, I., Kim, J., . . . Choi, G. (2002). AtMYB21, a gene encoding a flower-specific transcription factor, is regulated by COP1. *The Plant Journal*, 30(1), 23-32.
- Shin, D., Moon, S. -J., Han, S., Kim, B. -G., Park, S. R., Lee, S. -K., . . . Byun, M. -O. (2011). Expression of StMYB1R-1, a novel potato single MYB-like domain transcription factor, increases drought tolerance. *Plant Physiology*, 155(1), 421-432.
- Shindo, K., Kato, M., Kinoshita, A., Kobayashi, A., & Koike, Y. (2006). Analysis of antioxidant activities contained in the *Boesenbergia pandurata* Schult. rhizome. *Bioscience, Biotechnology and Biochemistry*, 70(9), 2281-2284.
- Simon, M., Lee, M. M., Lin, Y., Gish, L., & Schiefelbein, J. (2007). Distinct and overlapping roles of single-repeat MYB genes in root epidermal patterning. *Developmental Biology*, 311(2), 566-578.
- Sohn, J. H., Han, K. L., Lee, S. H., & Hwang, J. K. (2005). Protective effects of panduratin A against oxidative damage of tert-butylhydroperoxide in human HepG2 cells. *Biological and Pharmaceutical Bulletin*, 28(6), 1083-1086.
- Sonbol, F. -M., Fornalé, S., Capellades, M., Encina, A., Touriño, S., Torres, J. -L., . . . Caparrós-Ruiz, D. (2009). The maize *ZmMYB42* represses the phenylpropanoid pathway and affects the cell wall structure, composition and degradability in *Arabidopsis thaliana*. *Plant Molecular Biology*, 70(3), 283-296.
- Song, M. -S., Shim, J. -S., Gwon, S. -H., Lee, C. -W., Kim, H. -S., & Hwang, J. -K. (2008). Antibacterial activity of panduratin A and isopanduratin A isolated from *Kaempferia pandurata* Roxb. against acne-causing microorganisms. *Food Science and Biotechnology*, 17(6), 1357-1360.
- Steiner-Lange, S., Unte, U. S., Eckstein, L., Yang, C., Wilson, Z. A., Schmelzer, E., . . . Saedler, H. (2003). Disruption of *Arabidopsis thaliana* MYB26 results in male sterility due to non-dehiscent anthers. *The Plant Journal*, 34(4), 519-528.
- Stracke, R., Ishihara, H., Huep, G., Barsch, A., Mehrtens, F., Niehaus, K., & Weisshaar, B. (2007). Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the *Arabidopsis thaliana* seedling. *The Plant Journal*, 50(4), 660-677.
- Stracke, R., Werber, M., & Weisshaar, B. (2001). The R2R3-MYB gene family in *Arabidopsis thaliana*. *Current Opinion in Plant Biology*, 4(5), 447-456.

- Subramaniam, S., Mathiyalagan, R., Gyo, I. J., Bum-Soo, L., Sungyoung, L., & Chun, Y. D. (2011). Transcriptome profiling and insilico analysis of *Gynostemma pentaphyllum* using a next generation sequencer. *Plant Cell Reports*, 30(11), 2075-2083.
- Sun, C., Li, Y., Wu, Q., Luo, H., Sun, Y., Song, J., . . . Chen, S. (2010). *De novo* sequencing and analysis of the American ginseng root transcriptome using a GS FLX Titanium platform to discover putative genes involved in ginsenoside biosynthesis. *BMC Genomics*, 11(1), 262.
- Sun, Y., Luo, H., Li, Y., Sun, C., Song, J., Niu, Y., . . . Chen, S. (2011). Pyrosequencing of the *Camptotheca acuminata* transcriptome reveals putative genes involved in camptothecin biosynthesis and transport. *BMC Genomics*, 12(1), 533.
- Suo, J., Liang, X., Pu, L., Zhang, Y., & Xue, Y. (2003). Identification of *GhMYB109* encoding a R2R3 MYB transcription factor that expressed specifically in fiber initials and elongating fibers of cotton (*Gossypium hirsutum* L.). *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression*, 1630(1), 25-34.
- Tahvanainen, J., Helle, E., Julkunen-Tiitto, R., & Lavola, A. (1985). Phenolic compounds of willow bark as deterrents against feeding by mountain hare. *Oecologia*, 65(3), 319-323.
- Takos, A. M., Jaffé, F. W., Jacob, S. R., Bogs, J., Robinson, S. P., & Walker, A. R. (2006). Light-induced expression of a MYB gene regulates anthocyanin biosynthesis in red apples. *Plant Physiology*, 142(3), 1216-1232.
- Tamagnone, L., Merida, A., Parr, A., Mackay, S., Culianez-Macia, F. A., Roberts, K., & Martin, C. (1998). The AmMYB308 and AmMYB330 transcription factors from *Antirrhinum* regulate phenylpropanoid and lignin biosynthesis in transgenic tobacco. *The Plant Cell*, 10(2), 135-154.
- Tan, E. C., Karsani, S. A., Foo, G. T., Wong, S. M., Rahman, N. A., Khalid, N., . . . Yusof, R. (2012). Proteomic analysis of cell suspension cultures of *Boesenbergia rotunda* induced by phenylalanine: identification of proteins involved in flavonoid and phenylpropanoid biosynthesis pathways. *Plant Cell, Tissue and Organ Culture*, 111(2), 219-229.
- Tan, S. K., (2005). *Flavonoids from Boesenbergia rotunda (L.) Mansf.: chemistry, bioactivity and accumulation* (Doctoral dissertation). University of Malaya.

- Tang, Q., Ma, X., Mo, C., Wilson, I. W., Song, C., Zhao, H., . . . Qiu, D. (2011). An efficient approach to finding *Siraitia grosvenorii* triterpene biosynthetic genes by RNA-seq and digital gene expression analysis. *BMC Genomics*, *12*(1), 343.
- Tao, L., Zhao, Y., Wu, Y., Wang, Q., Yuan, H., Zhao, L., . . . You, X. (2016). Transcriptome profiling and digital gene expression by deep sequencing in early somatic embryogenesis of endangered medicinal *Eleutherococcus senticosus* Maxim. *Gene*, *578*(1), 17-24.
- Tatusov, R. L., Galperin, M. Y., Natale, D. A., & Koonin, E. V. (2000). The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Research*, *28*(1), 33-36.
- Terrier, N., Torregrosa, L., Ageorges, A., Vialet, S., Verriès, C., Cheynier, V., & Romieu, C. (2009). Ectopic expression of VvMybPA2 promotes proanthocyanidin biosynthesis in grapevine and suggests additional targets in the pathway. *Plant Physiology*, *149*(2), 1028-1041.
- Tewtrakul, S., Subhadhirasakul, S., & Kummee, S. (2003a). HIV-1 protease inhibitory effects of medicinal plants used as self medication by AIDS patients. *Songklanakarin Journal of Science and Technology*, *25*(2), 239-243.
- Tewtrakul, S., Subhadhirasakul, S., Karalai, C., Ponglimanont, C., & Cheenpracha, S. (2009). Anti-inflammatory effects of compounds from *Kaempferia parviflora* and *Boesenbergia pandurata*. *Food Chemistry*, *115*(2), 534-538.
- Tewtrakul, S., Subhadhirasakul, S., Puripattavong, J., & Panphadung, T. (2003b). HIV-1 protease inhibitory substances from the rhizomes of *Boesenbergia pandurata* Holtt. *Songklanakarin Journal of Science and Technology*, *25*(4), 503-508.
- Thompson, W., Meinwald, J., Aneshansley, D., & Eisner, T. (1972). Flavonols: pigments responsible for ultraviolet absorption in nectar guide of flower. *Science*, *177*(4048), 528-530.
- Tian, H., Xu, X., Zhang, F., Wang, Y., Guo, S., Qin, X., & Du, G. (2015). Analysis of *Polygala tenuifolia* transcriptome and description of secondary metabolite biosynthetic pathways by Illumina sequencing. *International Journal of Genomics*, *2015*, 1-11.
- Tian, Q., Wang, X., Li, C., Lu, W., Yang, L., Jiang, Y., & Luo, K. (2013). Functional characterization of the poplar R2R3-MYB transcription factor PtoMYB216 involved in the regulation of lignin biosynthesis during wood formation. *PLOS ONE*, *8*(10), e76369.

- Tominaga, R., Iwata, M., Sano, R., Inoue, K., Okada, K., & Wada, T. (2008). Arabidopsis CAPRICE-LIKE MYB 3 (CPL3) controls endoreduplication and flowering development in addition to trichome and root hair formation. *Development*, 135(7), 1335-1345.
- Tominaga-Wada, R., & Wada, T. (2014). Regulation of root hair cell differentiation by R3 MYB transcription factors in tomato and Arabidopsis. *Frontiers in Plant Science*, 5, 31-35.
- Tominaga-Wada, R., Nukumizu, Y., & Wada, T. (2013b). Flowering is delayed by mutations in homologous genes CAPRICE and TRYPTICHON in the early flowering Arabidopsis cpl3 mutant. *Journal of Plant Physiology*, 170(16), 1466-1468.
- Tominaga-Wada, R., Nukumizu, Y., & Wada, T. (2013c). Tomato (*Solanum lycopersicum*) homologs of TRIPTYCHON (SITRY) and GLABRA3 (SIGL3) are involved in anthocyanin accumulation. *Plant Signaling & Behavior*, 8(7), e24575.
- Tominaga-Wada, R., Nukumizu, Y., Sato, S., & Wada, T. (2013a). Control of plant trichome and root-hair development by a tomato (*Solanum lycopersicum*) R3 MYB transcription factor. *PLOS ONE*, 8(1), e54019.
- Trakoontivakorn, G., Nakahara, K., Shinmoto, H., Takenaka, M., Onishi-Kameyama, M., Ono, H., . . . Tsushida, T. (2001). Structural analysis of a novel antimutagenic compound, 4-hydroxypanduratin A, and the antimutagenic activity of flavonoids in a Thai spice, fingerroot (*Boesenbergia pandurata* Schult.) against mutagenic heterocyclic amines. *Journal of Agricultural and Food Chemistry*, 49(6), 3046-3050.
- Tsurumaru, Y., Sasaki, K., Miyawaki, T., Momma, T., Umemoto, N., & Yazaki, K. (2010). An aromatic prenyltransferase-like gene HIPT-1 preferentially expressed in lupulin glands of hop. *Plant Biotechnology*, 27(2), 199-204.
- Tsurumaru, Y., Sasaki, K., Miyawaki, T., Uto, Y., Momma, T., Umemoto, N., . . . Yazaki, K. (2012). HIPT-1, a membrane-bound prenyltransferase responsible for the biosynthesis of bitter acids in hops. *Biochemical and Biophysical Research Communications*, 417(1), 393-398.
- Tuchinda, P., Reutrakul, V., Claeson, P., Pongprayoon, U., Sematong, T., Santisuk, T., & Taylor, W.C. (2002). Anti-inflammatory cyclohexenyl chalcone derivatives in *Boesenbergia pandurata*. *Phytochemistry*, 59(2), 169-173.

- Tuntiwachwuttikul, P., Pancharoen, O., Reutrakul, V., & Byrne, L. (1984). (1'RS,2'SR,6'RS)-(2,6-Dihydroxy-4-methoxyphenyl)-[3'-methyl-2'-(3"-methylbut-2"-enyl)-6'-phenyl-cyclohex-3'-enyl]methanone (panduratin A)-a constituent of the red rhizomers of a variety of *Boesenbergia pandurata*. *Australian Journal of Chemistry*, 37(2), 449-453.
- Ülker, B., & Somssich, I. E. (2004). WRKY transcription factors: from DNA binding towards biological function. *Current Opinion in Plant Biology*, 7(5), 491-498.
- Urao, T., Noji, M. A., Yamaguchi-Shinozaki, K., & Shinozaki, K. (1996). A transcriptional activation domain of ATMYB2, a drought-inducible Arabidopsis Myb-related protein. *The Plant Journal*, 10(6), 1145-1148.
- Venkatesh, T. V., Karunanandaa, B., Free, D. L., Rottnek, J. M., Baszis, S. R., & Valentin, H. E. (2006). Identification and characterization of an Arabidopsis homogentisate phytyltransferase paralog. *Planta*, 223(6), 1134-1144.
- Villar, E., Klopp, C., Noirot, C., Novaes, E., Kirst, M., Plomion, C., & Gion, J.-M. (2011). RNA-Seq reveals genotype-specific molecular responses to water deficit in eucalyptus. *BMC Genomics*, 12(1), 538.
- Vimolmangkang, S., Han, Y., Wei, G., & Korban, S.S. (2013). An apple MYB transcription factor, MdMYB3, is involved in regulation of anthocyanin biosynthesis and flower development. *BMC Plant Biology*, 13(1), 176.
- Vom Endt, D., Kijne, J. W., & Memelink, J. (2002). Transcription factors controlling plant secondary metabolism: what regulates the regulators? *Phytochemistry*, 61(2), 107-114.
- Voravuthikunchai, S., Limsuwan, S., Supapol, O., & Subhadhirasakul, S. (2006). Antibacterial activity of extracts from family Zingiberaceae against foodborne pathogens. *Journal of Food Safety*, 26(4), 325-334.
- Wang, Q., Ding, Z. -H., Liu, J. -K., & Zheng, Y. -T. (2004). Xanthohumol, a novel anti-HIV-1 agent purified from Hops *Humulus lupulus*. *Antiviral Research*, 64(3), 189-194.
- Wang, R., Chen, R., Li, J., Liu, X., Xie, K., Chen, D., . . . Dai, J. (2014). Molecular characterization and phylogenetic analysis of two novel regio-specific flavonoid prenyltransferases from *Morus alba* and *Cudrania tricuspidata*. *Journal of Biological Chemistry*, 289(52), 35815-35825.

- Wang, S., & Chen, J. -G. (2014). Regulation of cell fate determination by single-repeat R3 MYB transcription factors in Arabidopsis. *Frontiers in Plant Science*, 5, 36-47.
- Wang, S., Hubbard, L., Chang, Y., Guo, J., Schiefelbein, J., & Chen, J. -G. (2008). Comprehensive analysis of single-repeat R3 MYB proteins in epidermal cell patterning and their transcriptional regulation in Arabidopsis. *BMC Plant Biology*, 8(1), 81.
- Wang, X., Li, S., Li, J., Li, C., & Zhang, Y. (2015). *De novo* transcriptome sequencing in *Pueraria lobata* to identify putative genes involved in isoflavones biosynthesis. *Plant Cell Reports*, 34(5), 733-743.
- Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, 10(1), 57-63.
- Wang, Z., Zhang, J., Jia, C., Liu, J., Li, Y., Yin, X., . . . Jin, Z. (2012). *De novo* characterization of the banana root transcriptome and analysis of gene expression under *Fusarium oxysporum* f. sp. *Cubense* tropical race 4 infection. *BMC Genomics*, 13(1), 650.
- Wani, M.C., Taylor, H. L., Wall, M. E., Coggon, P., & McPhail, A. T. (1971). Plant antitumor agents. VI. Isolation and structure of taxol, a novel antileukemic and antitumor agent from *Taxus brevifolia*. *Journal of the American Chemical Society*, 93(9), 2325-2327.
- Wanner, L. A., Li, G., Ware, D., Somssich, I. E., & Davis, K. R. (1995). The phenylalanine ammonia-lyase gene family in *Arabidopsis thaliana*. *Plant Molecular Biology*, 27(2), 327-338.
- Wätjen, W., Weber, N., Lou, Y. -J., Wang, Z. -Q., Chovolou, Y., Kampkötter, A., Kahl, R., & Proksch, P. (2007). Prenylation enhances cytotoxicity of apigenin and liquiritigenin in rat H4IIE hepatoma and C6 glioma cells. *Food and Chemical Toxicology*, 45(1), 119-124.
- Weaver, L. M., & Herrmann, K. M. (1997). Dynamics of the shikimate pathway in plants. *Trends in Plant Science*, 2(9), 346-351.
- Wei, H., Chen, X., Zong, X., Shu, H., Gao, D., & Liu, Q. (2015). Comparative transcriptome analysis of genes involved in anthocyanin biosynthesis in the red and yellow fruits of sweet cherry (*Prunus avium* L.). *PLOS ONE*, 10(3), e0121164.

- Wenping, H., Yuan, Z., Jie, S., Lijun, Z., & Zhezhi, W. (2011). *De novo* transcriptome sequencing in *Salvia miltiorrhiza* to identify genes involved in the biosynthesis of active ingredients. *Genomics*, 98(4), 272-279.
- Weston, K. (1998). Myb proteins in life, death and differentiation. *Current Opinion in Genetics & Development*, 8(1), 76-81.
- Whetten, R. W., MacKay, J. J., & Sederoff, R. R. (1998). Recent advances in understanding lignin biosynthesis. *Annual Review of Plant Biology*, 49(1), 585-609.
- Whetten, R., & Sederoff, R. (1995). Lignin biosynthesis. *The Plant Cell*, 7(7), 1001-1013.
- Wijayakusuma, H. M. H. (2001). *Tumbuhan berkhasiat obat Indonesia: rempah, rimpang dan umbi*. Jakarta: Milenia Populer.
- Wilkins, O., Nahal, H., Foong, J., Provart, N. J., & Campbell, M. M. (2009). Expansion and diversification of the *Populus* R2R3-MYB family of transcription factors. *Plant Physiology*, 149(2), 981-993.
- Win, N. N., Awale, S., Esumi, H., Tezuka, Y., & Kadota, S. (2007). Bioactive secondary metabolites from *Boesenbergia pandurata* of Myanmar and their preferential cytotoxicity against human pancreatic cancer PANC-1 cell line in nutrient-deprived medium. *The Journal of Natural Products*, 70(10), 1582-1587.
- Win, N. N., Awale, S., Esumi, H., Tezuka, Y., & Kadota, S. (2008). Panduratin D - I, novel secondary metabolites from rhizomes of *Boesenbergia pandurata*. *Chemical and Pharmaceutical Bulletin*, 56(4), 491-496.
- Wingender, R., Röhrig, H., Hörnicke, C., Wing, D., & Schell, J. (1989). Differential regulation of soybean chalcone synthase genes in plant defence, symbiosis and upon environmental stimuli. *Molecular General Genetics*, 218(2), 315-322.
- Winkel-Shirley, B. (2001). Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiology*, 126(2), 485-493.
- World Health Organization. (2012, August). *Global strategy for dengue prevention and control, 2012-2020*. Retrieved May 16, 2013, from <http://www.who.int/denguecontrol/9789241504034/en/>

- Xu, H. -X., & Lee, S. F. (2001). Activity of plant flavonoids against antibiotic-resistant bacteria. *Phytotherapy Research*, 15(1), 39-43.
- Xu, H., Zhang, F., Liu, B., Huhman, D. V., Sumner, L. W., Dixon, R. A., & Wang, G. (2013). Characterization of the formation of branched short-chain fatty acid: CoAs for bitter acid biosynthesis in hop glandular trichomes. *Molecular Plant*, 6(4), 1301-1317.
- Xu, Y., Gao, S., Yang, Y., Huang, M., Cheng, L., Wei, Q., . . . Hong, B. (2013). Transcriptome sequencing and whole genome expression profiling of chrysanthemum under dehydration stress. *BMC Genomics*, 14(1), 662.
- Yamamoto, H., Senda, M., & Inoue, K. (2000). Flavanone 8-dimethylallyltransferase in *Sophora flavescens* cell suspension cultures. *Phytochemistry*, 54(7), 649-655.
- Yamazaki, Y., Suh, D. Y., Sitthithaworn, W., Ishiguro, K., Kobayashi, Y., Shibuya, M., . . . Sankawa, U. (2001). Diverse chalcone synthase superfamily enzymes from the most primitive vascular plant, *Psilotum nudum*. *Planta*, 214(1), 75-84.
- Yang, C. Q., Fang, X., Wu, X. M., Mao, Y. B., Wang, L. J., & Chen, X. Y. (2012). Transcriptional regulation of plant secondary metabolism. *Journal of Integrative Plant Biology*, 54(10), 703-712.
- Yang, X., Jiang, Y., Yang, J., He, J., Sun, J., Chen, F., . . . Yang, B. (2015). Prenylated flavonoids, promising nutraceuticals with impressive biological activities. *Trends in Food Science & Technology*, 44(1), 93-104.
- Yanti & Hwang, J. -K. (2010). Suppressive effect of ethanolic *Kaempferia pandurata* Roxb. extract on matrix metalloproteinase-2 expression in *Porphyromonas gingivalis*-treated human gingival fibroblasts *in vitro*. *Journal of Oral Science*, 52(4), 583-591.
- Yanti, Anggakusuma, Gwon, S. -H., & Hwang, J. -K (2009d). *Kaempferia pandurata* Roxb. inhibits *Porphyromonas gingivalis* supernatant-induced matrix metalloproteinase-9 expression via signal transduction in human oral epidermoid cells. *Journal of Ethnopharmacology*, 123(2), 315-324.
- Yanti, Lee, M., Kim, D., & Hwang, J. -K. (2009a). Inhibitory effect of panduratin A on c-Jun N-terminal kinase and activator protein-1 signaling involved in *Porphyromonas gingivalis* supernatant-stimulated matrix metalloproteinase-9 expression in human oral epidermoid cells. *Biological and Pharmaceutical Bulletin*, 32(10), 1770-1775.

- Yanti, Oh, H. I., Anggakusuma & Hwang, J. K. (2009b). Effects of panduratin A isolated from *Kaempferia pandurata* Roxb. on the expression of matrix metalloproteinase-9 by *Porphyromonas gingivalis* supernatant-induced KB cells. *Biological and Pharmaceutical Bulletin*, 32(1), 110-115.
- Yanti, Rukayadi, Y., Lee, K. H., & Hwang, J. K. (2009c). Activity of panduratin A isolated from *Kaempferia pandurata* Roxb. against multi-species oral biofilms *in vitro*. *Journal of Oral Science*, 51(1), 87-95.
- Yazaki, K., Kuniyoshi, M., Fujisaki, T., & Sato, F. (2002). Geranyl diphosphate: 4-hydroxybenzoate geranyltransferase from *Lithospermum erythrorhizon*. *Journal of Biological Chemistry*, 277(8), 6240-6246.
- Yazaki, K., Sasaki, K., & Tsurumaru, Y. (2009). Prenylation of aromatic compounds, a key diversification of plant secondary metabolites. *Phytochemistry*, 70(15-16), 1739-1745.
- Ye, J., Fang, L., Zheng, H., Zhang, Y., Chen, J., Zhang, Z., . . . Wang, J. (2006). WEGO: a web tool for plotting GO annotations. *Nucleic Acids Research*, 34(suppl_2), W293-W297.
- Ye, J., Hu, T., Yang, C., Li, H., Yang, M., Ijaz, R., . . . Zhang, Y. (2015). Transcriptome profiling of tomato fruit development reveals transcription factors associated with ascorbic acid, carotenoid and flavonoid biosynthesis. *PLOS ONE*, 10(7), e0130885.
- Yoon, J. -H., Shim, J. -S., Cho, Y., Baek, N. -I., Lee, C. -W., Kim, H. -S., & Hwang, J. -K. (2007). Depigmentation of melanocytes by isopanduratin A and 4-hydroxypanduratin A isolated from *Kaempferia pandurata* RoXB. *Biological and Pharmaceutical Bulletin*, 30(11), 2141-2145.
- Yoshida, K., Iwasaka, R., Kaneko, T., Sato, S., Tabata, S., & Sakuta, M. (2008). Functional differentiation of *Lotus japonicus* TT2s, R2R3-MYB transcription factors comprising a multigene family. *Plant and Cell Physiology*, 49(2), 157-169.
- Yuan, Y., Chiu, L. -W., & Li, L. (2009). Transcriptional regulation of anthocyanin biosynthesis in red cabbage. *Planta*, 230(6), 1141-1153.
- Yun, J. M., Kweon, M. H., Kwon, H., Hwang, J. K., & Mukhtar, H. (2006). Induction of apoptosis and cell cycle arrest by a chalcone panduratin A isolated from *Kaempferia pandurata* in androgen-independent human prostate cancer cells PC3 and DU145. *Carcinogenesis*, 27(7), 1454-1464.

- Yun, J. M., Kwon, H., & Hwang, J. K. (2003). *In vitro* anti-inflammatory activity of panduratin A isolated from *Kaempferia pandurata* in RAW264. 7 cells. *Planta Medica*, 69(12), 1102-1108.
- Yun, J. M., Kwon, H., Mukhtar, H., & Hwang, J. K. (2005). Induction of apoptosis by panduratin A isolated from *Kaempferia pandurata* in human colon cancer HT-29 cells. *Planta Medica*, 71(6), 501-507.
- Yusuf, N. A., Annuar, M. S. M., & Khalid, N. (2011). Efficient propagation of an important medicinal plant *Boesenbergia rotunda* by shoot derived callus. *Journal of Medicinal Plants Research*, 5(13), 2629-2636.
- Yusuf, N. A., Annuar, M. S., & Khalid, N. (2011). Rapid micropropagation of *Boesenbergia rotunda* (L.) Mansf. Kulturpfl.(a valuable medicinal plant) from shoot bud explants. *African Journal of Biotechnology*, 10(7), 1194-1199.
- Zhang, F., Gao, Q., Khan, G., Luo, K., & Chen, S. (2014). Comparative transcriptome analysis of aboveground and underground tissues of *Rhodiola algida*, an important ethno-medicinal herb endemic to the Qinghai-Tibetan Plateau. *Gene*, 553(2), 90-97.
- Zhang, F., Gonzalez, A., Zhao, M., Payne, C. T., & Lloyd, A. (2003). A network of redundant bHLH proteins functions in all TTG1-dependent pathways of Arabidopsis. *Development*, 130(20), 4859-4869.
- Zhang, S., Ma, P., Yang, D., Li, W., Liang, Z., Liu, Y., & Liu, F. (2013). Cloning and characterization of a putative R2R3 MYB transcriptional repressor of the rosmarinic acid biosynthetic pathway from *Salvia miltiorrhiza*. *PLOS ONE*, 8(9), e73259.
- Zhang, X. -N., Han, Z. -H., Yin, L. -L., Kong, J., Xu, X. -F., Zhang, X. -Z., & Wang, Y. (2013). Heterologous functional analysis of the *Malus xiaojinensis* MxIRT1 gene and the His-box motif by expression in yeast. *Molecular Biology Reports*, 40(2), 1499-1504.
- Zhang, Z., Liu, X., Wang, X., Zhou, M., Zhou, X., Ye, X., & Wei, X. (2012). An R2R3 MYB transcription factor in wheat, TaPIMP1, mediates host resistance to *Bipolaris sorokiniana* and drought stresses through regulation of defense-and stress-related genes. *New Phytologist*, 196(4), 1155-1170.
- Zhou, D., Gao, S., Wang, H., Lei, T., Shen, J., Gao, J., . . . Liu, J. (2016). *De novo* sequencing transcriptome of endemic *Gentiana straminea* (Gentianaceae) to

identify genes involved in the biosynthesis of active ingredients. *Gene*, 575(1), 160-170.

Zhou, J., Lee, C., Zhong, R., & Ye, Z. -H. (2009). MYB58 and MYB63 are transcriptional activators of the lignin biosynthetic pathway during secondary cell wall formation in *Arabidopsis*. *The Plant Cell*, 21(1), 248-266.

Zhu, H. -F., Fitzsimmons, K., Khandelwal, A., & Kranz, R. G. (2009). CPC, a single-repeat R3 MYB, is a negative regulator of anthocyanin biosynthesis in *Arabidopsis*. *Molecular Plant*, 2(4), 790-802.

Zhu, L., Shan, H., Chen, S., Jiang, J., Gu, C., Zhou, G., . . . Chen, F. (2013). The heterologous expression of the chrysanthemum R2R3-MYB transcription factor CmMYB1 alters lignin composition and represses flavonoid synthesis in *Arabidopsis thaliana*. *PLOS ONE*, 8(6), e65680.

Zimmermann, I. M., Heim, M. A., Weisshaar, B., & Uhrig, J. F. (2004). Comprehensive identification of *Arabidopsis thaliana* MYB transcription factors interacting with R/B-like BHLH proteins. *The Plant Journal*, 40(1), 22-34.

Zou, B., Jia, Z., Tian, S., Wang, X., Gou, Z., Lü, B., & Dong, H. (2013). AtMYB44 positively modulates disease resistance to *Pseudomonas syringae* through the salicylic acid signalling pathway in *Arabidopsis*. *Functional Plant Biology*, 40(3), 304-313.

Zubieta, C., He, X. -Z., Dixon, R. A., & Noel, J. P. (2001). Structures of two natural product methyltransferases reveal the basis for substrate specificity in plant O-methyltransferases. *Nature Structural & Molecular Biology*, 8(3), 271-279.

Zuluaga, D. L., Gonzali, S., Loreti, E., Pucciariello, C., Degl'Innocenti, E., Guidi, L., . . . Perata, P. (2008). *Arabidopsis thaliana* MYB75/PAP1 transcription factor induces anthocyanin production in transgenic tomato plants. *Functional Plant Biology*, 35(7), 606-618.

Zvi, M. M. B., Shklarman, E., Masci, T., Kalev, H., Debener, T., Shafir, S., . . . Vainstein, A. (2012). PAP1 transcription factor enhances production of phenylpropanoid and terpenoid scent compounds in rose flowers. *New Phytologist*, 195(2), 335-345.

LIST OF PUBLICATIONS AND PAPERS PRESENTED

1. Md-Mustafa, N. D., Mohd-Yusuf, Y., Khalid, N., & Othman, R. Y. (2013, June). Transcriptome profiling data of *Boesenbergia rotunda* cell culture in response to exogenous phenylalanine treatment. Poster session presented at the 20th Scientific Meeting of the Malaysian Society for Molecular Biology and Biotechnology (MSMB), Research Management and Innovation Complex (RMIC), University of Malaya.
2. Md-Mustafa, N. D., Khalid, N., Gao, H., Peng, Z., Alimin, M. F., Bujang, N., . . . Othman, R. Y. (2014). Transcriptome profiling shows gene regulation patterns in a flavonoid pathway in response to exogenous phenylalanine in *Boesenbergia rotunda* cell culture. *BMC Genomics*, 15(1), 984-1009.