ISOLATION, CHARACTERIZATION AND OVEREXPRESSION OF *Chalcone synthase* GENE IN SUSPENSION CULTURES OF *Boesenbergia rotunda*

NURNADIAH BINTI ROSLAN

FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2017

ISOLATION, CHARACTERIZATION AND OVEREXPRESSION OF *Chalcone synthase* GENE IN SUSPENSION CULTURES OF *Boesenbergia rotunda*

NURNADIAH BINTI ROSLAN

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: Nurnadiah binti Roslan

Matric No: SHC120118

Name of Degree: Doctor of Philosophy (except mathematics & science)

Title of Thesis ("this Work"): ISOLATION, CHARACTERIZATION AND OVEREXPRESSION OF *Chalcone synthase* GENE IN SUSPENSION CULTURES OF *Boesenbergia rotunda*

Field of Study: Plant Molecular Biology

I do solemnly and sincerely declare that:

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

Candidate's Signature Date:

Subscribed and solemnly declared before,

Witness's Signature Date:

Name:

Designation:

ISOLATION, CHARACTERIZATION AND OVEREXPRESSION OF *Chalcone synthase* GENE IN SUSPENSION CULTURES OF *Boesenbergia rotunda*

ABSTRACT

Boesenbergia rotunda (L.) Mansf. is a herb plant from Boesenbergia genera and belongs to Zingiberaceae family. In some Asian countries, it is commonly consumed as a traditional meal or used as a spice ingredient and traditional medicine. Its ethnomedicinal properties derived from rhizome extracts have broadly drawn attention to further explore its medicinal and health values. The bioactive compounds from rhizome extracts are plant secondary metabolites synthesized from the flavonoid biosynthesis pathway by a number of enzymes. In most of the plants, chalcone synthase (CHS) is the first key enzyme that initiates the biosynthesis of flavonoid. In this study, at least four to five copies of *BrCHS* were successfully detected and predominantly expressed in a tissue-specific manner. *BrCHS2* isolated from the rhizomes was chosen to be introduced into B. rotunda suspension culture to observe the potential effects of different *BrCHS* levels on flavonoid accumulation and regulation of gene expression in the plant pathways. The transcript variant from rhizome was chosen because it is the most relevant variant with medicinal interest as it has been utilized to be expressed and produced compound that serve pharmaceutical benefits. The enzymatic properties of the BrCHS2 showed that it has utilized Cinnamoyl-CoA as substrate to produce compound for pinocembrin chalcone. The reaction was accompanied by the formation of unidentified products. Focus on elucidating potential genes related to flavonoid biosynthesis was further carried out using high-throughput RNA sequencing (RNA-Seq). The mass of transcript sequences obtained from RNA-Seq led to the annotation of functional genes affected by the influence of increasing the expression *BrCHS*. In order to determine and annotate their functions, a total of 67,251 (65.8%) unigenes were successfully cross referenced to several public databases. An inter-sectioned between three groups of wild-type (WT) and

transgenic (TL) produced 3,148 differentially expressed genes, of which 1,905 were up-regulated and 1,043 were down-regulated in the suspension culture library. Looking at the distribution of differentially expressed genes, the overall picture showed that there were relatively higher numbers of upregulated genes in the transgenic suspension cultures compared to wild-type as the result of *BrCHS* overexpression. However, further comparative analysis against the KEGG database highlighted the fact that for flavonoid biosynthesis and related pathways, there were actually more down-regulated genes. Due to the overexpression of *BrCHS*, unigenes that were responsible for plant defense pathway have also increased as a response towards stress, and thus obtained important function on plant resistance. With sufficient data and transcripts obtained, the potential unigenes that could be applied to synthesize targeted compounds, especially in the flavonoid biosynthesis pathway responsible for the potent medicinal properties production, could be identified for ultimate use in future drug analysis.

Keywords: *Chalcone synthase*, *Boesenbergia rotunda*, overexpression, flavonoid, RNA-Seq

PENGASINGAN, PENCIRIAN DAN EKSPRESI BERLEBIHAN OLEH GEN *Chalcone synthase* DI DALAM KULTUR TERAMPAI *Boesenbergia rotunda*

ABSTRAK

Boesenbergia rotunda (L.) Mansf. merupakan tumbuhan herba dari genus Boesenbergia dan tergolong dalam keluarga Zingiberaceae. Ia biasanya digunakan sebagai hidangan tradisi atau sebagai bahan rempah dan ubat-ubatan tradisional yang penggunaannya adalah meluas di kalangan penduduk di beberapa negara Asia. Nilai perubatan-etno yang diperoleh daripada ekstrak rizom telah menarik perhatian ke arah penerokaan nilai-nilai perubatan dan kesihatan berkaitan herba tersebut. Sebatian bioaktif daripada ekstrak rizom adalah metabolit sekunder tumbuhan tersintesis daripada laluan biosintesis *flavonoid* oleh beberapa enzim. Dalam kebanyakan tumbuh-tumbuhan, chalcone synthase (CHS) adalah enzim utama pertama yang memulakan biosintesis *flavonoid*. Dalam kajian ini, sekurang-kurangnya empat hingga lima salinan gen *BrCHS* telah berjaya dikesan dan kebanyakannya diekspresi dalam tisu yang tertentu. Gen *BrCHS2* yang diasingkan daripada rizom dipilih untuk diperkenalkan ke dalam kultur terampai *B. rotunda* untuk melihat sebarang kesan yang berkemungkinan berlaku disebabkan oleh pelbagai tahap *BrCHS* yang berbeza pada pengumpulan *flavonoid* dan pengawalan gen di dalam pelbagai laluan tumbuhan. Ciri-ciri enzimatik BrCHS2 menunjukkan bahawa ia telah menggunakan *Cinnamoyl-CoA* sebagai substrat untuk menghasilkan sebatian bagi *pinocembrin chalcone* dan tindak balasnya telah diiringi oleh pembentukan produk yang tidak dikenali. Fokus selanjutnya dalam menerangkan gen yang berpotensi berkaitan dengan biosintesis *flavonoid* telah dijalankan menggunakan urutan RNA celusan-tinggi (RNA-Seq). Kumpulan jujukan transkrip yang diperolehi daripada RNA-Seq membawa kepada anotasi gen berfungsi yang terkesan akibat penambahan gen *BrCHS*. Sebanyak 67.251 (65.8%) gen tunggal telah berjaya dirujuk silang pada beberapa pangkalan data awam untuk menentukan dan memberikan

keterangan tentang fungsi mereka. Lintasan antara tiga kumpulan jenis liar (WT-*wild-type*) dan transgenik (TL-*transgenic*) menghasilkan 3,148 gen terzahir berbeza, yang mana 1,905 diatur-menaik dan 1,043 diatur-menurun dalam daftar kultur ampaian. Melihat kepada pengagihan gen terzahir berbeza, gambaran keseluruhannya menunjukkan bahawa terdapat jumlah gen diatur-menaik yang agak tinggi di dalam kultur ampaian transgenik berbanding jenis liar (WT) disebabkan oleh ekspresi berlebihan *BrCHS*. Walau bagaimanapun, analisis perbandingan selanjutnya di dalam pangkalan data KEGG telah menemukan fakta bahawa bagi biosintesis *flavonoid* dan laluan yang berkaitan, sebenarnya gen yang diatur-menurun adalah lebih banyak. Disebabkan oleh penekanan-lampau gen *BrCHS*, gen-tunggal yang bertanggungjawab dalam menyumbangkan daya tahan tumbuhan, juga telah meningkat sebagai tindak balas terhadap tekanan, dan dengan demikian memperoleh fungsi yang penting dalam membantu daya tahan tumbuhan. Dengan memperoleh jumlah data dan transkrip yang mencukupi, sekurang-kurangnya kita akan dapat mengenal pasti gen-tunggal yang berpotensi yang boleh digunakan untuk mensintesis sebatian yang disasarkan terutamanya di dalam laluan biosintesis *flavonoid* yang bertanggungjawab untuk penghasilan sifat-sifat perubatan yang kuat, boleh dikenal pasti untuk kegunaan dalam analisa ubat pada masa depan.

Kata kunci: *Chalcone synthase*, *Boesenbergia rotunda*, ekspresi berlebihan, *flavonoid*, RNA-Seq

ACKNOWLEDGEMENTS

First and foremost, I am deeply grateful to Allah, the Great Almighty, for giving me the opportunity and ability to face this colourful journey and complete my graduate program. My sincere gratitude to my advisor, Prof. Dr. Zulqarnain Mohamed, who introduced this study to me and provided continuous support of my Ph.D study and related research. His guidance helped me in all the time of research and writing of this thesis.

Besides my advisor, I would like to thanks Dr. Teh Ser Huy, who patiently helped me to perform several laboratory tasks starting from the beginning I joined the research group. Her insightful comments and encouragement, incented me to widen my research from various perspectives.

My sincere thanks also goes to Prof. Dr. Noorzulaani Khalid, Prof. Dr. Norsaadah, Dr. Lee Yean Kee, and Dr. Wong Sher Ming, who gave access to the laboratory and research facilities. Without their precious support it would not be possible to conduct this research.

The completion of this thesis could not have been possible without the participation, assistance and endless supports of my beloved parents; Roslan bin Jani and Maimunah bt Ali and my bestfriends; Maisarah bt Mustapa, Norliyana bt Nor Hisham Shah and Ezzatul Alia bt Zubir Ansori.

Lastly, I thank my fellow labmates for the stimulating discussions, troubleshooting workbench problems, for the sleepless nights we were working together before deadlines and for all the fun we have had in the last few years. Thank you very much friends, I will never forget all of your kindness.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

2,4-D	:	2,4-dichlorophenoxyacetic acid
MOPS	:	3-(N-morpholino) propanesulfonic acid
X-gluc	:	5-bromo-4-chloro-3-indolyl-beta-D-glucuronic acid
X-gal	:	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
BCIP	:	5-bromo-4-chloro-3'-indolyphosphate p-toluidine salt
BAP	:	6-Benzylaminopurine
AS	:	Acetosyringone
bp	:	Basepair
B. rotunda	:	*Boesenbergia rotunda*
BSA	:	Bovine serum albumin
CO_2	:	Carbon dioxide
°C	:	Celsius
cm	:	Centimeter
CHS	:	*Chalcone synthase*
CI	:	Chloroform: Isoamyl-alcohol
CDS	:	Coding DNA sequence
cDNA	:	Complementary DNA
DNA	:	Deoxyribonucleic acid
dNTP	:	Deoxyribonucleotide triphosphate
Na-EDTA	:	Disodium Ethylenediaminetetraacetic acid
DTT	:	Dithiothreitol
E. coli	:	*Escherichia coli*
EtBr	:	Ethidium Bromide
x g	:	g-force

GPS	:	Global Positioning System
CTAB	:	Hexa-decyltrimethylammonium bromide
HCl	:	Hydrochloric acid
IC_{50}	:	Inhibitory concentration
IPTG	:	Isopropyl β-D-1-thiogalactopyranoside
kb	:	Kilobases
kDa	:	Kilodalton
L	:	Liter
LiCl	:	Lithium chloride
LB	:	Luria Bertani
$MgSO_4$	:	Magnesium chloride
CH_3OH	:	Methanol
µ	:	Micro
µg	:	Microgram
m	:	Milli
mg	:	Milligram
MIC	:	Minimal inhibitory concentration
min	:	Minute
M	:	Molar
MS	:	Murashige and Skoog
nm	:	Nanometer
nM	:	Nanomolar
NBT	:	Nitro-blue tetrazolium chloride
OD	:	Optical density
PC	:	Phenol:Chloroform
PCI	:	Phenol:Chloroform:Isoamyl-alcohol

H_3PO_4	:	Phosphoric acid
KOAc	:	Potassium acetate
pH	:	Potential of hydrogen
psi	:	Pounds per square inch
rpm	:	Revolutions per minute
RNA	:	Ribonucleic acid
SSC	:	Saline sodium citrate
sec	:	Second
SCV	:	Settled cell volume
NaOAc	:	Sodium acetate
NaCl	:	Sodium chloride
SDS	:	Sodium dodecyl sulphate
NaOH	:	Sodium hydroxide
sdH_2O	:	Sterile distilled water
™	:	Trademark
UV	:	Ultraviolet
V	:	Voltage
v/v	:	Volume per volume
w/v	:	Weight per volume

LIST OF APPENDICES

CHAPTER 1: INTRODUCTION

1.1 Current episode on dengue in Malaysia

Dengue is a very serious problem in Malaysia. Every year, the number of cases is increasing since the first reported case in 1902 (Skae, 1902). Dengue has emerged as a public health burden in Southeast Asia and reported an epidemic of dengue fever that caused morbidity and mortality in Malaysia. With just one bite of *Aedes aegypti*; a type of mosquito that carries infectious virus which can spread the dengue fever to human. There are three types of dengue fever in order of least severe to most: the typical uncomplicated dengue fever, dengue hemorrhagic fever (DHS) and dengue shock syndrome (DSS). Dengue can spread easily throughout the country, especially in urban areas due to the close proximity of people to one another with the *Aedes aegypti* mosquito for most urban transmission of dengue in Malaysia. The World Health Organization (WHO) estimated that 40% of the world population is at risk of this infectious mosquito-borne disease.

The latest report from the Malaysia's Ministry of Health showed the total number of dengue cases was 37,654 with 93 fatalities from 1st January until 29th May 2017 (Latest dengue information, 2017). This statistics shows that the increasing numbers of dengue cases in Malaysia are at an alarming rate. At this point of time, the best method of controlling this disease is by prevention, as no other better method to overcome the disease directly. The Malaysian authorities have tried to fight the dengue for decades by relying on outmoded mosquito control techniques such as insecticide and fogging to control the population of mosquitoes. However, these methods do not bring any significant impact in slowing or reducing dengue.

Currently, scientists are working on genetic engineering to save lives. Genetically modified (GM) mosquitoes have been created to fight against infectious mosquitoes. Institute for Medical Research (IMR) under Ministry of Health, Malaysia is evaluating 'GM mosquitoes' as a new solution to combat the vectors. These GM mosquitoes are genetically modified with self-limiting gene, which before the larvae reach maturity, it will be killed by the gene itself. Despite that, only female mosquitos are dangerous as the male will not pose any threat to human as it is not able to bite human or transmit any diseases. However, the possibilities that the mosquitoes become more resistant towards insecticides and spread another outbreak of human disease in the future are still under investigation.

Another approach to control dengue is the introduction of a vaccine for dengue called Dengavaxia® in 2015. This is a part of the dengue prevention program. However, it is not a cost-effective method as the vaccine is expensive and the target group of people for vaccination is huge. The vaccination program might be stopped due to budget limitation or only particular age-group or sub group will be selected for the vaccination program. Thus, with no definite and promising treatment from the creation of GM mosquitoes and the availability of modern medicine such as vaccines, many researches have turned to an alternative plant-based medicine for the treatment of dengue.

1.2 *Boesenbergia rotunda*, healing herb that lies in nature

The search for traditional herbs to treat dengue infection has increased. Since the existing antiviral drugs that often lead to viral resistance and high contraindication of their usage, alternative treatments using herbal plants become high in demand as they are generally considered not to have adverse effects compared with synthetic drugs. Scientists have reported that there are various plant species which have been identified to possess potential bioactive compounds. These natural bioactive compounds are derived

from extracts of different parts of medicinal plants and this further fits into the trajectory of current research on *Boesenbergia rotunda* (*B. rotunda*), a fingerroot herb plant.

According to previous studies, both compound of cyclohexenyl chalcone derivatives (CCDs) such as panduratin A and hydroxypanduratin A as examples of compounds isolated from rhizomes of *B. rotunda*, was suggested to have anti-dengue properties. They are mainly attributed to their strong inhibitory action against the NS3 protease of DEN-2 virus as compared to pinocembrin, pinostrobin, cardamonin, and alpinetin compounds (Tan et al. 2006). Therefore, elucidation of flavonoid biosynthesis at the molecular level could lead to a better understanding of the process and to aid in enhancing the production of these valuable secondary metabolites. Despite the potential of these compounds, the main challenge faced by researchers is the limited amount of natural bioactive compounds that is specifically presented in the edible rhizome of *B. rotunda* plants. Additionally, there is limited continuous supply of plant source since *B. rotunda* is propagated by vegetative method. Due to its slow and time consuming propagation, it continues to be a significant challenge for the application of rhizome segment for propagation (Tan et al. 2016). Attempt has also been made with chemical solvents to synthesize some flavonoid compounds; however, the production is still limited due to the use of toxic chemical solvents and extreme reaction conditions (Chee et al. 2010).

Therefore, as an alternative approach, genetic manipulation of its biosynthetic pathways that can help to improve the yield of targeted compounds by utilizing suspension culture as a plant host can be considered. However, this method requires in-depth knowledge of biosynthetic pathway which at present remains limited and unclear. Since most of the important compounds are extracted from rhizomes; therefore, to overcome the low yield, gene from widely branched pathway, *Chalcone synthase* (*CHS)* gene is isolated from rhizomes and introduced in suspension cultures of *B. rotunda*. The

overexpression of *CHS* can determine the enzyme functionality and increase the synthesis of desired compounds as well as discover gene regulatory networks, especially in the flavonoid biosynthesis pathway through mRNA transcriptome profiling.

The findings of this study will rebound to the benefit of society considering that *B. rotunda* is one of the underrated but essential herbs as it is useful in preventing and treating several diseases. A promising solution towards dengue disease can be achieved from the investigation of anti-dengue properties exhibited by panduratin A and hydroxypandutin A. In addition to that, it is also shown how this plant can offer the researchers to further exploit the richness of biochemistry by metabolic engineering and applying suspension culture as a plant host.

In this study, the transcript variant from rhizome was chosen because it is the most relevant variant with medicinal interest as it has been utilized to be expressed and produced compound that serve pharmaceutical benefits. The idea of utilizing and overexpressing the *BrCHS2* (isolated from rhizomes) could give a clue to exploit the synthesis of targeted compounds, to understand how the plant biosynthesis them and how the plant genes control the desired compound in a particular plant pathway.

The general objective of this study is to investigate the effect of overexpression of *BrCHS2* (isolated from rhizome) on the flavonoid biosynthesis and the related plant pathways in suspension cultures of *B. rotunda.* To achieve this main objective, this work has been divided into three parts with their specific objectives.

Part 1: Cloning and characterization of *BrCHS* isolated from different parts of *B. rotunda* plant.

1) To isolate and characterize *BrCHS* from different parts of *B. rotunda.*

2) To determine copy number of *BrCHS* in the plant genome.

Part 2: Overexpression of *BrCHS*2 in suspension cultures of *B. rotunda*.

1) To construct recombinant DNA containing *BrCHS*2 (isolated from rhizomes of *B. rotunda*) fused with 35S promoter and subsequent transformation into suspension cultures.

2) To determine *BrCHS* expression level in transgenic suspension cultures.

3) To investigate *BrCHS2* enzyme functionality in transgenic suspension cultures.

Part 3: Transcriptome profiling of suspension cultures of *B. rotunda*, overexpressed with *BrCHS*.

1) To determine the effect of overexpression of *BrCHS* in flavonoid biosynthesis and related plant pathways at the transcriptomic level.

1.3 Hypothesis

Overexpressing of *BrCHS2* transcript variant in the suspension cultures of *B. rotunda* could result in the enhancement of the flavonoid production and change the transcript expression level in flavonoid biosynthesis pathway and related plant pathways.

CHAPTER 2: LITERATURE REVIEW

2.1 *Boesenbergia rotunda* (L.) Mansf.

2.1.1 Background

Boesenbergia rotunda is an herb plant from the Boesenbergia genera and belongs to the Zingiberaceae family. Its scientific name has been accepted and recorded in The Plant List database by the World Checklist of Selected Plant Families ("The Plant List," 2016). It is widely cultivated throughout Southeast Asia, where it grows in dense forest and shaded lowland areas. It is a type of ginger species which is botanically known as *Boesenbergia cochinchinensis* (Gagnep.) Loes., *Boesenbergia pandurata* (Roxb.) Schltr., *Curcuma rotunda* L., *Gastrochilus panduratus* (Roxb.) Ridl., *Gastrochilus rotundus* (L.) Alston, *Kaempferia cochinchinensis* (Gagnep.), *Kaempferia ovate* (Roscoe), and *Kaempferia pandurata* (Roxb.); however, it is usually known as *Boesenbergia rotunda* (L.) Mansf. (The Plant List, 2016). This plant is more commonly recognized as Chinese keys or fingerroot in English, "Temu Kunci" in Malay and "Krachai" or "Krachai-Dang" in Thailand.

2.1.2 Morphological description

B. rotunda is a small herb plant of about 15 to 40 cm in height. It has three to four broad leaves which are ovate-oblong in shape and are about 7 to 9 cm in width and 10 to 20 cm in length (Figure 2.1a) (Yusuf, 2011). Its rhizome can be found in the underground portion of the plant. It consists of a small globular shaped central rhizome or sometime known as mother rhizome (1.5 to 2 cm in diameter) that sprouts out bright yellow, strongly aromatic slender and long tubers (1.0 to 1.5 cm thick in diameter and 5 to 10 cm long) in the same direction that resemble the fingers of a hand (Figure 2.1b) (Yusuf, 2011). The tissue of the tuber is softer and more watery than the central rhizome. Yellow variety of fingerroot plant produces bright yellow color of central rhizome and the tubers, while the other variety produces red and black rhizomes. Even though they are different

from each other, they produce strong aromatic odor of fingerroot plant. The flowers are tubular, orchid-like flowers and pink in color. These scarlet flowers are unnoticeable because they are hidden at the base of the foliage. It is extremely delicate and short-lived, usually bloom once at a time in tropical countries (Figure 2.1c) (Yusuf, 2011).

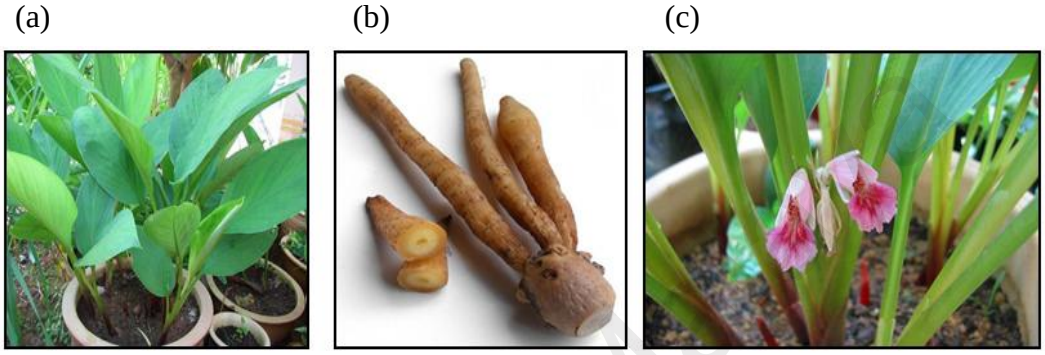


Figure 2.1: Different parts of *B. rotunda* plant. Picture Courtesy: www.davesgarden.com

Whole plants of *B. rotunda* cover a) light green foliage, b) yellow variety of rhizomes and c) darker and light shed of pink flowers.

2.1.3 Common uses of *B. rotunda*

Most people in many Asian countries use this fingerroot plant as a condiment in food due to its aromatic flavor. It is a common edible ingredient in which its young rhizomes and shoots are cooked as vegetable or eaten raw. Traditionally, fingerroot plant is also used for medicinal purposes. The fresh rhizomes are practically used in post-partum tonic mixtures such as the popular Indonesian tonic, 'jamu' for women after childbirth, as a stomachic to improve digestion, to treat illness such as muscle pain, as a remedy for coughs, wounds and mouth ulcers, swelling, tooth and gum disease (Tewtrakul et al. 2009). Crushed rhizomes are also applied externally to treat rheumatism. The popularity of its ethnomedicinal usage that is particularly derived from rhizome extracts has drawn

wide attention for researchers to further explore its pharmaceutical values that could therefore explain the significance of its traditional usage.

2.1.4 Pharmaceutical values of *B. rotunda*

Over the years, researchers have successfully isolated and elucidated nearly a hundred of bioactive compounds extracted from different parts of *B. rotunda* such as leaves, roots, stems and rhizomes. An array of bioactive compounds of *B. rotunda* has shown to execute medicinal benefits especially for treatment of several diseases. The benefit ranges from flavonoid derivatives, chalcone derivatives, esters, essential oils, terpenes and terpenoids that can be used to treat different illness (Cheenpracha et al. 2006, Bhamarapravati et al. 2006) . More focus has been put on the role of this fingerroot plant in medical treatment as it has exhibited antiviral, antimicrobial, antioxidant, anticancer, antimutagenic and antileukemia activities that will be described in more details on the following subtopics.

2.1.4.1 Antiviral Activity

(a) *Anti-HIV-1 Protease Activity*

A group of researchers has successfully characterized anti-HIV-1 protease inhibitor to exhibit antiviral activity against HIV disease (Cheenpracha et al. 2006). Several compounds were extracted from rhizomes of *B. pandurata* and classified as chalcone derivatives such as panduratin C, panduratin A, hydroxypanduratin A, helichrysetin, 2',4',6'-trihydroxyhydrochalcone, and uvangoletin. Based on their results, hydroxypanduratin A and panduratin A were active against HIV-1 PR, possessed with IC_{50} values of 5.6μM and 18.7μM, respectively. The inhibition of these compounds was related to their structure-activity relationships such as hydroxylation and prenylation of chalcone that gave higher activity.

Tewtrakul et al. (2003) investigated chloroform-, and methanol-and water-extracts from different parts of twelve Thai medicinal plants. According to their report, chloroform extract from rhizome part of *B. pandurata* exhibited the most potent inhibition against HIV-1 protease (HIV-1 PR) (64.92±4.75%), followed by methanol extract with the inhibition of 51.92±0.22%, at concentration of 100μg/mL as compared to other Thai local plant species such as *Alpinia galangal* (rhizome part, methanol extract; 48.70±1.21) and water extract from whole plant of *Eclipta prostrata* (42.53±2.30). In another study, antiviral assay showed that among four flavonoids (pinocembrin, pinostrobin, cardamonin and alpinetin) that was isolated from an ethanol extract of *B. pandurata* rhizomes, cardamonin showed most potent against HIV-1 PR with an IC_{50} value of 31.0μg/mL as compared with other compounds that only exhibited mild inhibitory activities (Tewtrakul et al. 2003).

(b) *Inhibition of Dengue NS2B/NS3 Protease*

The inhibition of dengue virus has been studied against dengue NS2B/NS3 protease cleavage (Tan et al. 2006). Compounds such as 4-hydroxpanduratin A and panduratin A showed the strongest inhibition, with inhibition constant, K*i*, values of 21μM and 25μM, respectively, as compared to the other compounds such as pinocembrin, pinostrobin, cardamonin, and alpinetin. For protease cleavage, different substrates were tested using substrate 1 (Boc-Gly-Arg-Arg-MCA), with 4-hydroxypanduratin A showing most potent inhibition for NS2B/NS3 protease with (78.1±0.1%) than panduratin A (66.7±0.1%) at 80ppm. Meanwhile, when using substrate 2 (Boc-Gln-Arg-Arg-MCA), 4-hydroxypanduratin A still showed the highest inhibition at 90% at a concentration of 120ppm, compared to pinostrobin and panduratin A.

2.1.4.2 Antimicrobial activities

(a) *Anti-Helicobacter pylori Activity*

Bioactive components extracted from plant could potentially serve as drug development for inhibition of *H. pylori* infection. Two compounds extracted from roots of *B. rotunda* were isolated for *in vitro* anti- *H. pylori* test; pinostrobin and red oil against several different isolates of *H. pylori*. The test result showed that minimum inhibitory concentration (MIC) for pinostrobin and red oil were 125µg/ml and 150µg/ml, respectively, which were comparable to the positive control, clarithromycin (120µg/ml); meanwhile, the minimum bactericidal concentration (MBC) was determined to be 150µg/ml and 175µg/ml, respectively (Bhamarapravati et al. 2006).

(b) *Pathogenic bacteria inhibition activity*

Although some bacteria were previously considered to be non-pathogens, several of them are known to cause diseases. Pathogenic bacteria could induce diseases in humans and plants. Different strains of pathogenic bacteria such as *Salmonella enterica*, *Escherichia coli*, *Listeria monocytogenes*, *Bacillus cereus*, and *Staphylococcus aureus* were tested with ethanolic extracts from four different species; *Zingiber officinale* Rosc., *Cucurma longa* Linn., *Alpinia galangal* Stuntz., and *Boesenbergia pandurata* Schltr to study antimicrobial activity by using agar dilution method (Pattaratanawadee et al. 2006). Among four species, fingerroot showed the highest bacteriostatic effect against three pathogenic bacteria; *L. monocytogenes*, *B. cereus*, and *S. Aureus*, with MIC value of 0.2-0.4% (v/v). In addition, Rukayadi et al. (2009) reported that Panduratin A isolated from *B. rotunda* possessed antimicrobial activity against *Staphylococcus* strains with MIC_{50} of 0.5µg/mL and MIC_{90} of 1µg/mL, in which both values were comparable to the most potent antibiotic, vancomycin.

2.1.4.3 Antioxidant activity

(a) *Inhibition of Lipid Peroxidation in Brain*

Six bioactive compounds were extracted from rhizome of *B. rotunda* to examine the spread of free-radical scavenging activities on rat brain homogenate model. These compounds included panduratin A, 4-hydroxypanduratin A, 2',6'-dihydroxy-4'-methoxychalcone, 2',4'-dihydroxy-6'-methoxychalcone, 5-hydroxy-7-methoxyflavanone, and 5,7-dihydroxyflavanone. The lipid peroxidation in tested rat brain was successfully inhibited and revealed two potent inhibitors, panduratin A with IC_{50} value of 15μM and 4-hydroxypanduratin A with IC_{50} value of 4.5μM (Shindo et al. 2006). Both compounds also possessed high neuroprotective effect with an effective concentration, EC_{50} value of 13μM and 14μM, respectively, as compared to (+)-catechin (160μM) through L-glutamate toxicity.

2.1.4.4 Anticancer

(a) *Breast Cancer and Colon Cancer Prevention*

Eleven species of Zingiberaceae were examined for anticancer activity (Kirana et al. 2007). Among them, *Boesenbergia rotunda* and *Zingiber aromaticum* were found to exhibit the highest inhibition towards MCF-7 breast cancer (MCF-7) with IC_{50} value of 21.3±0.3μg/mL and 32.5±1.5μg/mL, respectively. Moreover, both species also showed potent activity on human HT-29 colon cancer (HT-29) with the IC_{50} value of 20.2±1.8μg/mL and 11.8±1.0μg/mL, respectively. Morphological studies showed the apoptosis of cell death by observing the appearance of membrane blebs, nuclear condensation, and formation of apoptotic bodies.

Additionally, the growth of MCF-7 breast cancer cell and HT-29 colon cancer cell was potentially inhibited by panduratin A extracted from *B. rotunda* with the IC_{50} value for

MCF-7 and HT-29 cells were determined to be 3.75μg/mL and 6.56μg/mL, respectively. After treating with panduratin A, 71% of the cells were arrested at G0/G1 as compared to 33% for untreated cells according to cell cycle and proliferation studies. Meanwhile, cytotoxic activities of *B. rotunda* volatile oils against breast cancer MCF-7 (IC_{50} value of 31.7±5.4μg/mL) and LS174T colon cancer (IC_{50} value of 12.0±1.6μg/mL) cell lines were also reported by Zaeoung et al. (2005).

With the flurry of publications and extensive ongoing research in searching useful treatment to combat human diseases, researchers have changed route to focus on *B. rotunda* plants as a potential alternatives to cure disease and infection. Based on the evidences given, plant-derived compounds extracted from its fresh rhizomes can be further developed into plant-based drugs together with deeper knowledge on biosynthetic pathways.

2.2 Chalcone synthase (CHS)

2.2.1 Background

With the remarkable of various pharmaceutical activities of bioactive compounds found in *B. rotunda* plants, chalcone synthase (CHS) has been the most exclusively studied enzymes as it generates backbones of a variety of plant secondary metabolites in flavonoid biosynthesis pathways. Chalcone synthase (CHS; EC 2.3.1.74) belongs to the CHS superfamily of plant type III polyketide synthases (PKSs) which resembles closely the condensing enzyme of type-II fatty acid synthesis with respect to several of its catalytic and structural properties (Austin & Noel, 2003). Since the first isolation of *CHS* from parsley (*Petroselium hortense*) in 1983, the study of this enzyme is now rapidly progressing with more than 1000 of different plant species successfully cloned and characterized CHS enzyme (Kreuzaler et al. 1983).

2.2.2 Structure of CHS

CHS is the most abundant of all enzymes and has a conserved homodimeric protein with a monomer size of 40 to 45kDa. Generally, it has a size about 1170bp, coding for 390 amino acids (Sun et al. 2015; Wiriyaampaiwong et al. 2010). When comparing *CHS* from different plant species, it is revealed that *CHS* are structurally conserved with most of them containing two exons (exon 1 encodes about 37 to 64 amino acid residues while exon 2 encodes about 340 amino acid residues) (Yang et al. 2002). The most characteristic feature of the core region is its location in the second exon, where the majority active residues of the catalytic region are contained.

The length of the intron varies from less than 100 bp to several kilobases (Oberholzer et al. 2000). In general, *CHS* consists of single intron flanked by two exons in most plant species, such as in *Curcuma longa* (Wannapinpong et al. 2013) and *Gingko biloba* (Pang et al. 2005) with the only exception of *CHS* from *Antirrhinum majus* which contains two introns and three exons (Radhakrishnan et al. 2010) and *Physcomitrella patens* which can be intronless, one intron or two introns (Jiang et al. 2006). Members of CHS family are homodimers of 40-45kDa subunits, and share high degree of similarity in their amino acid sequence, structure, and catalytic principles, and possess a conserved of Cys-His-Asn catalytic triad in the buried active sites (Austin & Noel, 2003).

2.2.3 Multiple copy variants of *CHSs*

There is a high variation in the number of CHS copies among plants in which the expression is directly related to the functional diversity of several flavonoids. Based on previous literatures, only a single copy of the *CHS* was found in *Arabidopsis thaliana* (Feinbaum & Ausubel, 1992). Meanwhile, in *Zea mays* and *Hordeum vulgare*, most of them were found with two copies of the *CHS* (Franken et al. 1991) while five copies were found in morning glory (Durbin et al. 2000). In addition, a total of eight-member

CHS family were detected in *Glycine max*. The first subfamily; *CHS1* to *CHS6* contributed 93% nucleotide sequence identity meanwhile, the second subfamily; *CHS7* and *CHS8* contributed to most of *CHS* transcripts and shared 97.0% similarity at the nucleotide level (Tuteja et al. 2004).

In other plants such as *Petunia hybrida*, eight complete *CHS* and four partial genes were cloned and sequenced (Koes et al. 1989). At least eight *CHS* members were identified in pea (*Pisum sativum*) (Ito et al. 1997) and fourteen *CHS* were identified in maize that present unevenly located on six chromosomes (Han et al. 2016). However, the main problem discussed in previous studies is the difficulty to estimate the expression of individual members of a multigene family because of the high similarities in the gene coding region and flanking sequences that rendered difficulty in having differences between them. On the other hand, *B. rotunda's* number of *CHS* are lack in report and this study is working on to characterize the gene from different plant sources and analyze those sequences. All of the CHS family participates in the biosynthesis of various secondary metabolites which play significant roles in flower pigmentation, pollen fertility, protection against ultraviolet (UV) and defense against herbivores and pathogens (Dao et al. 2011; Jenkins et al. 2001; Li et al. 2016; Mo et al. 1992).

2.2.4 Enzymatic reaction of CHS

CHS-superfamily type III PKS is a well-studied enzyme that catalyzes sequential decarboxylative condensation of malonyl unit with a CoA-linked starter molecule. It commonly catalyzes the condensation of one molecule of (a) p-coumaroyl-CoA and three (b) malonyl-CoA molecules to produce a new aromatic ring system, (c) naringenin chalcone, the key intermediate in the biosynthesis of flavonoids as shown in Figure 2.2 (Jez et al. 2000).

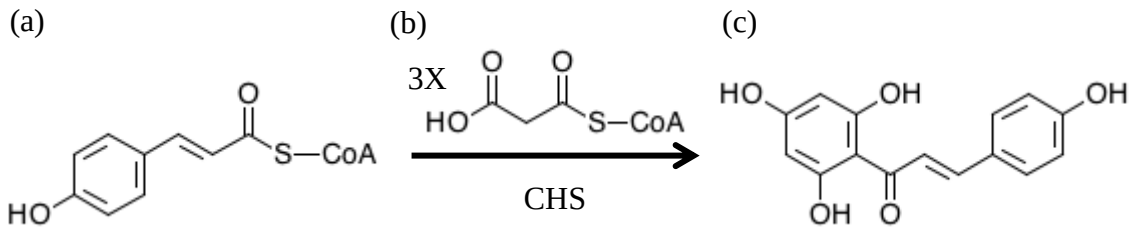


Figure 2.2: Schematic diagram of a condensation reaction

Condensation reaction is catalyzed by CHS enzyme with one molecule of a) p-coumaroyl-CoA and three molecules of b) malonyl-CoA to produce c) intermediate naringenin chalcone.

The functional diversity of CHS-superfamily type III PKSs depends on i) the number of malonyl-CoA it condenses as the numbers vary between one to eight molecules, ii) the utilization of different CoA-linked starter substrates ranging from aliphatic-CoA to aromatic-CoA, from small acetyl-CoA to bulky N-methylanthraniloyl-CoA substrates, and from polar malonyl-CoA to nonpolar n-hexanoyl-CoA substrates as well as iii) the mechanisms by which the resulting intermediates are cyclized and aromatized (Morita et al. 2010). Examples of other plants reacted with different types of substrates and their products are presented in Table 2.1.

Claisen and aldol cyclization are two different types of cyclization reaction in which Claisen-type cyclization is catalyzed by CHS (Abe & Morita, 2010). Taking stilbene synthase (STS) as an example, it condenses three acetate units of Malonyl-CoA with a phenylpropanoid CoA ester. Unlike CHS, STS cyclizes the intermediate by aldol-type, in which CO_2 is lost to generate resveratrol, a phytoalexin stilbene (Abe & Morita, 2010).

Table 2.1: Examples of substrates react with different type of CHS to produce different products

CHS - type (Claisen, aromatic)	**Substrate**	**Product**	**Plants**	**Author(s)**
Chalcone synthase (CHS)	p-Coumaroyl-CoA	Naringenin chalcone	*Medicago sativa*	Austin & Noel, 2003; Ferrer et al. 1999
Phlorisovalerophenone synthase (PVS)	Isovaleryl-CoA	Phlorisovalerophenone	*Humulus lupulus*	Okada & Ito, 2001; Paniego et al. 1999
Isobutyrophenone synthase (BUS)	Isobutyryl-CoA	Phlorisobutyrophenone	*Hypericum calycinum* *Centaurium erythraea*	Klingauf et al. 2005
Benzophenone synthase (BPS)	3-Hydroxybenzoyl-CoA Benzoyl-CoA	2,3',4,6-Tetrahydroxybenzophenone 2,4,6-Trihydroxybenzophenone	*Hypericum androsaemum*	Beerhues, 1996; Liu et al. 2003
Acridone synthase (ACS)	*N*-Methylanthraniloyl-CoA	1,3-Dihydroxy-*N*-methylacridone	*Ruta graveolens* *Huperzia serrata*	Junghanns et al. 1998; Springob et al. 2000; Wanibuchi et al. 2007
Homoeriodictyol/eriodictyol synthase (HEDS or HvCHS)	Feruloyl-CoA Caffeoyl-CoA	Homoeriodictyol chalcone Eriodictyol chalcone	*Hordeum vulgare*	Christensen et al. 1998

2.3 Flavonoids as secondary metabolites in plants

2.3.1 Structure and subclasses of flavonoids

Phenolic structures of flavonoids come in diversified form and are ubiquitous in nature. They have been documented from higher plants and microbes with an estimated of 6,000 different flavonoids (Ferrer et al. 2008). The core flavonoids structure is based upon a fifteen-carbon skeleton with two aromatic benzene rings that contain six carbon atoms (ring A and B) interconnected by a heterocyclic pyrene ring (ring C) that contains three carbon atoms (Shashank & Abhay, 2013). Therefore, their structure is also referred to C6-C3-C6 as shown in Figure 2.3.

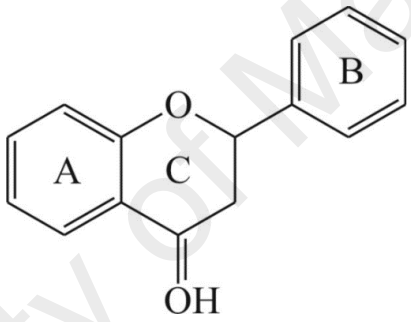


Figure 2.3: Basic structure of flavonoid

Heterocyclic pyrene Ring A and B with 6 carbon atoms and Ring C with 3 carbon atoms, build a link of C6-C3-C6 flavonoid structure.

Various flavonoids can be subdivided into different subclasses depending on the pattern of substitution of the C ring on which B ring is attached, and according to the degree of unsaturation and oxidation of the C ring (Halbwirth, 2010). They can be divided into a variety of subclasses of bioactive compounds such as flavanols (catechin, epicatechin), flavonols (quercetin, myricetin, kaempferol, rutin), flavones (apigenin, luteolin, diosmin), flavanones (naringenin, alpinetin, pinostrobin, pinocembrin, naringin, hesperidin), chalcones (boesenbergin, cardamonin, panduratin A, hydroxypanduratin A),

aurone, anthocyanidins and proanthocyanidin (Falcone et al. 2012; Shashank & Abhay, 2013). The different classes are shown in Figure 2.4.

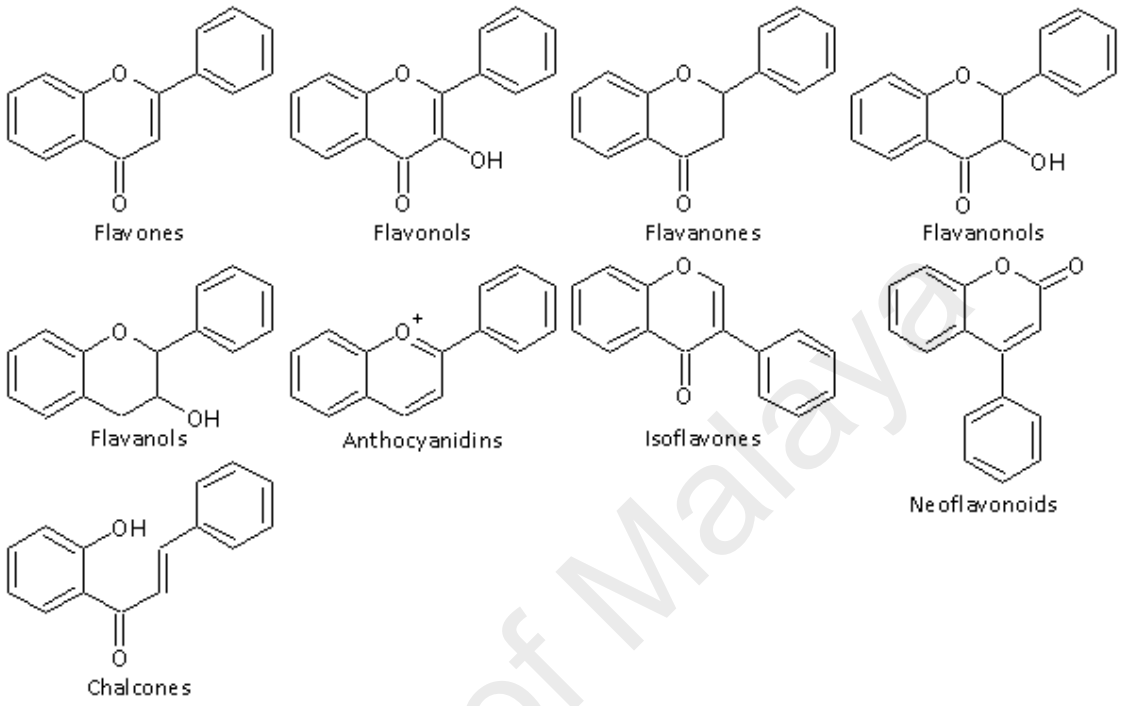


Figure 2.4: Flavonoid subclasses

According to the substitution patterns of the central C ring, flavonoids can be divided into multiple subclasses such as flavones, flavonols, flavanones and more.

2.3.2 Biosynthesis pathway of flavonoids

Plants contain countless metabolic pathways that are responsible for the biosynthesis of several secondary metabolites that occur in a broad diversity and functionality. In general, flavonoids are synthesized through the phenylpropanoid and polyketide pathway, converting phenylalanine into p-coumaroyl-CoA and integrate within flavonoid biosynthesis pathway (Austin & Noel, 2003). When entering flavonoid biosynthesis pathway, chalcone synthase (CHS) performs as a starting enzyme that regulates the pathway and creates fifteen structure backbone from which all different flavonoids are derived from (Shashank & Abhay, 2013).

CHS catalyzes sequential condensation and decarboxylation reactions of two substrates (malonyl-CoA and p-Coumaroyl-CoA) in the course of which a polyketide intermediate is formed (Ferreyra et al. 2012). The polyketide undergoes cyclization and aromatization leading to the formation of naringenin chalcone (2′,4,4′,6′-tetrahydroxychalcone), the first precursor to be synthesized by plants (Falcone et al. 2012; Winkel-Shirley, 2001a).

Various biosynthetic enzymes further down this pathway such as chalcone isomerase (CHI), isoflavone synthase (IFS), flavanone-3-hydroxylase (F3H), dihydroflavonol reductase (DFR), flavonol synthase (FLS), glucosyltransferase (GT) and many other important enzymes are accountable for catalyzing the conversion of chalcones into different flavonoid molecules (Mouradov & Spangenberg, 2014). Investigations on molecular aspects of flavonoid biosynthesis pathway have been reviewed by researchers (Saito et al. 2013; Winkel-Shirley, 2001a, 2001b). An overview of the flavonoid biosynthesis pathway is presented in Figure 2.5.

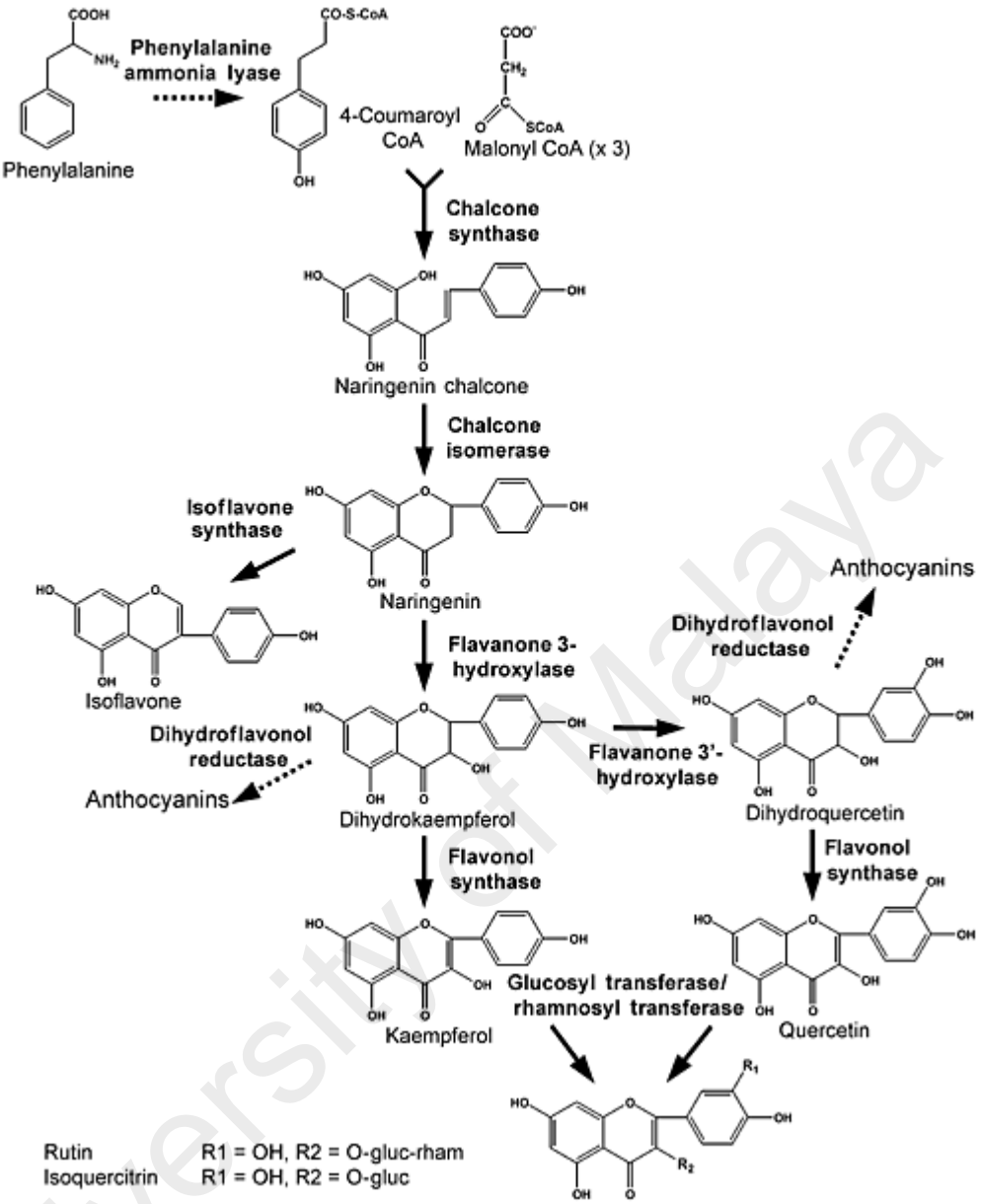


Figure 2.5: Plant pathways relevant to CHS and resulting classes of natural products

Different types of compounds are synthesized via reaction of the key enzyme; CHS with 4-coumaroyl-CoA and malonyl-CoA as substrates.

B. rotunda is rich in flavonoids and these compounds are produced in the rhizomes. The structure of flavonoid compunds in *B. rotunda* are classified into two main groups, flavanones (alpinetin, pinostrobin, and pinocembrin), chalcones (pinocembrin chalcone, pinostrobin chalcone, boesenbergin, cardamonin) and chalcone cyclohexenyl derivatives

(CCDs) such as panduratin A, and 4-hydroxypanduratin A (Tan et al. 2015; Yusuf et al. 2013).

Flavonoids serve diverse biological functions including playing a major role in plant responses to environmental cues, in particular during biotic and abiotic stresses. It protects plant against abiotic stress (ultraviolet (UV) radiation and heat) and biotic stress (herbivores, pathogens) (Dao et al. 2011). It induces root nodulation when excreted by symbiotic nitrogen-fixing rhizobia (Mandal et al. 2010), as well as the coloration of flowers as a visual signal that attracts pollinators (Winkel-Shirley, 2001a).

2.4 Plant cell culture as a source of high-value natural metabolites

The production of secondary metabolites such as terpenoids, alkaloids, polyketides, phenylpropanoids, and flavonoids derived from plant cells offer an alternative approach to chemical synthesis. Extraction of these plant compounds is valuable and economical for pharmaceutical drug sources and is traditionally obtained from plants growing in the wild or from field cultivated sources. For example, ginsenoside that is used for pharmaceutical applications, extracted from *Panax quinquefolium*, usually takes four to six years to harvest in large-scale (Zhang et al. 1996). However, the extraction of these molecules from plant sources does not result in quantities large enough to meet the increasing market demands.

In addition, there have been immense challenges for chemists to synthesize plant-derived compounds via organic chemistry. An attempt has been made to develop an efficient method for the synthesis of panduratin A, through Diels-Alder cyclization reaction; however, the overall yield actually depended on variety of diene under moderate environment and it continued to hinder large-scale production of this compound and hampered by the cost of their synthesis that outweighs their commercial availability (Chee et al. 2010). Where natural harvesting of plants is limited due to slow growing and non-

feasible for the chemical synthesis from an economic point of view, plant cell culturing offers potential strategy and cost-effective approach to meet popular market demand (Rao & Ravishankar, 2002, Jedinak et al. 2004).

2.5 Plant suspension cultures as an alternative method for compound synthesis

A flurry of researches on plant organs and suspension cultures has been investigated for the production of desired natural products. When aiming for large-scale production of desired compound, the yield remains a bottleneck especially for root and shoots cultures (plant organs). These differentiated cells produce the same compound as the parent plant, however, during dedifferentiation stage, it results in loss of production capacity (Verpoorte et al. 2002). Among different approaches of plant tissue cultures techniques, suspension culture is capable of accelerating production of specific medicinal compounds at a rate similar or superior to that of intact plants. Suspension cultures can be initiated by transferring fragments of friable callus tissue to a liquid medium, which is constantly agitated and cause cells of friable callus tissue to readily dissociate.

Due to its homogeneity in the cell population, undifferentiated suspension cultures owing to the fast propagating rate, ease of scaling-up by growing under controlled conditions and having high rate of cell growth (Rasche et al. 2016, Ramachandra Rao & Ravishankar, 2002). Previously, an approximate of 0.01% of the dry weight of paclitaxel was produced in bark of *Taxus brevifolia* (Rao et al. 1995; Witherup et al. 1990). Due to the limited supply of paclitaxel in plants, *Taxus* cell lines was established to produce paclitaxel by elicitation with methyl jasmonate in a range of 140-295 mg/L (Tabata, 2006). Examples of other high-value secondary metabolites obtained from variety of plants species, cultivated via suspension cultures are listed in Table 2.2.

Table 2.2: List of high-value natural bioactive compounds isolated from plants, cultivated via suspension cultures

Compounds	**Main use**	**Species**
Abietane diterpenoids	Antitumour	*Cephalotaxus fortune*
Alkaloids	Antimicrobial	*Ailanthus altissima*
Ajmalicine	Antihypertensive	*Catharanthus roseus*
Artemisinin	Antimalarial	*Artemisia annua*
Anthocyanin	Antioxidative	*Vitis vinifera* *Fragaria ananassa*
Anthraquinones	Antimicrobial	*Morinda elliptica* *Rubia tinctorum* *Cassia acutifolia*
Antifungal monoterpene	Antifungal	*Piqueria trinervia*
Berberine	Intestinal ailment	*Coptis japonica* *Thalictrum minus*
Camptothecin	Antitumour	*Camptotheca acuminate*
Canthinone alkaloids	Antitumour Antimalarial	*Brucea spp.*
Capsaicin	Counterirritant	*Capsicum frutescens*
Cerebroside	Regulation of cell growth	*Lycium chinense*
Cholecalciferol	Calcium absorption	*Solanum malacoxylon*
Chlorogenic acid	Antimicrobial Antioxidative	*Eucommia ulmoides*
Codeine	Sedative	*Papaver somniferum*
Colchicine	Antitumour	*Colchium autumnale*
Coumarins	Anticoagulant	*Ammi majus*
Crocin	Anticancer	*Crocus sativus*
Cryptotanshinone	Antioxidative Antimicrobial	*Salvia miltiorrhiza*
Digoxin	Heart stimulant	*Digitalis lanata*
Diosgenin	Steroidal precursor	*Dioscorea deltoidea*
Dipyranocoumarins	Anti-HIV	*Calophyllum inophyllum*
Eleutheroside	Analgesic Anti-inflammatory Antipyretic Diuretic	*Eleutherococcus sessiliflorus*
Ellipticine	Antitumour	*Ochrosia elliptica*
Emetine	Antiparasitic	*Cephaelis ipecacuanha*
Forskolin	Bronchial asthma	*Coleus forskolii*
Furanocoumarin	Antitumour Antioxidative	*Glehnia littoralis*
Furoquinoline alkaloids	Antitumour Antimicrobial	*Choisya ternata*
Ginkgolides	Health tonic	*Ginkgo biloba*
Ginsenosides	Health tonic	*Panax ginseng*
Gymnemic acid	Antidiabetic	*Gymnema sylvestre*
Hispidulin	Antitumor	*Saussurea medusa*
Homoisoflavonoids	Antimicrobial Antitumour	*Caesalpinia pulcherrima*

Table 2.2, continued. (Adapted from Yue et al. (2014))

Hypericin	Antidepressive	*Hypericum perforatum*
Inulin	Diabetics	*Helianthus tuberosus*
Isoquinoline alkaloids	Antitumour Antioxidative	*Fumaria capreolata*
Jaceosidin	Antitumour	*Saussurea medusa*
L-dihydroxyphenylalanine	Anti-Parkinson	*Mucuna pruriens*
Morphine	Sedative	*Papaver somniferum*
Phenolic compound	Arthritis Digestive disorders Rheumatism Venereal diseases	*Larrea divaricata*
Phenylethanoid glycosides	Anhrodisiacs Antioxidative Anti-inflammatory Antihypertensive	*Cistanche salsa* *Olea europaea*
Phytoestrogens	Health tonic	*Psoralea corylifolia*
Plumbagin	Anticancer Antimicrobial Antifertil	*Drosophyllum lusitanicum*
Podophyllotoxin	Antitumour	*Podophyllum*
Polyphenols	Hpyerglycemic Antimicrobial	*Cornus kousa*
Quassin	Antiphlogistic	*Picrasma quassioides*
Reserpine	Antihypertensive	*Rauwolfia serpentina*
Resveratrol	Health tonic	*Vitis vinifera*
Robustaquinones	Antimalarial	*Cinchona robusta*
Rosmarinic acid	Antioxidative	*Anchusa officinalis* *Salvia officinalis* *Coleus blumei* *Orthosiphon aristatus*
Sanguinarine	Antiplaque	*Sanguinaria canadensis*
Shikonin	Antibacterial	*Lithospermum erythrorhizon*
Silymarin	Liver ailment	*Silybum marianum*
Tanshinone	Cardiac disorders	*Salvia miltiorrhiza*
Taxol	Anticancer	*Taxus brevifolia*
Theamine	Antihypertensive	*Camellia Sinensis*
Triterpene	Antiinflammation Antitumor Antidiabetes	*Eriobotrya japonica*
Vinblastine	Antileukemic	*Catharanthus roseus*
Withanolide A	Health tonic Antistress	*Withania somnifera*

As mentioned in Table 2.2, there are many different plant-derived compounds which have been extracted from different plant species. Even though propagation of suspension cultures can be scaled up with established tissue culture techniques, the yield of valuable compounds is often hindered in suspension cultures. Notably, the accumulation of secondary metabolites is controlled in a tissue-specific manner (Verpoorte et al. 2002); thus, the genetically unstable and undifferentiated cells often lose, partially or totally, their ability to produce secondary products (Hendrawati et al. 2012). For example, artimisinin was not discovered in suspension cultures of *Artemisia annua*; however, considerable amounts were detected in shoot cultures (Liu et al. 2006). Meanwhile, other study also reported that trace amounts of Deoxypodophylltoxin were detected in callus and suspension cultures of *Anthriscus sylvestris* (Koulman et al. 2003).

Due to its limitation, it is vital to take other approaches to ensure continuous supply of compounds in order to meet the commercial and market demand. Through the advancement of molecular biology techniques, which include cloning, recombinant DNA, flooding of genome information and knowledge of various metabolic pathways in plant cell, metabolic engineering approach offers an advantage for production of the desired natural products specifically in suspension cultures.

2.5.1 Aim and strategies in metabolic engineering

Generally, metabolic engineering is defined as the redirection of one or more enzymatic reactions to yield a novel plant-derived compound by introducing the appropriate heterologous genes, to improve the production of existing compounds, or to mediate the degradation of compounds (DellaPenna, 2001; Verpoorte et al. 2002). On top of that, this method helps to improve the understanding of plant pathways regulation since many biosynthetic pathways are complex and complicated, which requires multiple enzymatic steps to produce desired high yield compound (Kirakosyan & Kaufman, 2009).

Therefore, in this study, the target is to increase the amount of desired compounds. There are several metabolic engineering strategies applied in plants to enhance the production level of a compound of interest that is present in trace amounts. The appropriate strategies are demonstrated in Figure 2.6.

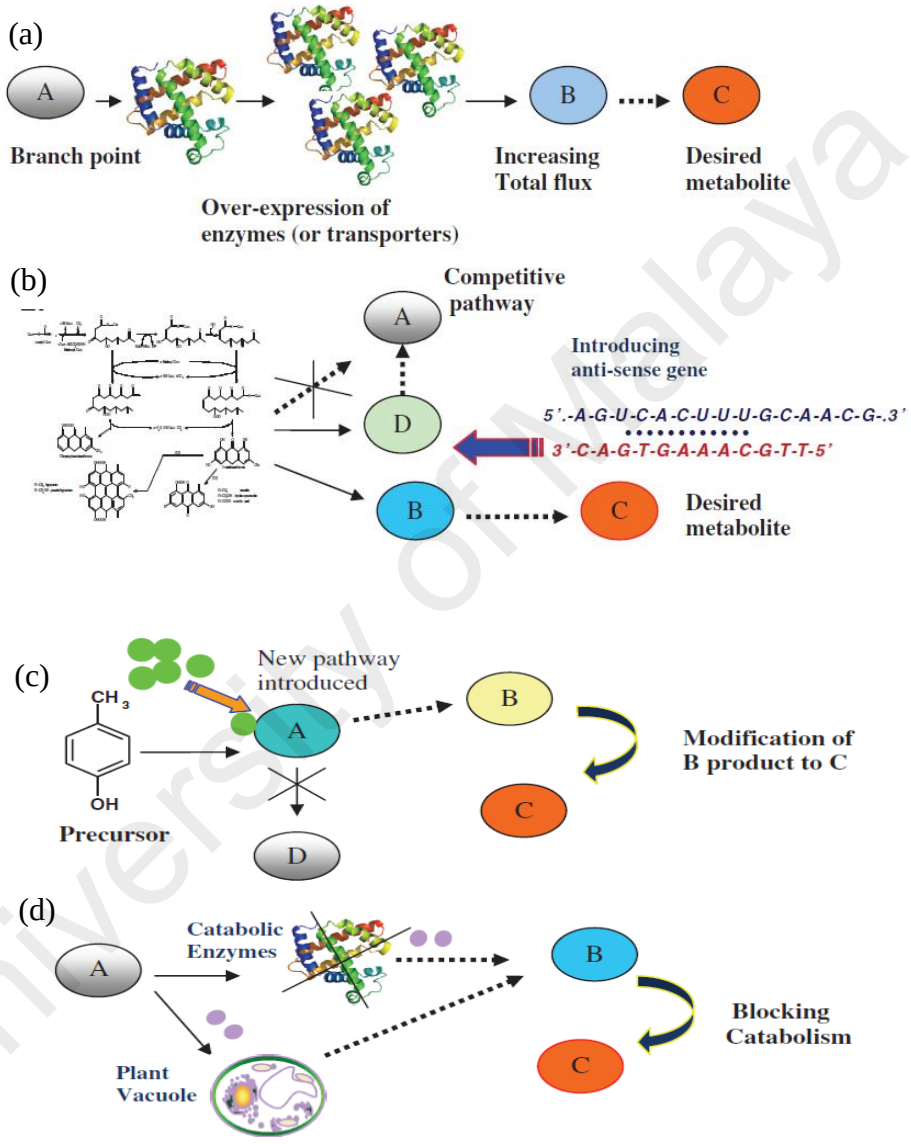


Figure 2.6: Strategies in metabolic engineering

Different strategies in metabolic engineering involves (a) increasing in the total carbon flux in the pathway, (b) decreasing in the total flux through competitive pathways or introduce antisense gene, (c) modifying desired metabolite by targeting of rate-limiting steps or by introducing a new pathway; and (d) blocking catabolism by increasing the metabolites transport into vacuole (Adapted from Kirakosyan & Kaufman, 2009).

Such studies on understanding the basic metabolic processes involved in metabolites biosynthesis should be focused on systematic metabolic engineering process. Thus, armed with the available tools in sequencing and bioinformatics, exploring and discovering the targeted gene has become easier. Integration of targeted gene into the plant genome can be mediated by *Agrobacterium tumefaciens* to facilitate gene engineering of secondary metabolic pathways (Rajesh et al. 2013; Ribas et al. 2011). Gene engineering is feasible to be successfully achieved by the highest possible levels of desirable compounds by the overexpression of genes controlling the limiting steps or by suppressing the undesired product biosynthesis.

Metabolic engineering by regulating *CHS* can be positively correlated with the metabolites accumulation. For example, the regulation of *CHS* correlates with the accumulation of metabolites as these genes are constitutively expressed during plant development which is restricted to floral tissues and caused accumulation of anthocyanin compounds. For example, Sun et al. (2015) conducted overexpression of *FhCHS1* gene in *Petunia hybrida* plants and observed that the white flower color changed to pink due to the accumulation of anthocyanin metabolites. Similarly, in *Nicotiana tabacum*, deeper red petal color was observed when *McCHS* in the plant was overexpressed (Tai et al. 2014). Meanwhile, in Guoqing No. 4 satsuma mandarin, decrease of *CHS* expression gradually in peels caused lower yield of total flavonoid, trans-chalcone, narirutin and hesperidin compounds (Wang et al. 2010).

All these examples demonstrated the wide range of metabolic engineering approaches that can be applied to plant pathways. It can also help to illustrate the range of problems for the production of novel high-value pharmaceutical compounds that could not be detected in wild type plants.

2.6 Gene expression and pathway exploration

2.6.1 Emergence of NGS platform

More than 20 years ago, obtaining thousands of bases of sequence information from a core facility marked a productive day. Technologies such as quantitative real-time polymerase chain reaction (qPCR) and microarrays are powerful tools for analyzing gene expression. However, the sensitivity and the accuracy of measured fold changes in the gene expression arrays are limited. Thus, both approaches cannot be scaled to transcriptome-level complexity and lack the ability to discover novel information on a large scale.

Noticeably, gene expression patterns drive the molecular mechanisms that govern the broad range of biological questions of interest. To address these barriers, NGS platform has continued to generate hundreds of millions of bases and produced knowledge regarding the quantitative and qualitative aspects of transcriptomes in most RNA species inside a cell (Wang et al. 2009). Since the first report of NGS, the application is growing exponentially and later the birth of this technology is announced to the world of RNA-Seq (Snyder, 2010). With the comprehensive and dramatically improved technology as well as data acquisition and analysis, this strategy has brought high reproducible results (Martin & Wang, 2011; Ozsolak & Milos, 2011; Z. Wang *et al.* 2009).

Good experimental designs before applying RNA-Seq as a platform for sequencing, are well established to achieve validity and efficiency of every experiment which can be employed by choosing the library type or preparation, sequencing depth and number of replicates. To perform a library from the RNA of interest, one important aspect during RNA extraction protocol is to remove the abundant ribosomal RNA (rRNA) that are generally not of interest. Poly-A enrichment for messenger RNA (mRNA) samples is desirable and can be achieved by targeting the polyadenylated (poly(A)) tails to ensure

that rRNA is separated from polyadenylated transcripts, leaving only the 1-2% mRNA (Guo et al. 2015). The removal of rRNA can also be done by ribosomal depletion especially for non-polyadenylated mRNA of bacterial samples (Ciulla et al. 2010). Therefore, it is crucial to produce high quality poly(A) RNA-Seq libraries and high proportion of mRNA integrity with minimal degradation as this can be measured by RNA integrity number (RIN).

Furthermore, RNA-Seq can involve single-end (SE) or paired-end (PE) reads. The latter is preferable as it allows users to sequence both ends of a fragment and generate high-quality, alignable sequence data. PE sequencing facilitates different applications such as detecting gene fusions in cancer, characterizing novel splice isoforms or *de novo* transcript discovery (Li et al. 2015; Sakarya et al. 2012; Yarden, 2011). PE sequencing together with longer reads improve 'mappability' and transcript identification of poorly annotated transcriptomes (Łabaj et al. 2011). Figure 2.7 showed an example of paired-end sequencing and the alignment of reads.

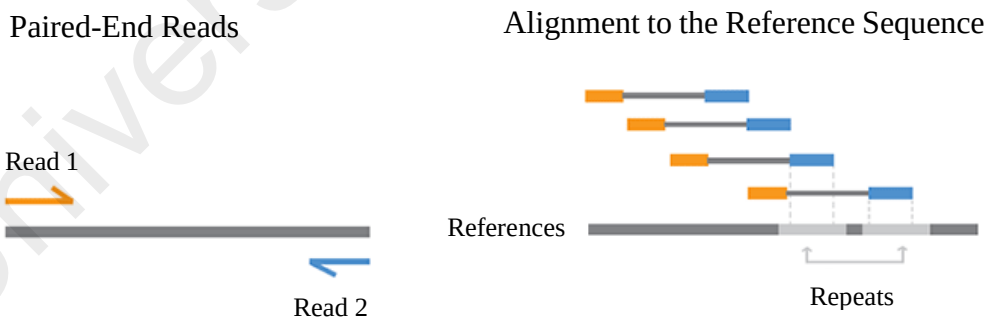


Figure 2.7: Paired-end sequencing and alignment

Paired-end (PE) sequencing allows both ends of fragment to be sequenced and produce PE reads with known distance between each paired read. Thus, PE reads are aligned over repetitive regions more precisely across difficult-to-sequence and repetitive regions of genome (Treangen & Salzberg, 2013).

Another important aspect is sequencing depth or coverage, which is referred as the number of total mapped sequences for a given sample. Sequencing depth is chosen based on a given budget and experimental target that produce a large number of transcripts. Different sequencing depth was set out for the purpose of detecting transcript expression level. As an example, 5 million mapped reads were reported to sufficiently quantify expressed genes in eukaryotic transcriptome; in contrast, approximately 100 million reads were sequenced to quantify low expression level of transcript (Conesa et al. 2016). Eventhough high sequencing depth is mostly chosen as it generates more reads; however, these high amount of informational reads could detect noise within data and off-targets transcripts (Tarazona et al. 2011). Deep sequencing approach comes with cost, and resources will be wasted in scenarios where more sequencing brings diminishing returns as a saturation level is approached (Liu et al. 2014).

Finally, the number of replicates is also an important aspect in RNA-Seq experiment. There are two levels of replication that should be included in the experiment; namely, biological replicates and technical replicates. The choice of type of replicate depends on the goal of RNA-Seq study, whether to evaluate the technology/process or to examine the biological differences between conditions/tissues/treatments (Lu et al. 2010; Martin et al. 2010; Snyder, 2010) that can significantly effects on differentially expressed (DE) genes. Previously, differential expression analyses were investigated by applying limited biological replication, instead of utilizing deep sequencing strategy (Hah et al. 2012).

Moreover, a critical investigation was also studied by Liu and colleagues, (2014) to provide the answer for the necessity to sequence more samples with low depth or to sequence fewer samples with high depth. Based on their report, the number of DE genes was significantly increased with the increased number of biological replicates and number of reads in each sample. It generated 2709 DE genes when applying three biological

replicates for a total of 30 million combined reads (10 million each) compared with two biological replicates for a total of 30 million combined reads (15 million each) that generated 2139 DE genes. Therefore, by sequencing less reads and performing more biological replication are preferable and more effective as a strategy to increase accuracy in differential expression studies.

With a given budget and goal of sequencing, an efficient planning for determining the number of replicates is important since it helps in estimating and decreasing experimental error during analysis. Therefore, to have an accurate measurement on the genes expressed at low level, one needs to critically decide whether to increase the sequencing depth per sample or to increase the biological replicates by increasing sample size together but with limited sequencing depth.

2.6.2 Gene expression study in *B. rotunda* and other plants

Transcriptome sequencing can be an efficient approach to obtain functional genomic information, which could contribute to pathway mining. A study was attempted to evaluate the gene regulation pattern in flavonoid biosynthesis pathway in phenylalanine-treated suspension cultures of *B. rotunda*. According to their DE genes analysis, with the addition of phenylalanine, the data generated 14,644 up-regulated genes and 14,379 down-regulated genes (Md-Mustafa et al. 2014). Looking at the data, there was not much difference in gene expression level throughout this study. As phenylalanine performed as one of the precursors that can induce the expression of genes in the flavonoid pathway that includes *CHS*, it was found that only one *CHS* unigene was upregulated at 2.5 higher compared to control and 7 *CHS* unigenes were downregulated (range between -1.1 to -2.2-fold lower compared to control).

In addition to their findings, there were several unigenes that may be indirectly involved in the panduratin A biosynthesis, differentially expressed in the

phenylpropanoid and flavonoid pathway. The identified unigenes encoded for shikimate O-hydroxycinnamoyltransferase (4 up-regulated, 12 down-regulated), flavonol synthase (5 up-regulated, 6 down-regulated), naringenin 3-dioxygenase/flavanone-3-hydroxylase (2 up-regulated, 5 down-regulated) and chalcone synthase (1 up-regulated, 7 down-regulated). Even though there was not much difference between DE genes in treated and untreated samples, this data would bring attention in discovering important potential genes and determining the risk of compromised functional pathways in the transcriptome level of *B. rotunda.*

During stress conditions, a number of disease resistance genes are expressed as part of its plant defense, besides being part of the plant developmental program. *CHS* is also induced in plants under different form of stresses such as UV, wounding and in the presence of herbivory or microbial pathogens (Dao et al. 2011). This induction results in the production of compounds that have several activities of antimicrobial (phytoalexins), insecticidal and antioxidant and also compound that quench UV light. As an example, transcriptome from wilt-sensitive *Zingiber officinale* Rosc. and wilt-resistant *Curcuma amada* Roxb. was generated to identify defense related genes that resists *Ralstonia solanacearum* and further identified potential pathways responses to pathogen (Prasath et al. 2014).

Other than that, potential genes associated with fruit and flower pigmentation have also been identified in other plant species. As reported by Wu et al. (2015), in yam tubers, pigment genes were found to produce purple-flesh trait in anthocyanin pathway. Meanwhile, in *Herbaceous peony*, pigment genes that produced yellow flowers were identified via anthocyanin pathway (Zhao et al. 2014). Besides, identification of 8 unigenes characterized as seed-specific oleosins in *Carthamus tinctorius* L. via *de novo* transcriptome was previously found and this study further revealed genes related to

flavonoid biosynthesis since this flower is one of the main source of oil crops in the world (Li et al. 2012).

A large number of transcriptomic and genomic sequences became available in model organisms, such as *Arabidopsis thaliana, Oryza sativa* and *Zea mays*, which have greatly improved the understanding of the complexity of growth and development in higher plants (Dukowic-Schulze et al. 2014; Wakasa et al. 2014; Wang et al. 2014). Accordingly, transcriptome sequencing offered an attractive alternative to generate genetic resources for species that has economic importance to the world. Besides sunflower as mentioned in previous paragraph, examples of such initiatives includes garlic (Sun et al. 2012), fern (Der et al. 2011), garden pea (Franssen et al. 2011), chickpea (Garg et al. 2011) and Japanese knotweed (Hao et al. 2011).

With the advancement of technology, molecular level of the biosynthesis of the targeted plant metabolites, the mechanisms and pathways that contribute to its desired properties can be explored and further characterized. Entire plant genomes can now be sequenced and facilitating the discovery of biosynthetic gene clusters.

CHAPTER 3: METHODOLOGY

3.1 Samples collection and preparation of suspension cultures

B. rotunda plants were collected from a field located in Temerloh, Pahang (GPS: 3.636385, 102.415142). For the purpose of DNA and RNA extraction, plant leaves, flowers and rhizomes were harvested, frozen immediately in liquid nitrogen and stored at -80°C until required. Meanwhile, *B. rotunda* suspension cultures were established from a single bud that emerged from rhizomes. To generate a complete medium for growth, suspension cultures were maintained in Murashige and Skoog (1962) liquid medium (MS) (refer to Appendix A), supplemented with 1 mg/L of 2,4-dichlorophenoxyacetic acid (2,4-D), 0.5 mg/L of 6-Benzylaminopurine (BAP) and 3% (w/v) of sucrose (Wong et al. 2013).

Homogenous suspension cultures were obtained from starter cultures through a sterile filter and were used as inoculum. The suspension cultures were aerated on shakers monitored at 100rpm, and incubated at 25 ± 2°C under a 16 hours photoperiod with a light intensity of 31.4 µmol m^{-2} s^{-1} with cool fluorescent lamps in a growth room. Subcultures were done every ten days. Whenever the amount of settled cell volume (SCV) was high, the suspension cultures were kept maintained by dividing the initial culture in two flasks, and that medium was replaced by a sterilized fresh MS liquid medium. Propagation of suspension cultures was carried out at the Plant Biotechnology Incubator Unit (PBIU), University of Malaya.

3.2 RNA extraction

Total RNA was isolated from frozen leaves, rhizomes, flowers and suspension cultures using a CTAB-based method with a slight modification from a protocol described by Chan et al. (2007). To ensure success working with RNA, it is important to maintain RNase-free environment (refer to Appendix B).

About 250 mg of samples were finely ground in liquid nitrogen and transferred to a tube containing freshly prepared 500 µL RNA extraction buffer (refer to Appendix C(a)) and an equal volume of Chloroform: Isoamyl-alcohol (CI; 24:1). The sample mixture was vortexed vigorously and centrifuged at 12,857 x g for 2 min at room temperature. Supernatant was transferred to a new microcentrifuge tube and purified with an equal volume of Phenol:Chloroform:Isoamyl-alcohol (PCI; 25:24:1), centrifuged at 12,857 x g for 2 min at room temperature and repeated until a clean interphase was observed. Next, supernatant was transferred to a new microcentrifuge tube.

One-tenth volume of 3 M sodium acetate (pH 5.2) was added together with 2.5 volume of cold absolute ethanol and mixed well. The mixture was incubated at 4°C for one hour and centrifuged at 12,857 x g for 20 min at 4°C. After one hour, a pellet was obtained and washed with 70% (v/v) ethanol, centrifuged at 12,857 x g for 5 min, air-dried and 200 µL RNAse-free water (refer to Appendix C(b)) was added to dissolve the pellet. The dissolved pellet was transferred to a new microcentrifuge tube and 2 M Lithium chloride (LiCl) was added and kept at 4°C for overnight incubation followed by centrifugation at 18,514 x g for 20 min at 4°C.

The following day, pellet was washed with 70% (v/v) ethanol, centrifuged at 12,857 x g for 5 min at 4°C and air-dried. Pellet was dissolved by adding 20 µL of RNAse-free water. RNA samples were treated with DNase I, Amp Grade (Invitrogen, USA) with the addition of 1X DNase I reaction buffer and incubated for 15 min at 37°C to remove residual genomic DNA. DNase I was inactivated by the addition of 1 µL of 25 mM EDTA solution to the reaction mixture followed by heating for 10 min at 65°C. Treated RNA sample was then kept at -80°C for longer storage. The total RNA was quantified with NanoPhotometer™ Pearl (IMPLEN, Germany) at 230, 260, 280nm.

3.2.1 Agarose gel electrophoresis

The quality and integrity of RNA were determined using agarose gel electrophoresis. Agarose powder (Promega, USA) was weighed and boiled in 1X Tris Borate EDTA (TBE) buffer to prepare 1% (w/v) agarose gel (refer to Appendix C(c)). Buffer was prepared with RNAse-free water. After the molten gel was cooled down to 55°C, 1µL of Ethidium Bromide (EtBr; Sigma, USA) was added and mixed. Subsequently, the molten gel was poured into a tray containing comb. Equal volume of 2X RNA Loading Dye (Thermo scientific, Lithuania) was added for both; RNA sample and RiboRuler High Range RNA Ladder (Thermo scientific, Lithuania). Both mixtures were heated at 70°C for 15 min, chilled on ice and loaded into the well. The electrophoresis was carried out at voltage of 100V for 25 min and the gel was visualized using Alpha Imager system (Alpha Innotech Corp., USA).

3.3 Reverse transcription-PCR (RT-PCR) of cDNA *BrCHS* fragment

SuperScript™ III One-Step RT-PCR System with Platinum®Taq High Fidelity (Invitrogen, USA) was used for RT-PCR analysis to perform cDNA synthesis and PCR amplification using gene-specific primers at one step. A reaction containing 2X Reaction Mix, 500 ng of template RNA from all parts of *B. rotunda* plants and 10 µM of sense and anti-sense primers was prepared. Primers used were gene specific primers; CHS_F and CHS_R, designed using Sequence Manipulation Suite (Version 2) as listed in the Appendix D. These primers were designed by referring to the *B. rotunda* CHS sequence deposited in NCBI database (Accession no: HQ176338). The reaction was carried out in 25 µL reaction volume following three-steps cycling condition: (a) cDNA synthesis and pre-denaturation: 1 cycle of 50°C for 30 min followed by 94°C for 2 min, (b) PCR amplification: 40 cycles of denaturation at 94°C for 15 sec, annealing at 55°C for 30 sec and extension at 68°C for 1 min per kb, and (c) Final extension: 1 cycle of 68°C for 5 min.

3.3.1 Agarose gel electrophoresis

The quality and integrity of cDNA were determined using agarose gel electrophoresis. Agarose powder (Promega, USA) was weighed and boiled in 1X Tris Borate EDTA (TBE) buffer to prepare 1% (w/v) agarose gel (refer to Appendix C(c)). After the molten gel was cooled down to 55°C, 1 µL of Ethidium Bromide (EtBr; Sigma, USA) was added and mixed. The molten gel was then poured into a tray containing comb. The sample was mixed with 1 µL of 6X loading dye (0.25% of bromophenol blue, 0.25% of xylene cyanole FF) and later loaded into the well. The size of PCR product was estimated using a 1kb DNA ladder (Thermo scientific, Lithuania) as a molecular weight marker. The electrophoresis was carried out at voltage of 120V for 25 min and the gel was visualized using Alpha Imager system (Alpha Innotech Corp, USA).

3.3.2 Gel extraction and purification of *BrCHS* cDNA fragment

After electrophoresis, the gel was then visualized and specific band corresponding to the expected size was sliced out under ultraviolet (UV) illumination. To avoid DNA damage by UV, this step was carried out as fast as possible. The sliced gel was put in a new microcentrifuge tube and purified using QIAquick Gel Extraction Kit according to manufacturer's protocol (Qiagen, USA). Briefly, 3 volumes of Buffer QG were added to 1 gel volume of the sliced gel (100 mg ≈ 100 µL). The gel was then solubilized by incubation at 50°C for 1 min or until the gel slice was completely dissolved. Later, 1 gel volume of isopropanol was added to the sample and the mixture was transferred into a spin column followed by centrifugation at 18,928 x g for 1 min. Then, the flow-through was discarded.

To remove all traces of agarose, 0.5 mL of Buffer QG was added and centrifuged for 1 min. After the flow-through was discarded, 0.75 mL of Buffer PE was added to wash the column and centrifuged for 1 min. After the flow-through was discarded for the

second time, additional centrifugation for 1 min was performed to remove residual ethanol. Next, elution of DNA was done by adding 30 µL sterile distilled water (sdH_2O) to the center of the QIAquick membrane and left to stand for 1 min before centrifuged for 1 min at high speed. Finally, agarose gel electrophoresis was carried out to confirm the present of the purified *BrCHS* cDNA fragment (refer to Section 3.3.1).

3.4 Cloning of BrCHS into *E. coli* (pGEM-T Easy vector system)

3.4.1 Preparation of Luria-Bertani (LB) broth and agar media

Broth and agar medium of Luria Bertani (LB) for bacterial cultures were prepared (refer to Appendix C(d)). For colony selection, agar media was cooled down at least until 55°C before adding 100 µg/mL of ampicillin, 0.5 mM of Isopropyl β-D-1-thiogalactopyranoside (IPTG) and 80 µg/mL of 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal). The media was then poured into petri dishes and stored at 4°C.

3.4.2 Preparation of competent cells, *E. coli* strain JM109

Escherichia coli (*E. coli*), strain JM109 was used in cloning purposes. To prepare competent cells, a single colony of *E. coli* from strain JM109 was inoculated into 10 mL of LB broth medium and grown overnight using incubator shaker at 37°C for 220rpm. A total of 1 mL of overnight culture was subcultured in a new 10 mL LB broth medium at 37°C for 3 hours at 220rpm until it reached OD_{600} = 0.4-0.6. Bacterial culture was transferred to a new 15 mL falcon tube and incubated at 0°C for 20 min followed by a centrifugation at 4°C for 5 min at 1,000 x g. The supernatant was discarded and the cell pellet was dissolved in 5 mL of RF1 solution (refer to Appendix C(e)).

The mixture was incubated on ice for 20 min and centrifuged at 1,000 x g, 4°C for 10 min. Subsequently, the supernatant was removed and 400 µL of RF2 solution (refer to Appendix C(e)) was added to dissolve the pellet. Finally, the mixture was transformed to competent cells and aliquoted to a new microcentrifuge tube before storage in -80°C.

3.4.3 Ligation

A reaction containing pGEM-T vector (Promega, USA) and purified *BrCHS* fragment (insert DNA) was prepared with the addition of 1X T4 DNA Ligase reaction buffer and 1U of T4 DNA ligase (1 U/µL) (New England Biolabs, UK). To increase the ligation efficiency, the ratio of vector and insert DNA was set to 1:7 and final volume was adjusted to 10 µL. The mixture was incubated overnight at 4°C prior to *E.coli* transformation step.

3.4.4 Transformation into competent cells, *E. coli* strain JM109

Competent cells that were prepared previously (refer Section 3.4.2) were thawed on ice. An aliquot of 5 µL of ligated product was transferred into a new microcentrifuge tube containing 100 µL of competent cells on ice. The mixture was mixed gently and incubated for 30 min. Next, the mixture was heat-shocked at 42°C for 50 sec. After that, the tube was immediately incubated on ice for 5 min. A total of 900 µL of LB broth medium was added to the tube and incubated for 2.5 to 3 hours at 37°C with a shake of 220rpm. For colony screening, the cells were plated on the LB agar (refer to Appendix C(d)) containing 100 µg/mL of ampicillin, 80 µg/mL of X-gal and 0.5 mM of IPTG followed by overnight incubation (16 to 18 hours) at 37°C.

3.4.5 Colony screening of DNA inserts in *E.* coli by PCR analysis

Positives colonies were selected based on blue and white screening selection system. White colonies indicated successful transformation of ligated product while blue colonies represented a transformation without gene of interest. A library of positive colonies were made by transferring individual selected white colonies onto a new LB agar medium containing 100 µg/mL of ampicillin, 80 µg/mL of X-gal and 0.5 mM of IPTG using an inoculate loop. The remaining colonies on the inoculate loop were dipped into 0.5 mL tube containing 100 µL of sdH_2O. The mixture was boiled at 99°C for 10 min and

centrifuged for 5 min prior to colony screening using PCR analysis. Primers used were designed and labelled as M13_F and M13_R as listed in Appendix D.

The PCR analysis was performed in a 25 µL reaction containing 500 ng of DNA, 1X of Buffer A with 1.5 mM of $MgCl_2$, 1.25 mM of dNTPs, 0.4 µM of sense primer and anti-sense primer and 1U of Taq DNA polymerase (EuRx, Poland). PCR conditions were described according to the stage: (a) initial denaturation step: 95°C for 3 min, (b) amplification step for 32 cycles (denaturation: 95°C for 45 sec, annealing: 55°C for 45 sec, extension: 72°C for 1 min and (c) final extension step: 72°C for 5 min. Amplified products were kept at 4°C for storage. Finally, agarose gel electrophoresis was carried out to confirm the present of amplified DNA fragment (refer to Section 3.3.1).

3.4.6 Plasmid DNA extraction and DNA sequencing

The growing colonies on media represent resistant colonies towards ampicillin antibiotic and thus carry the insert. Therefore, selected colony that was positively checked by PCR was inoculated into LB broth medium containing 100 µg/mL of ampicillin and incubated overnight at 37°C, 220rpm. For long term storage, 850 µL overnight culture was mixed with 150 µL sterilized glycerol and kept in -80°C. The balance of bacterial cultures was transferred to 15 mL falcon tubes and centrifuged at 13,000 x g for 5 min. The supernatant was discarded and bacterial cell pellet was resuspended with 200 µL of Solution I (refer to Appendix C(f)) by vortexing.

After the bacterial pellet was fully dissolved, the suspension was transferred to a new microcentrifuge tube. A total of 200 µL of Solution II (refer to Appendix C(f)) was freshly prepared, added and mixed gently at room temperature for 4 min. Next, 200 µL of Solution III (refer to Appendix C(f)) was added, mixed gently and incubated at 0°C for 15 min. The suspension was then centrifuged at 19,000 x g for 10 min. Subsequently, the supernatant was transferred to a new microcentrifuge tube. RNase A (10 mg/mL)

(Thermo Scientific, Lithuania) was added to a final concentration of 0.25 µg/µL and further incubated at 37°C for 3 hours.

Next, 1 volume of phenol was added to the mixture, mixed gently and centrifuged at 19,000 x g for 3 min. The supernatant was transferred to a new microcentrifuge tube. This step was repeated by adding an equal volume of chloroform. One-tenth volume of 5 M of NaCl and 2.5 volumes of isopropanol were added to the supernatant followed by incubation at 0°C for 20 min. The suspension was centrifuged at 19,000 x g for 15 min and later the supernatant was carefully discarded without disrupting the pellet.

Next, 1 mL of 70% (v/v) ethanol was added and centrifuged at 19,000 x g for 5 min. Pellet was dried at 50°C for 10 min and resuspended with 50 µL of sdH_2O. Agarose gel electrophoresis was carried out to confirm the present of plasmid DNA (refer to Section 3.3.1). The yield and purity of plasmid were measured using NanoPhotometer™ Pearl (IMPLEN, Germany). Next, DNA sequencing was carried out using BigDye® Terminator v3.1 Cycle Sequencing Kit and performed by 3130*xl* Genetic Analyzer (Applied Biosystem, USA) according to the manufacturer's protocol.

3.5 Multiple sequence alignment and phylogenetic analysis

There were more than ten transcript variants of BrCHS, amplified from each part of the *B. rotunda* plants and all were aligned based on their amino acid sequences using MEGA 6 software. Phylogenetic tree analysis was also performed using Neighbour-Joining method and was subjected to 1000 times bootstraps replication with *Cucurma longa* (ClPKS9; JN017186.1) was selected as an outgroup. The distances were computed using Tamura 3-parameter method in the units of number of base substitutions per site. The analysis involved 63 amino acid sequences (refer to Appendix E(a)). All ambiguous positions were removed for each sequence pair. There were a total of 1191 positions in

the final dataset. Conserved regions of BrCHS were identified in order to verify the true sequence of CHS.

Meanwhile, amino acid alignment from each *BrCHS*s was aligned with other various monocot plants containing CHS and was conducted using MEGA 6 software. The monocots plants were *Cucurma longa* (ClPKS9; JN017186.1), *Curcuma alismatifolia* (CaCHS; GU140082.1), *Musa acuminata* (MaPKSIII3; GU724609.1), *Oryza sativa* (OsCHS; AB000801.2), *Zea mays* (ZmCHS; NM_001155550.1) and *Medicago sativa* (MsCHS2; L02902.1). Amino acid sequences were subjected to 1000 bootstraps replication with *Arabidopsis thaliana* CHS was selected as an outgroup.

The distances were computed using the JTT matrix-based method. The variation rate among sites was modelled with a gamma distribution (shape parameter = 0.4). The analysis involved 20 amino acid sequences (refer to Appendix E(b)). All ambiguous positions were removed for each sequence pair. There were a total of 394 positions in the final dataset. Additionally, alignment of deduced amino acid sequences among BrCHSs and monocot's CHS were generated to identify the catalytic residues and active site of the gene.

3.6 Southern blot analysis

3.6.1 DNA extraction

Genomic DNA was isolated from fresh leaves of *B. rotunda* plant according to the method described by Khayat et al. (2004). An approximate of 0.5 to 1.0 g of leaves was ground using mortar and pestle with the addition of liquid nitrogen. The ground samples and 500 µL of DNA extraction buffer were incubated together (refer to Appendix C(g)) at 55°C for 30 min. Samples were centrifuged at 2,800 x g for 5 min and the supernatants were transferred to a new microcentrifuge tube. This step was done in order to have more compact interphase. RNase A (Thermo Scientific, Lithuania) with a final concentration

of 0.2 µg/L was added to the samples and incubated at 37°C for 3 hours. Next, equal volume of Phenol:Chloroform:Isoamyl alcohol (PCI; 25:24:1) was added, vortexed and centrifuged at 2,800 x g for 5 min. The aqueous phase was then transferred to a new microcentrifuge tube. The same step was repeated by adding equal volume of Chloroform:Isoamylalcohol (CI; 24:1) and upper aqueous phase was transferred to a new microcentrifuge tube.

Later, the DNA was precipitated by adding 1 volume of ice-cold isopropanol followed by incubation at -80°C for 30 min. Later, the samples were thawed in 4°C and centrifuged at 8,000 x g for 20 min. The pellet was collected and washed with 1 mL of 70% (v/v) ethanol and centrifuged at 2,800 x g for 5 min. Ethanol was discarded and the pellet was air-dried. To dissolve the pellet, 30 µL of sdH_2O was added. Agarose gel electrophoresis was carried out to confirm the success of genomic DNA isolation (refer to Section 3.3.1). The yield and purity of genomic DNA were measured using NanoPhotometer™ Pearl (IMPLEN, Germany).

3.6.2 Probe labelling with DIG

3.6.2.1 Preparation of *CHS* as a probe

The coding region of *CHS* was amplified using specific primers (CHS_F and CHS_R, as listed in Appendix D) targeting its core fragment. PCR was carried out as mentioned in Section 3.4.5. PCR products were purified (as mentioned in section 3.3.2) and was visualized by agarose gel electrophoresis (as mentioned in section 3.3.1). About 10 ng to 3 µg of purified DNA was used for probe labelling. The DNA was denatured by heating in boiling water for 10 min and immediately chilled on ice. Complete denaturation was essential for efficient labelling. A reaction containing 2 µL of 1X Hexanucleotide mix, 2 µL of dNTP labelling mix, 1 µL of Klenow enzyme labelling grade from DIG DNA Labeling and Detection Kit (Roche, USA) and freshly denatured DNA was prepared. The

mixture was incubated at 37°C for 20 hours (overnight) and 2 µL of 0.2 M EDTA (pH 8.0) was added to terminate the reaction.

3.6.2.2 Determination of probe labelling efficiency

An optimal probe concentration is important in hybridization to avoid background effects or weak signals. A series of dilutions of Digoxygenin (DIG)-labeled *BrCHS* DNA was prepared and applied to a small strip of a positively charge Hybond™ N+ nylon membrane (Amersham Pharmacia Biotech, Ireland). A defined solution of DIG-labeled control DNA was preloaded into the nylon membrane as standard. The nucleic acid was fixed to the membrane by heating at 120°C for 30 min. The membrane was then incubated with 20 mL of Maleic acid buffer (0.1 M of Maleic acid, 0.15 M of NaCl; pH 7.5) at 25°C with shaking for 2 min.

Next, the membrane was incubated in 10 mL of 1X Blocking solution (provided in the kit) for 30 min to block the unspecific binding and further incubation in 10 mL of Antibody solution (1:5000 (150 mU/mL) of Anti-Digoxigenin-AP in 1X Blocking solution) for 30 min. Then, the membrane was washed twice with 10mL Washing buffer (0.1 M of maleic acid, 0.15 M of NaCl (pH 7.5), 0.3% (v/v) of Tween 20) for 15 min to remove unbound antibodies. For visualization, the membrane was equilibrated in 10 mL freshly prepared Detection buffer (0.1 M of Tris-HCl, 0.1 M of NaCl, pH 9.5) for 2 to 3 min and subsequently incubated in 2 mL freshly prepared colour substrate solution; NBT (nitro-blue tetrazolium chloride)/BCIP (5-bromo-4-chloro-3'-indolyphosphate p-toluidine salt) (NBT/BCIP) in the dark without shaking. The reaction was terminated by rinsing the membrane with sdH_2O for 5 min. The colour intensity of the DIG-labeled DNA dilution series was compared with control DNA, and the probe DNA was estimated and calculated.

3.7 Southern hybridization

3.7.1 Restriction enzyme digestion

A total of 10 µg genomic DNA was digested using four different restriction enzymes (RE); 1 U/µL *Eco*RI (Promega, USA), 1 U/µL *Eco*RV (Promega, USA), 1 U/µL *Dra*I (Promega, USA) and 1.2 U/µL *Hin*dIII (New England Biolabs, UK). The DNA was incubated with 1X respective enzyme buffer and Bovine Serum Albumin (BSA) (Promega, USA) together with RE and incubated overnight at 37°C, to achieve a complete digestion.

3.7.2 Agarose gel electrophoresis

The completely digested DNA was separated via agarose gel electrophoresis. About 0.7% (w/v) of agarose gel was prepared as mentioned in Section 3.2.1, except for the addition of EtBr in the gel. The DNA was mixed with 1 µL of 6X loading dye (0.25% of bromophenol blue and 0.25% of xylene cyanole FF) and later loaded into the wells. A 1kb DNA ladder (Thermo scientific, Lithuania) was used as a molecular weight marker. The electrophoresis was carried out at voltage of 100V for 1 hour. Later, the gel was stained with 0.5 µg/mL of ethidium bromide (EtBr; Sigma, USA) in sdH_2O for 30 min and destained with sdH_2O for another 30 min. Next, the gel was visualized using Alpha Imager system (Alpha Innotech Corp, USA).

3.7.3 Upward capillary transfer

Before the DNA is transferred to charged nylon membranes, the DNA was firstly denatured by soaking in a denaturing solution which was an alkaline transfer buffer (0.4 M of NaOH with 1 M of NaCl) for 15 min with constant gentle agitation on a belly dancer. The solution was changed and the gel was soaked again for another 20 min. A fresh scalpel was used to cut piece of nylon membrane approximately 1mm larger than the gel size in each dimension. Two sheets of thick blotting paper were also cut to the same size

as the membrane. The membrane was then floated on the surface of a container of deionized H_2O until it became completely wet from beneath. Then, the membrane was immersed in the alkaline transfer buffer for at least 5 min.

While denaturing the DNA, the compartment was filled with the alkaline transfer buffer until the level of the liquid reached almost to the top of the support. When the blotting paper on the top of the support was thoroughly wet, the glass rod was rolled on the paper to remove all air bubbles. Next, the gel was removed from the alkaline buffer solution and the gel was inverted so that its underside will be at the top. The inverted gel was placed on the support and centered on the wet blotting paper. During this step, no air bubbles were allowed between the blotting paper and the gel. To serve as a barrier to prevent liquid from the reservoir to flow to the paper towels which were placed on the top of the gel; the gel was surrounded with parafilm. The top of the gel was then wetted with alkaline transfer buffer.

Next, the two pieces of thick blotting paper were wetted in the alkaline transfer buffer and placed on top of the wet membrane. To remove the air bubbles, a pipette was rolled across the surface of the membrane. A stack of paper towels (5 to 8 cm) high was cut and folded to the size smaller than the blotting papers and placed on the blotting papers. A 400 g weigh was placed on the top of the stack and weighed it down. The DNA was allowed to transfer for overnight.

The entire stack of towels was ensured that they were not wet. On the next day, the paper towels were removed as well as the blotting papers above the gel. The gel and membrane were turned and put on a dry sheet of a blotting paper and the position of the gel slots on the membrane was marked with a pencil. Later, the gel was peeled from the membrane and then it was proceeded with the fixation of the DNA to the membrane.

Figure 3.1 shows upward capillary transfer with reference to Sambrook's protocol (Sambrook et al. 1989).

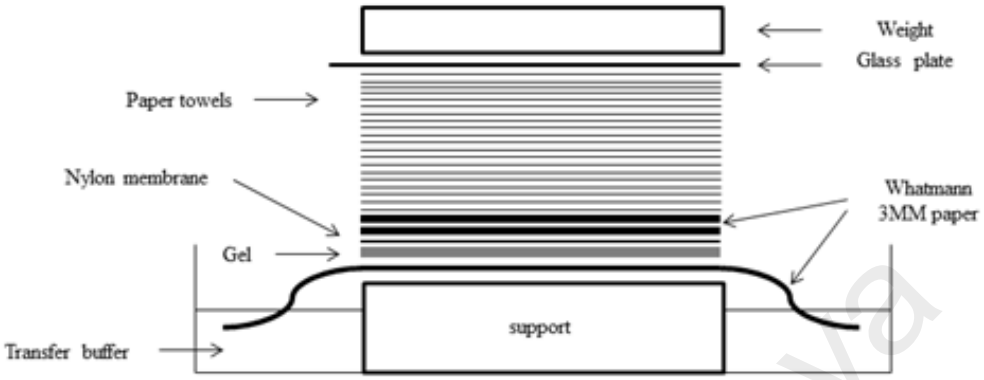


Figure 3.1: Upward capillary transfer of DNA from agarose gel

Buffer drawn from a reservoir passed through the gel into a stack of paper towels with a weight which was applied on the top of the paper towels. DNA transfer and elution from the gel were carried out by the moving of buffer stream that was deposited onto a nylon membrane.

3.7.4 Preparation for hybridization

Since positively charged nylon membrane was used during the DNA transfer, the membrane was further fixed by soaking in neutralization buffer II (0.5 M of Tris-HCl, pH 7.2 and 1 M of NaCl) and proceeding with pre-hybridization. This step will remove any piece of agarose that was stuck on the membrane and neutralize the membrane. Alkaline transfer formed covalent attachment of DNA to positively charged nylon membrane. The solution for hybridization was prepared in formamide buffer. After thorough mixing, the solution was filtered through a 0.45 µm disposable cellulose membrane. Background/noise of the membrane was decreased when hybridizing under condition of reduced stringency. Pure formamide was used to overcome the development of background.

The membrane containing the target DNA was floated on the surface of a tray of 6X of SSC (diluted from 20X of SSC; 3 M of NaCL and 0.3 M of $Na_2Citrate.2H_2O$) until the membrane became thoroughly wet from beneath. The membrane was submerged for 2 min. Later, pre-hybridization was done in a roller bottle. The wet membrane was gently rolled into the shape of cylinder and was placed inside a hybridization roller bottle with the addition of 0.1 mL prehybridization solution (6X of SSC, 5X of Denhardt's reagent, 0.5% (w/v) of SDS, 100 µg/ml of salmon sperm DNA and 50% (v/v) of formamide) for each square centimeter of membrane. Then, the bottle was closed tightly. The hybridization tube was placed inside a pre-warmed hybridization oven at 42°C for 1 to 2 hours.

Since the labelled *CHS* probe was double-stranded DNA, the strands were denatured by heating for 5 min at 100°C and the probe was rapidly chilled in ice water. Next, the *CHS* probe was hybridized to a blot containing genomic DNA by pouring off the pre-hybridization solution and replacing it with fresh hybridization solution (6X of SSC, 5X of Denhardt's reagent, 0.5% (w/v) of SDS, 100 µg/ml of salmon sperm DNA and 50% (v/v) of formamide) containing *CHS* probe. The bottle was sealed and incubated overnight in hybridization oven.

Next, the membrane was removed from the hybridization bottle and the excess hybridization solution was briefly drained out from the membrane. The membrane was washed with several hundred milliliters of 2X of SSC and 0.5% of SDS at room temperature and agitated gently on a slowly rotating platform. After 5 min, the first rinsed off took placed and several hundred milliliters of 2X of SSC and 0.1% of SDS was added again to the membrane. The membrane was incubated for 15 min at room temperature with gentle agitation. Next, the solution was rinsed off and several hundred milliliters of fresh 1X of SSC and 0.1% of SDS were added and incubated overnight at 65°C with

gentle agitation. Lastly, the membrane was washed briefly with 0.1X of SSC at room temperature before being subjected to immunological detection.

3.7.5 Immunological detection

Prior to detection, the membrane was rinsed briefly in Washing buffer for 5 min and further incubated in the 1X of Blocking solution for 30 min with gentle agitation. Next, the membrane was incubated in 20mL of Antibody solution for 30 min and washed twice in Washing buffer for 15 min each. The membrane was then equilibrated with Detection buffer for 5 min and later incubated with 10 mL of freshly prepared colour substrate solution and incubated overnight. Finally, the reaction was stopped with sdH_2O and the membrane was analysed.

3.8 Cloning of *BrCHS2* transcript variant into *E. coli* (pCAMBIA1304 vector system)

3.8.1 Restriction enzymes digestion

By referring to phylogenetic tree, one *BrCHS* transcript variant was chosen for the purpose of transformation in suspension cultures. This isolated transcript variant from rhizome was chosen because it is the most relevant variant with medicinal interest as it has been utilized to be expressed and produced compound that exhibit inhibitory activities such as an antimutagenic effect (Trakoontivakorn et al. 2001), anticancer activity (Kirana et al. 2007) and antioxidant activity (Shindo et al. 2006). Thus, the pGEM-T easy vector harbouring entire coding region of *BrCHS2* transcript variant was digested with restriction endonucleases, *Nco*I (Fermentas, USA) and *Spe*I (Fermentas, USA). The digested plasmid DNA was incubated at 37°C for overnight (16 to 18 hours) to obtain complete digestion. The complete digestion of plasmid DNA was confirmed using agarose gel electrophoresis (refer to Section 3.3.1) and gel extracted to obtain the specific *BrCHS2* fragment (refer to Section 3.3.2).

3.8.2 Ligation

The pCAMBIA1304 expression vector (GenBank accession no. AF234300.1) (refer to Appendix F) was used for ligation. A reaction consists of 1 µL of T4 DNA ligase (New England Biolabs, UK) and 1X of ligase reaction buffer, further added into a mixture containing the vector and digested *BrCHS2* fragment in a microcentrifuge tube. To increase the ligation efficiency, the ratio of vector and DNA fragment was set to 1:7. Later, for complete ligation, the mixture was incubated overnight at 4°C. The purified *BrCHS2* transcript variant was cloned in position between *Nco*I and *Spe*I restriction sites in pCAMBIA1304 vector. The expression vector pCAMBIA1304 contained T-DNA region that carried CAMV 35S promoter, *hptII* encoding hygromycin phosphotransferase and a fusion gene encoding mGFP5-GUSA protein.

Next, LB agar was prepared by adding 50mg/mL of Kanamycin, poured into petri dishes and stored at 4°C (refer Section 3.4.1). At the same time, the competent cells, *E. coli* strain TOP 10 were prepared prior to transformation of recombinant gene (refer Section 3.4.2). The ligated product was then transformed with competent cells, TOP 10 strain (refer to Section 3.4.4) and incubated overnight at 37°C.

3.8.3 Screening of recombinant *BrCHS* in *E. coli* by PCR analysis

Positive colonies carrying the insert recombinant gene were selected to make a library by transferring each of the white colonies onto a new gridded LB agar medium containing 50 mg/mL of Kanamycin (refer to Section 3.4.5) and screened the positive clones using PCR analysis. Primers used were 35S_F and CHS_R as listed in the Appendix D. The presence of amplified DNA was determined using agarose gel electrophoresis (refer to Section 3.3.1).

3.8.4 Plasmid DNA extraction

Positive single colonies that carried the *BrCHS2* transcript variant fused with the expression vector; pCAMBIA1304, were cultured in LB broth medium containing 50mg/mL of Kanamycin and followed by the plasmid DNA extraction steps (refer to Section 3.4.6). Agarose gel electrophoresis was carried out to confirm the present plasmid DNA (refer to Section 3.3.1). The plasmid DNA yield and purity were measured using NanoPhotometer™ Pearl (IMPLEN, Germany).

3.8.5 DNA sequencing and multiple sequence alignment analysis

Sequence analysis was carried out using DNA sequencing on selected clones obtained from PCR analysis (refer to Section 3.4.6). Primers used for DNA sequencing were 35S_F, CHS_R1, CHS_F1 and pCAMBIA_R as listed in the Appendix D. The nucleotide sequences were analyzed and translated into amino acid sequences using MEGA 6 program. Alignment of deduced amino acid sequences was also generated to confirm the orientation of the recombinant gene.

3.9 *Agrobacterium* mediated transformation method

3.9.1 Preparation of Luria-Bertani (LB) broth and agar media

LB broth and agar media were prepared by adding antibiotics; 30 mg/L of Rifampicin and 50 mg/L of Kanamycin (refer to Appendix C(d) and Section 3.4.1).

3.9.2 Preparation of *Agrobacterium* competent cells

Agrobacterium tumefaciens (*A. tumefaciens*) strain 4404 was used in this study. This strain was selected from glycerol stock and streaked on LB agar containing 30 mg/L of Rifampicin and incubated at 28°C for two days. After two days, single colony was selected to be inoculated in 10 mL LB broth medium containing 50 mg/L of Rifampicin. The colony was incubated overnight at 28°C with shaking at 180rpm. The following day, 1mL of the overnight culture was transferred into 10 mL LB broth medium and incubated

for 3 hours to obtain OD reading about 0.5-0.6. The culture was centrifuged at 6000 x g for 5 min and the supernatant was discarded. Pellet was resuspended in 5 mL storage buffer (0.2% of glucose, 12 mM of $MgSO_4$, 12% of PEG and 36% of glycerol). These *Agrobacterium* competent cells were kept at -80°C for long-term storage.

3.9.3 Plasmid transformation into *Agrobacterium*

The pCAMBIA1304 carrying *BrCHS2* transcript variant was added to *Agrobacterium* competent cells and kept on ice for 30 min. The mixture was kept frozen in liquid nitrogen for 30 sec to 1 min and immediately after, it was heat-shocked at 37°C for 4 min. Subsequently, the cells were cooled down to 0°C and 1 mL of LB broth medium was added. Later, the mixture was incubated at 28°C for 2 to 4 hours. Next, the mixture was centrifuged at 16,000 x g for 2 min and supernatant was discarded. Finally, the pellet was resuspended in 100 µL LB broth medium prior to culture on LB agar containing 30 mg/L of Rifampicin and 50 mg/L of Kanamycin. The cells were streaked on LB agar and incubated at 28°C for 2 days.

3.9.4 Selection of positive transformants by PCR analysis

Positive colonies carrying the insert gene were selected to make a library by transferring each of the white colony onto a new grided LB agar containing 30 mg/L of Rifampicin and 50 mg/L of Kanamycin (refer to Section 3.4.5) and the positive clones were screened using PCR analysis. The presence of amplified DNA in *Agrobacterium* colonies was screened (refer to Section 3.3.1) using vector-gene primers; 35S_F and CHS_R as listed in the Appendix D.

3.9.5 Plant infection, co-cultivation and selection

B. rotunda suspension cultures were used as the target explant in this project. Suspension cultures were prepared prior to *Agrobacterium* transformation (refer to Section 3.1). Transformed *Agrobacterium* carrying recombinant construct was cultured

overnight for 16 to 18 hours in LB broth medium containing 30mg/mL of Rifampicin and 50 mg/L of Kanamycin. On the next day, *Agrobacterium* was re-cultured for another 3 hours until OD600nm: 0.5-0.6 was reached. The bacteria were collected by centrifugation at 2,800 x g for 5 min, washed with LB broth medium and MS liquid medium. Next, the pellet was resuspended in 10 mL MS medium for the inoculation.

Approximately, 2 mL of settled cell volume (SCV) of suspension cultures were briefly air-dried by sucking out the liquid media and left in laminar flow cabinet for 10 to 15 min. Cells were infected by bacteria suspensions at room temperature for 10 min without shaking. During the infection, 100 µM of Acetosyringone (AS) was added to the MS liquid medium. After the infection, the cells were blotted-dry on filter paper in a petri dish containing solid MS medium supplemented with 1 mg/L of 2,4-D, 0.5mg/L of BAP and 100 µM of AS. The cells were then co-cultivated for 3 days in dark condition.

Next, the infected cells were cultured in liquid MS medium supplemented with 1 mg/L of 2,4-D, 0.5mg/L of BAP and 300 mg/L of Cefotaxime (Duchefa Biochemie, Netherlands) on a rotary shaker (100rpm) at 25±1°C for 1 week. The cells were then grown on semi-solid MS medium supplemented with 1.0 mg/L of 2,4-D , 0.5 mg/L of BAP, 300 mg/L of Cefotaxime and 15 mg/L of Hygromycin (Duchefa Biochemie, Netherlands) for selection of resistant callus in 4 weeks intervals, respectively.

3.10 Verification on transgenic lines

3.10.1 Callus morphology and GUS assay on resistant callus

Callus morphology of three resistant callus was observed under Stereo Microscope Leica M205 A (Leica Microsystem Ltd, Germany). The colour as well as the morphology of callus including wild-type were observed. The magnification used was 10X. Prior to GUS assay, histochemical staining reagent was prepared. Resistant callus was stained in histochemical reagent (refer to Appendix C(h)). Explants were then incubated at 37ºC in

darkness, overnight until blue coloration appeared. Stained samples were washed in 70% (v/v) ethanol and fixed in Formalin/Acetic/Alcohol (FAA) solution as described in the method by Jefferson et al. (1987) (refer to Appendix C(i)). Based on the blue coloration, ten resistant callus that showed positive blue colour were selected and transferred to liquid MS media. Subculture of suspension culture was then carried out every 10 days up to two months to acquire sufficient amount of cell cultures. After two months of subculture, suspension cultures were harvested and 100 µL of SCV of cell cultures were subjected to GUS assay.

3.10.2 Multiplex PCR analysis

After two months of subculture, ten transgenic suspension cultures (including wild-type) were harvested, extracted and subjected to PCR analysis. Genomic DNA was isolated following a method described by (Khayat et al. 2004) (refer to Section 3.6.1) prior to multiplex PCR analysis using primers labelled as 35S_F, CHS_R, HptII_F, HptII_R, EF1α_F and EF1α_R as listed in the Appendix D. The PCR reaction was then carried out (refer to Section 3.4.5). Later, the presence of amplified DNA was determined using agarose gel electrophoresis (refer to Section 3.3.1).

3.11 Quantification of *BrCHS* expression level

3.11.1 Primer efficiency testing

Different primer concentrations ranged from 50 nM to 100 nM were firstly tested for the identification of primer combination. Four different primer concentrations were evaluated and the best primer combination was chosen for quantifying the expression level of *BrCHS* as well as *EF1α,* a housekeeping gene in ten transgenic suspension cultures later. The selected primer combination indicated abundant of target nucleic acid successfully detected in real-time and have a strong reaction at the earlier cycle to detect

a real signal from sample (exceed threshold level).The primers tested were CHS_rt_F4, CHS_rt_R4, EF1α_F and EF1α_R as listed in the Appendix D.

3.11.2 RT-qPCR analysis on ten transgenic suspension cultures

After two months of subculture, ten transgenic suspension cultures were harvested, extracted and RT-qPCR was carried out to detect the expression level of *BrCHS*. The expression was detected using *Power* SYBR® Green RNA-to-C_T ™ *1-Step* Kit Protocol (Applied Biosystem, USA). RT-PCR reaction mix was prepared by mixing the Power SYBR® Green RT-PCR Mix (2X) and the RT Enzyme mix (125X). Then, the RT Enzyme Mix tube was centrifuged briefly. The volume for each component was measured and prepared in a final volume of 20 µL. Forward and reverse primers with optimized concentration were added to the mixture as well as 100 ng of RNA template. The solution was mixed and the tube was centrifuged briefly to span down the contents and air bubbles were eliminated from the solution.

Next, the PCR reaction plate was prepared by transferring 20 µL of the reaction mixture to each well of the plate. The plate was sealed and centrifuged briefly and ran according to the thermal cycling conditions: (a) reverse transcription step at 48°C for 30 min, (b) enzyme activation at 95°C for 10 min, (c) 40 cycles of denaturation at 95°C for 15 sec and (d) annealing and extension at 60°C for 1 min. The dissociation curve was established at the end of PCR cycle at 95°C for 15 sec, 60°C for 1 min followed by 95°C for 15 sec.

Elongation factor 1α (*EF1a*) gene was used as an endogenous control and three technical replicates were included in this experiment. The primers used were CHS_rt_F4, CHS_rt_R4, EF1α_F and EF1α_R as listed in Appendix D. The fold change of expression was calculated using the comparative C_T method ($2^{-\Delta\Delta Ct}$), where $\Delta\Delta Ct$ = (Ct(target,

transgenic – Ct(target, transgenic)) - (Ct(target, wild-type) – Ct(ref, wild type)). RT-qPCR was performed on a QuantStudio™ 12K Flex system (Applied Biosystem, USA).

3.11.3 Statistical analysis

Experimental data for RT-qPCR were subjected to the analysis of one-way ANOVA using IBM SPSS Statistics 21.0. The significant differences were tested with Tukey's Multiple Range Test.

3.12 CHS enzymatic assay

3.12.1 Protein extraction

Total protein was extracted from transgenic and wild-type suspension cultures, respectively as described in Carpentier et al. (2005) with minor modification. Suspension cultures were dried and weighed about 250 µg. These cells were ground with the addition of liquid nitrogen and transferred to a tube containing protein extraction buffer (refer to Appendix C(j)). The cells were vortexed for 30 sec and equal volume of phenol was added immediately. The mixture was vortexed for 10 min at 4°C and centrifuged for 3 min, 7,000 x g at 4°C. Samples were re-extracted by adding extraction buffer and briefly vortexed. Next, the samples were centrifuged again for 3 min, 7,000 x g at 4°C. The phenolic phase (supernatant) was then transferred into a new tube and protein was precipitated overnight with the addition of 5 volumes of 100 mM of Ammonium acetate in methanol.

On the next day, the samples were centrifuged for 1 hour, 19,000 x g at 4°C. Supernatant was discarded and 2 mL of rinsing solution was added to the pellet. Next, the samples were centrifuged for 30 min, 19,000 x g at 4°C to obtain the pellet. Finally, the pellet was dissolved in 50 mM of Tris-HCl buffer, pH 7.6 containing 1 M of urea. Standards for Bradford assay were prepared from 0.5 mg/mL of bovine serum albumin (BSA) with different concentrations (refer to Appendix C(k)). These standards were used

as references to estimate the concentration of total protein for each transgenic suspension cultures.

3.12.2 CHS enzymatic assay

To test the activity of CHS enzyme, a total of 100 µg protein extract was assayed using 0.8 mM of Malonyl-CoA (Sigma-Aldrich, USA) and 0.4 mM of Cinnamoyl-CoA (MicroCombiChem, Germany) in a final volume of 500 µL of 100mM of potassium phosphate buffer (pH 7.2) containing 0.1% Triton X-100. After incubation at 30°C for 2 hours, the reaction was stopped by acidification with 7.5 µL of 1 N of HCl.

3.12.3 High Performance Liquid Chromatography (HPLC) analysis

The products produced by the reaction were partitioned with 800 µL of ethyl acetate (EA) and concentrated by nitrogen gas. The residues were dissolved in 100 µL of methanol and separated by reversed-phase HPLC system equipped with Waters 1525 Binary HPLC pump and Waters 2998 photo diode array detector (Waters, USA) on a Kinetex RP C_{18} (150mm x 4.6mm) column (Phenomenex, USA). Two solvent systems were used for elution; solvent A (methanol); CH_3OH containing 0.01% of phosphoric acid; H_3PO_4 and solvent B (water); H_2O containing 0.01% of H_3PO_4. The gradient elution profile was as followed: an isocratic step of 50% of solvent A for 1 min, a linear gradient from 50% to 100% of solvent A for 10 min, followed by an isocratic step at 100% of solvent A for 10 min with a flow rate of 0.8ml/min. Peak detection was monitored at 290nm.

In this analysis, six standards were applied to check the present of respective compounds. Standards such as Pinostrobin, Pinostrobin Chalcone, Pinocembrin and Pinocembrin Chalcone (provided by Dr. Lee Yean Kee, Chemistry Department, UM), Naringenin (Sigma, St. Louis, MO) and Naringenin Chalcone (MicroCombiChem, Germany) were applied to the HPLC machine. The retention time for each standard was

recorded and used as reference for compound produced after enzymatic assay in transgenic suspension cultures.

3.12.4 Statistical analysis

Experimental data for HPLC analysis were subjected to one-way ANOVA statistical analysis using IBM SPSS Statistics 21.0. The significant differences were tested with Tukey's Multiple Range Test.

3.13 RNA-Seq analysis

3.13.1 Preparation of transgenic suspension cultures for RNA-Seq analysis

In order to have accurate measurement on the genes expressed at low level, to increase number of differentially expressed genes and to decrease experimental error during RNA-Seq analysis, three biological replicates of wild-type that serve as negative control and three biological replicates of transgenic suspension cultures were propagated in MS liquid medium (1962) supplemented with 0.5 mg/L of BAP and 1 mg/L of 2,4-D in a 250 mL conical flask. The suspension cultures were incubated at 25 ± 2°C with shaking at 100rpm, under a 16 hours photoperiod with a light intensity of 31.4 μmol/m^2/s provided by the cool fluorescent lamp in the plant growth room. Subcultures were carried out every 10 days up to two months to attain the sufficient number of suspension cultures.

3.13.2 Preparation for RNA extraction, cDNA library, and RNA-Seq

All six suspension cultures were harvested after two months, dried and rapidly frozen in liquid nitrogen and stored at -80°C. The samples were transported to Beijing Genome Institute (BGI), Hong Kong, China for RNA extraction, cDNA library preparation and RNA-Seq analysis, following manufacturer's instructions. RNA quantity was determined with a Qubit Fluorometer and RNA integrity was confirmed using an Agilent 2100 Bioanalyzer (Agilent, Waldbronn, Germany).

RNA concentration with more than 150 ng/µL with optical density reading OD260/280: 1.8-2.21, rRNA ratio (28S/18S)≥1.9-2.4, and RNA Integrity Number (RIN)≥7.5 were used for cDNA library preparation (refer to Appendix J). Next, total RNA was converted into a library of template molecules suitable for high-throughput DNA sequencing for subsequent cluster generation. The protocol involved by purifying the poly-A containing mRNA molecule using poly-T oligo-attached magnetic beads.

After purification step, the mRNA was fragmented into small pieces using divalent cations under elevated temperature. Then, the cleaved RNA fragments were copied into first strand cDNA using reverse transcriptase and random primers and further synthesized second strand cDNA. These cDNA fragments had gone through process of an end repair, which single 'A' base was added, and the fragments were then ligated with adapters. Next, these products were purified and enriched with PCR to produce the final cDNA library. Finally, high-throughput sequencing of six libraries clones with 200bp short-insert library was carried out on the platform of Illumina HiSeq™ 2000 (Illumina Inc., USA).

3.13.3 Processing and evaluation of sequence data

Sequencing data output from Illumina HiSeq™ 2000 (Illumina Inc., San Diego, CA, USA) was transferred into a base calling using raw data or raw reads and stored in fastq format. The poor quality raw reads generated were filtered to obtain high-quality read data. All raw reads with adaptor contamination, ambiguous sequence "N" bases, unknown nucleotides with more than 5% bases and low quality reads where the rate of reads which the quality value ≤ 10 were more than 20% were removed and discarded. The output of the sequencing data, including total raw reads, total clean reads, Q20 percentage, N percentage and GC percentage were statistically analysed and evaluated. The remaining

sequence fragments called clean data were then used for subsequent analysis in the RNA-Seq module.

3.13.4 Transcript assembly by Trinity and annotation

Transcriptome of *de novo* was carried out by assembling the clean data into contigs by short reads assembly program called Trinity (http://trinityrnaseq.sourceforge.net/). Trinity combines three independent software modules: Inchworm, Chrysalis, and Butterfly, applied sequentially to process large volumes of RNA-Seq reads (Grabherr et al. 2013)The resulting sequence of the trinity was called unigenes. Unigenes from each assembly of the sample were used for sequence splicing and removing redundancy within sequence clustering software to acquire long non-redundant unigenes.

Next, clustering of the gene family was applied and further divided unigenes into two classes. The first class called clusters (CL), and the cluster ID was put behind wherein each cluster; there were several unigenes that have more than 70% similarity between them. Meanwhile, the other class was called singletons, which known as unigene. For the final step, Blastx alignment was annotated (e-value cutoff of <0.00001) between the unigenes generated by integrating all the transcripts from transgenic and wild-type samples against public protein databases such as Nr, Swiss-Prot, KEGG and COG. With the results of the best-aligned sequences, the sequence directions of the unigenes were then determined. However, for unaligned unigenes, sequence orientations as well as the coding regions were predicted using ESTscan software.

3.13.5 Unigene functional annotation

Functional annotation of the assembled unigenes was conducted and aligned against public protein databases such as non-redundant (Nr), Swiss-Prot, Kyoto Encyclopedia of Genes and Genomes (KEGG), Gene Ontology (GO) and Cluster of Orthologous Groups (COG) by performing a BlastX homology search with a cut-off e-value <$1.00E^{-05}$ as well

as a nucleotide (Nt) database by BlastN with a cut-off e-value <$1.00E^{-05}$. To predict possible functions of the unigenes, pathway annotation can be developed after getting KEGG annotation. From KEGG database, biological complex of all genes can be studied. Next, Blast2GO program (Conesa et al. 2005) was used to acquire Gene Ontology (GO) annotation of unigenes and further classify unigenes for GO functional classification using WEGO software (Ye et al. 2006). The unigenes were classified under three GO-terms, namely molecular function, cellular component, and biological process.

3.14 Differentially expressed (DE) genes analysis

The calculation of unigenes expression was determined by using Fragments Per kb per Million reads (FPKM) method. The FPKM method was chosen to eliminate the influence of different gene length and sequencing level while calculating the gene expression from both samples. Therefore the calculated gene expression can be directly used for comparing the difference of gene expression between the samples.

In order to identify significantly expressed genes between control and transgenic suspension cultures, analysis of differentially expressed genes was done by employing Poisson distribution calculations (Audic & Claverie, 1997). In this analysis, ratio of two-fold increase ($|\log2Ratio| \geq 1$) and having a lower false discovery rate (FDR), FDR ≤ 0.001 was used in all samples for the whole experiment. The smaller the FDR, the greater the ratio and the larger the difference was observed on the expression level between the two samples. DE genes were then carried out into KEGG Pathway analysis.

3.14.1 KEGG pathway analysis on DE genes

The unigenes expression profiles between three replicate experiments of transgenic and wild-type had resulted in formation of three main groups; Group 1: WT-R1 vs TL-R1, Group 2: WT-R2 vs TL-R2 and Group 3: WT-R3 vs TL-R3. Since overexpression of *BrCHS*s could give effects on gene expression in flavonoid and related plant pathways;

therefore, differentially expressed (DE) genes between three groups were filtered according to several parameters against KEGG databases. The parameters were set with all DE genes conforming to be a) present in three groups with b) a fold change, $\log_2$ ratio of $\geq +1$ and ≤ -1, c) FDR of ≤ 0.001 and have d) significance, p-values of $p \leq 0.01$. These parameters were used as thresholds to obtain the significance up and down-regulated unigenes between both samples in the three groups. Flavonoid and related plant pathways with significantly enriched DE genes were investigated after comparing with the whole transcriptome background.

3.14.2 Validation analysis

To experimentally validate the transcriptional abundance results from sequencing and computational analysis, 5 unigenes were selected for qPCR analysis in all three groups. The unigenes include two random up-regulated unigenes (Unigene 28275_All and CL4317.Contig3_All) and three random down-regulated unigenes (CL1414.Contig2_All, Unigene 5234_All and CL12450.Contig1_All). Primers that were used for the validation were shown in Appendix I. Elongation factor, *EF1α* was used as a reference gene.

RT-qPCR was performed on QuantStudio™ 12K Flex realtime PCR platform (Applied Biosystem, USA) using *Power* SYBR® Green RNA-to-C_T ™ *1-Step* (Applied Biosystem, USA) to detect transcript abundance. The amplification was conducted according to the thermal cycling conditions: reverse transcription step at 48°C for 30 min, enzyme activation at 95°C for 10 min, followed by 40 cycles of denaturation at 95°C for 15 sec, finally annealing and extension at 60°C for 1 min. The dissociation curve was established at the end of PCR cycle at 95°C for 15 sec, 60°C for 1 min followed by 95°C for 15 sec. The relative expression levels of the selected unigenes normalized to elongation factor was calculated using $2^{-\Delta\Delta Ct}$ method. All reactions were performed in three replicates.

CHAPTER 4: RESULTS

The results are divided into two main parts: 1) Isolation and characterization of *BrCHS* and 2) Overexpression of *35Sp::BrCHS2* in suspension cultures and its effect on flavonoid biosynthesis and related plant pathways. For the first part, isolation and characterization of *BrCHS* were achieved by extraction of total RNA followed by cloning and sequencing experiment.

4.1 Total RNA extraction

Total RNA was successfully extracted from four parts of *B. rotunda* plants; leaves (L), rhizomes (R), flowers (F) and suspension cultures (SC). The presence of clear 28S rRNA (size: about 3.6 to 3.7kb) and 18S rRNA (size: about 1.9kb) indicated intact RNA was successfully isolated as shown in Figure 4.1. The band below 18S rRNA represents 16S rRNA (size: about 1.5kb). The smearing of RNA was appeared even though flash frozen in liquid nitrogen was applied immediately. The samples might be disrupted by the endogenous RNAse released internally or during loading the RNA samples into a gel. Extra precaution can be applied once the tissue or sample is ready for extraction. The sample should be homogenized as quickly as possible and performed with sufficient amount of liquid nitrogen.

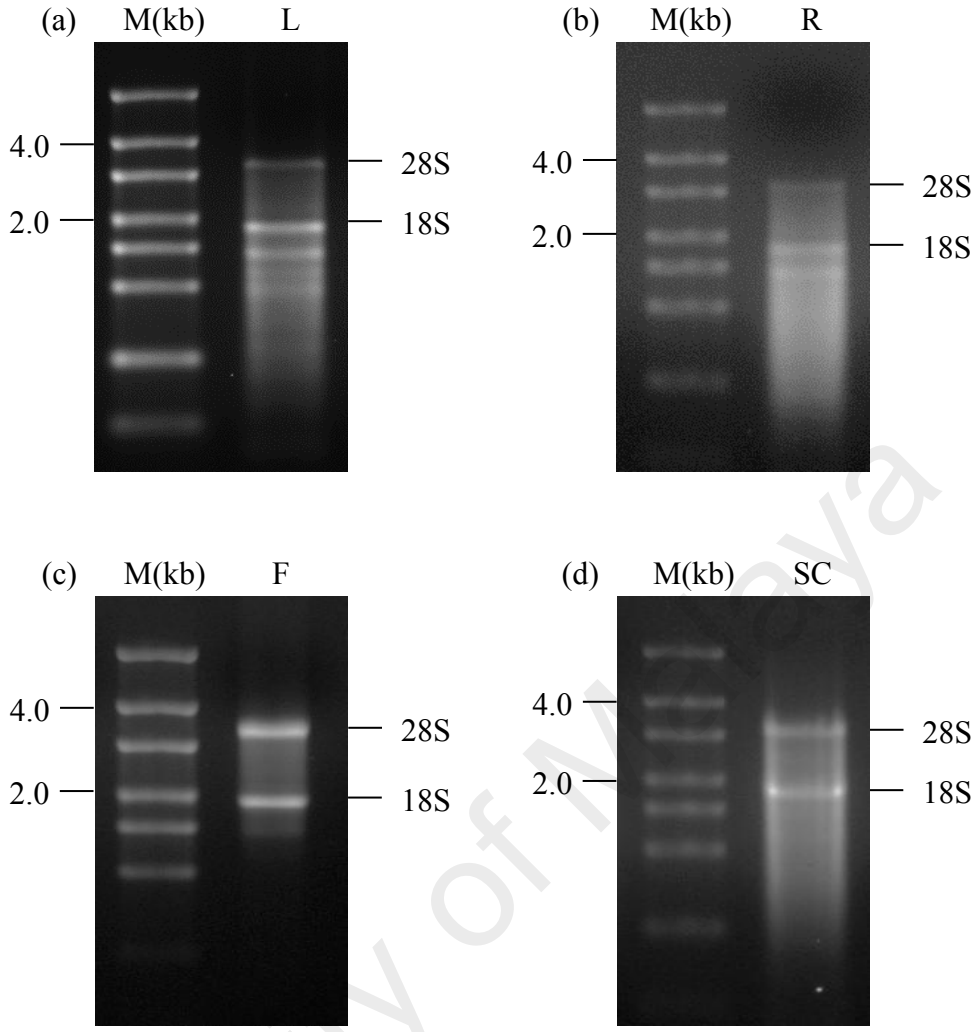


Figure 4.1: Total RNA extraction

Total RNA was successfully extracted from (a) leaves; L (b) rhizomes; R (c) flowers; F and (d) suspension cultures; SC. Lane M: RNA Ladder. The integrity of total RNA was quantified with NanoPhotometer™ Pearl (IMPLEN, Germany) at 230, 260, 280nm and subjected to RT-PCR analysis.

4.2 Reverse transcription-PCR (RT-PCR) analysis and gene purification

Chalcone synthase (CHS) is the key enzyme that is important for the production of secondary metabolites in flavonoid biosynthesis pathway. By using RT-PCR analysis, cDNA of *BrCHS* was successfully amplified by CHS_F and CHS_R primers. A fragment of 1.2kb was obtained from each of the RNA sources (Figure 4.2).

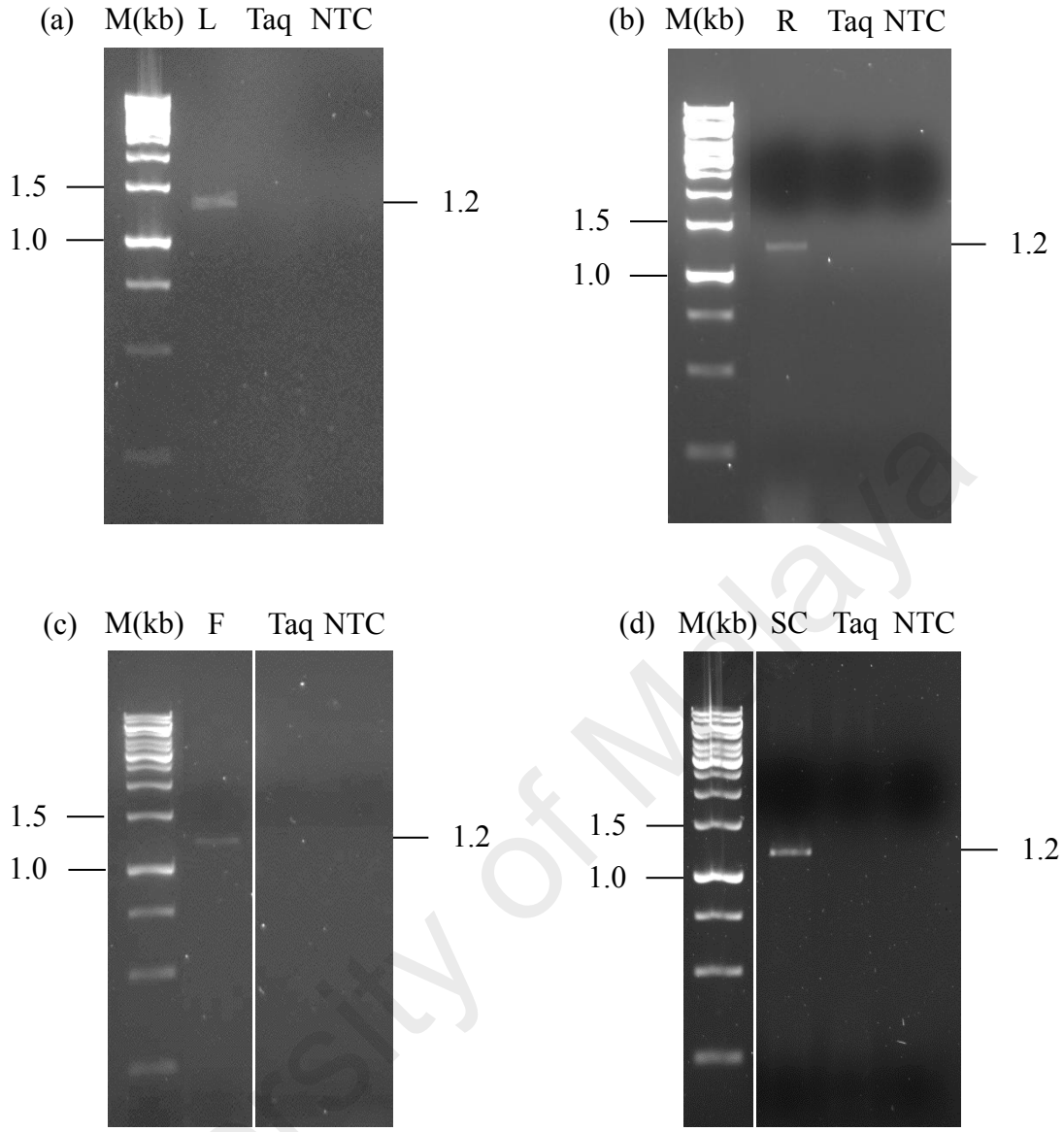


Figure 4.2: Amplified *BrCHS* cDNA product

The cDNA of *BrCHS* was successfully amplified from (a) leaves, L (b) rhizomes, R (c) flowers, F and (d) suspension cultures, SC. Lane M: 1kb DNA Ladder, Taq: amplification with Taq: Taq polymerase (genomic control) and NTC: Non-Template Control.

For cloning and subsequent use, the cDNA of *BrCHS*s were further purified by gel extraction and purification. About 1.2kb fragment was successfully purified (Figure 4.3). An extra band was observed after purification which indicates denatured (single stranded) DNA. Under certain conditions, chaotropic reagents can denature DNA fragments.

Therefore, to prevent this, the tube was allowed to cool slowly at room temperature before downstream application.

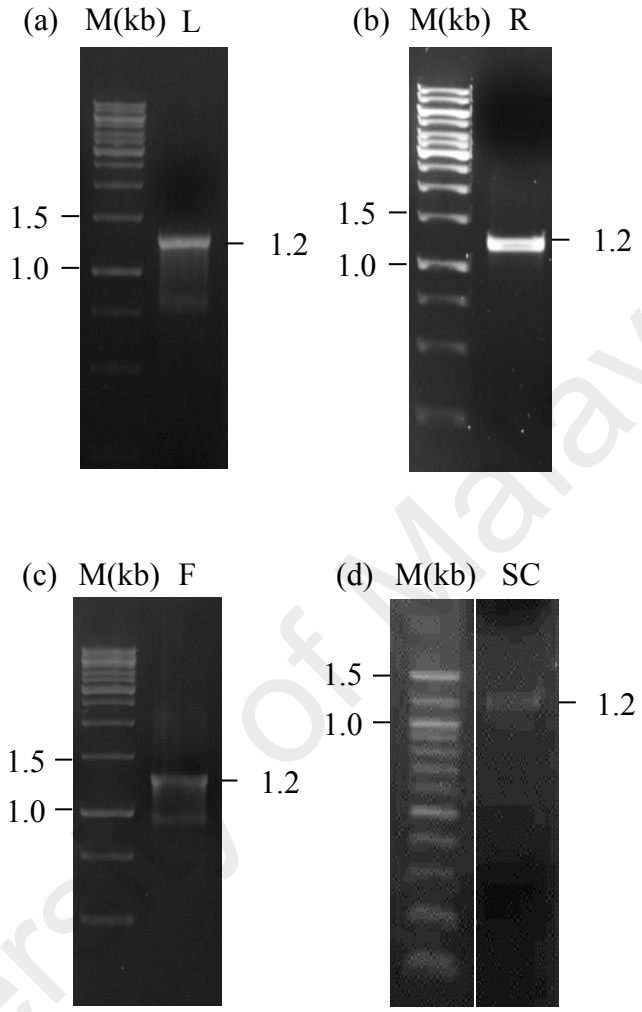


Figure 4.3: Purified cDNA of *BrCHS*

The cDNA of *BrCHS* was successfully purified from (a) leaves, L (b) rhizomes, R (c) flowers, F and (d) suspension cultures, SC. Lane M: 1kb DNA Ladder for (a), (b) and (c), 100bp DNA Ladder for (d).

4.3 Cloning of *BrCHS*s into *E.coli* (pGEM-T easy vector system)

Prior to cloning purposes, each of the purified *BrCHS* was cloned into *E.coli*. Colonies that resisted with ampicillin antibiotic were selected and amplified to obtain DNA insert (1.2kb) as shown in Figure 4.4. These colonies were cultured overnight and plasmid

DNAs extracted with desired *BrCHS* successfully shows high molecular weight of DNA (Figure 4.5).

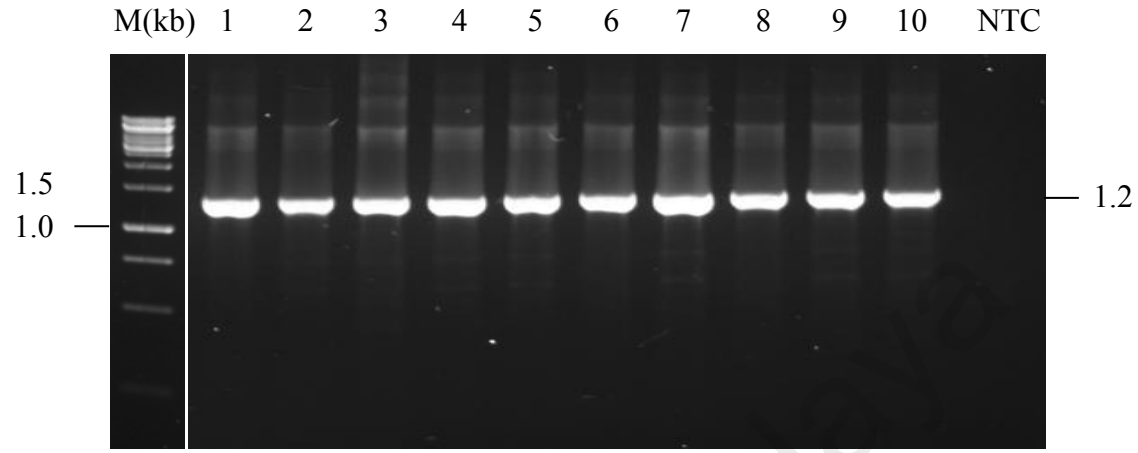


Figure 4.4: Positive colonies showing the inserted DNA

BrCHS was successfully amplified from each colony by using M13_F and M13_R primers. Lane M: 1kb DNA Ladder. Lanes 1-10: Positive colonies with DNA insert. NTC: Non-Template Control.

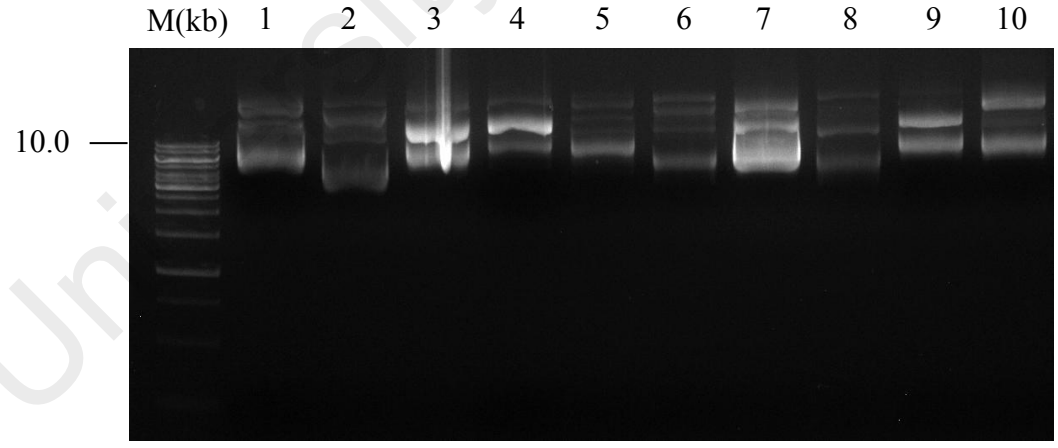


Figure 4.5: Plasmid DNA extracted with desired *BrCHS*

Colonies harboring desired *BrCHS* were resisted with ampicillin antibiotic and cultured overnight to further isolate plasmid DNA. Lane M: 1kb DNA Ladder. Lanes 1-10: High molecular weight of plasmid DNA.

4.4 DNA Sequencing analysis

Each of the cloned *BrCHS* was selected for DNA sequencing to verify the sequences. A total of 62 clones were sequenced. These sequences were then aligned for comparison. It was noted that the variability between the clones are the same. These variability can be preliminary grouped into 5 groups, namely *BrCHS1*, *BrCHS2*, *BrCHS3*, *BrCHS4* and *BrCHS5* transcript variant. This aligned sequence grouping is given in Figure 4.6.

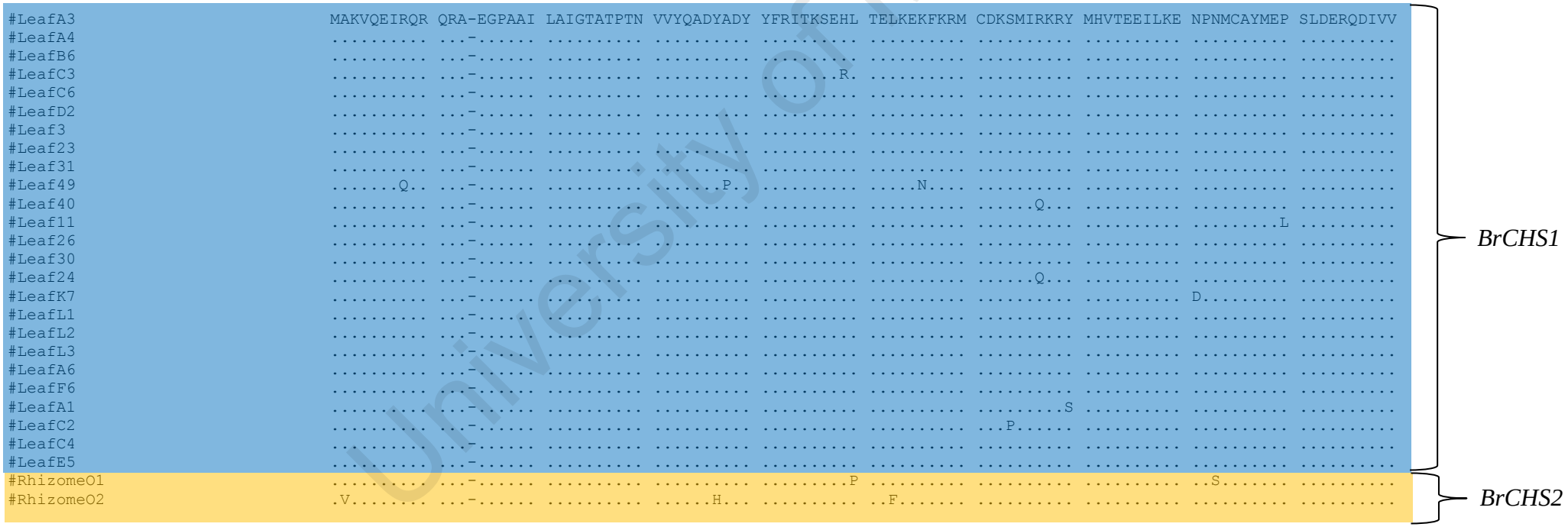


Figure 4.6: Multiple sequence alignment of *BrCHS* transcript variants from different parts of *B. rotunda* plant

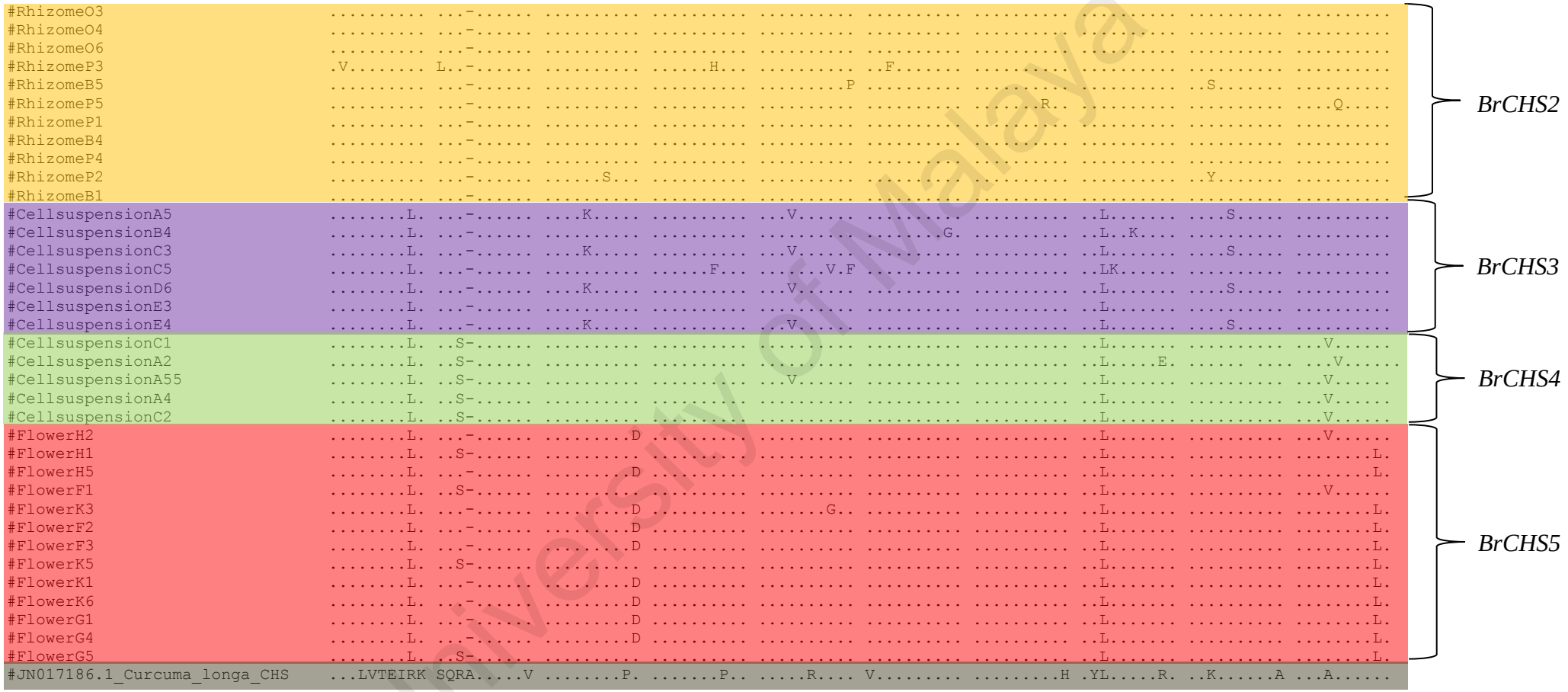


Figure 4.6, continued.

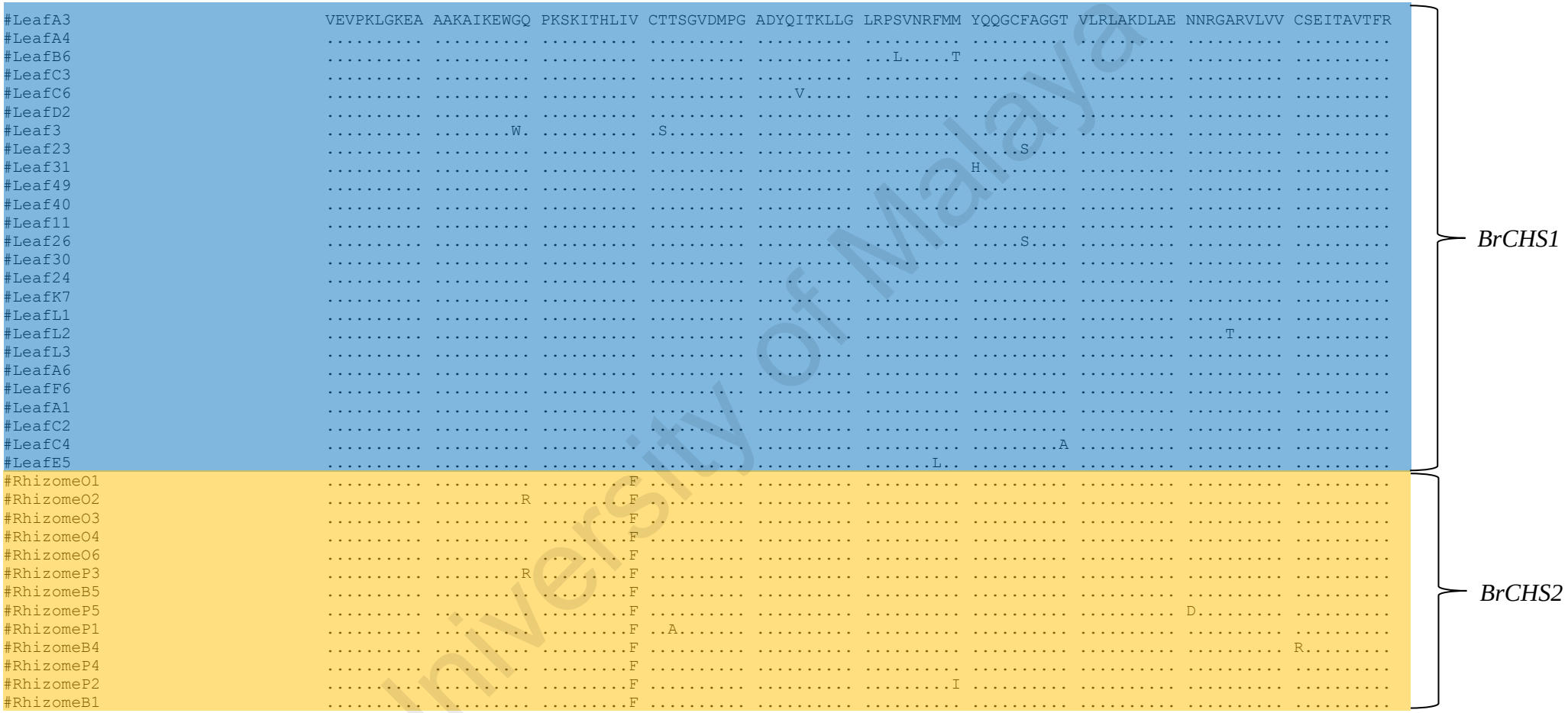


Figure 4.6, continued.

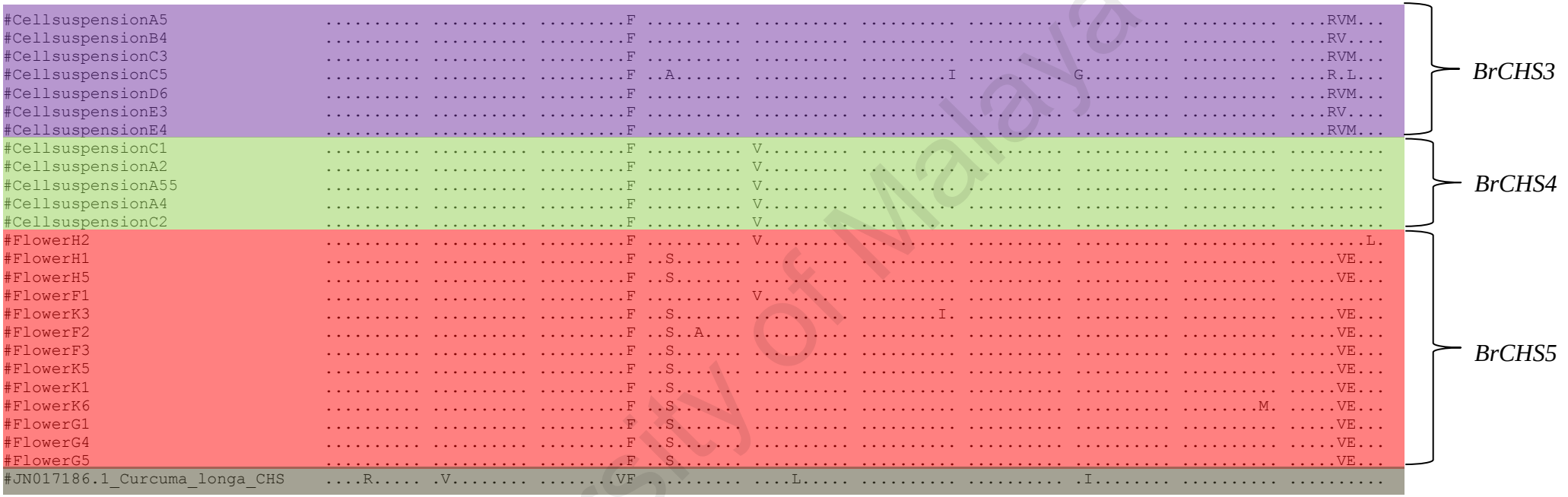


Figure 4.6, continued.

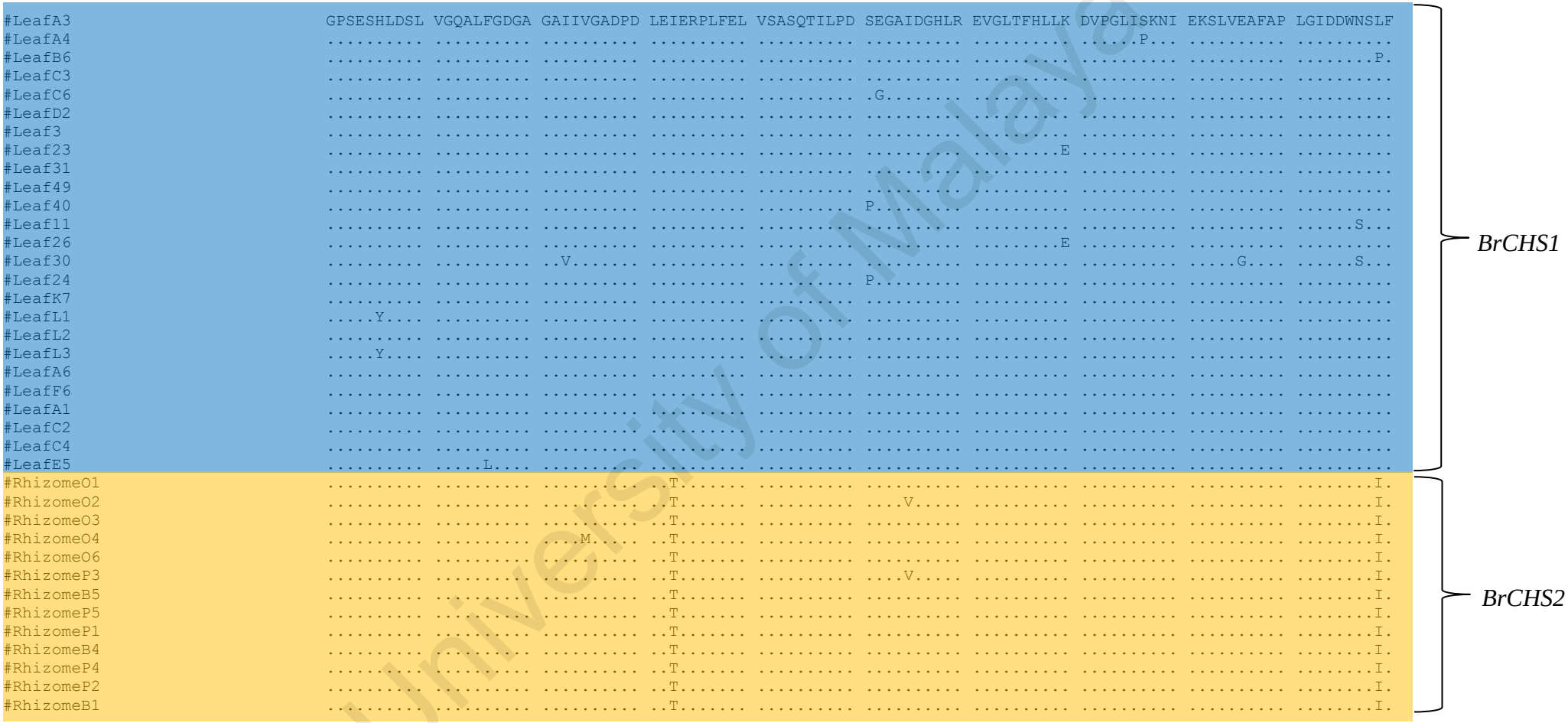


Figure 4.6, continued.

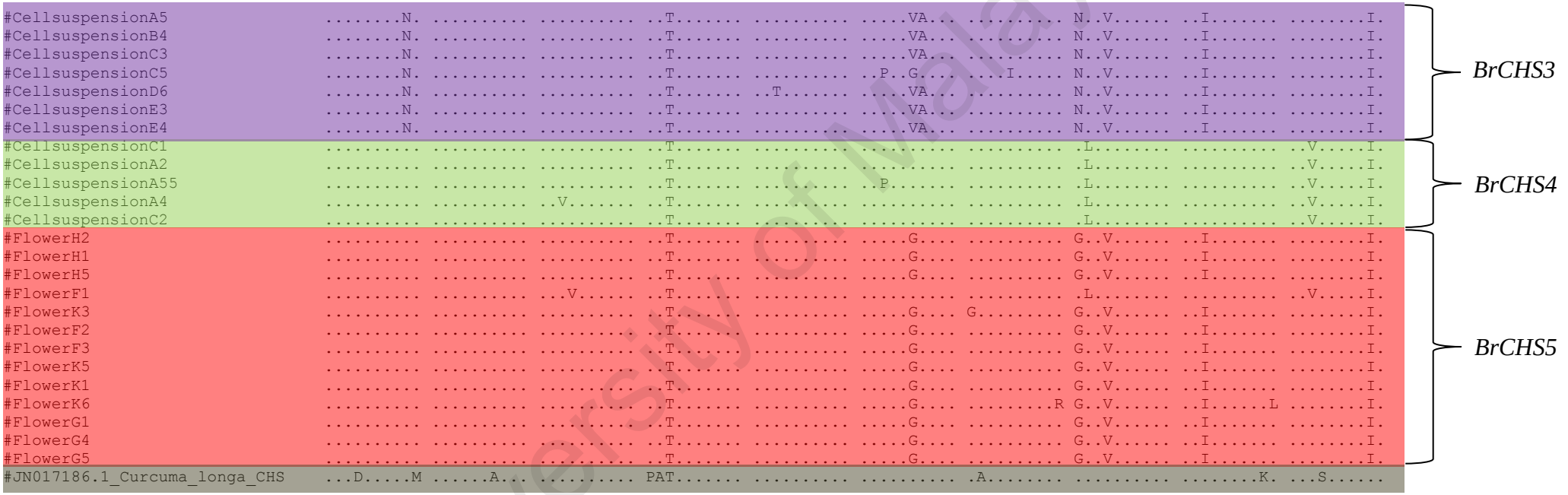


Figure 4.6, continued.

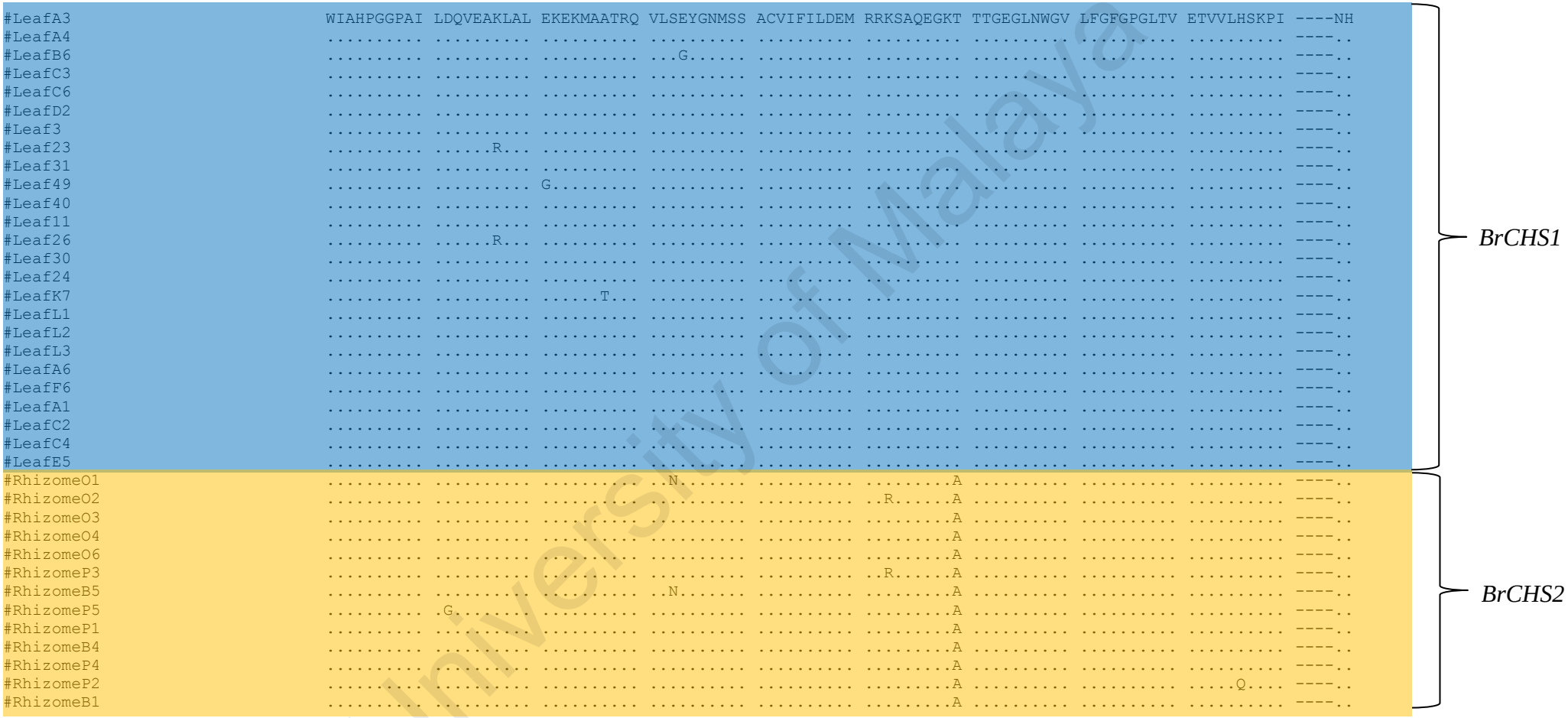


Figure 4.6, continued.

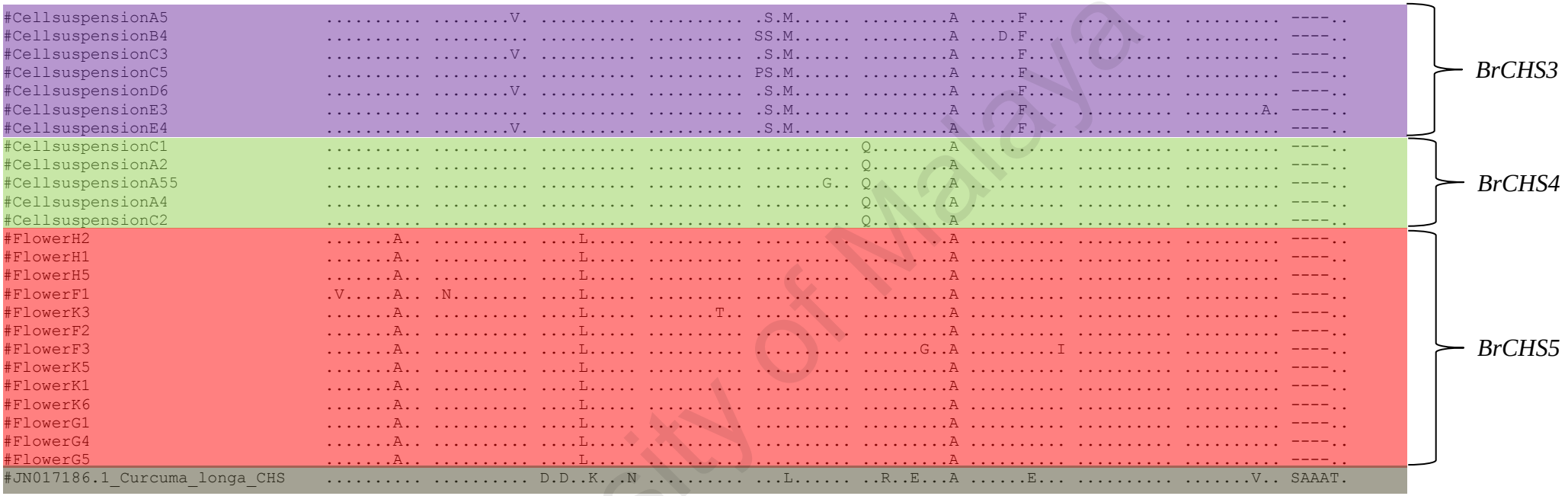


Figure 4.6, continued.

The blue box represents a group of *BrCHS1* transcript variants isolated from leaves, the orange box represents a group of *BrCHS2* transcript variants isolated from rhizomes, the purple and green boxes represent a group of *BrCHS3* and *BrCHS4* transcript variants isolated from suspension cultures and the red box represents a group of *BrCHS5* transcript variants isolated from flowers. The grey box represents CHS isolated from *Cucurma longa*.

The most common transcript variant from each group was selected and amino acid alignments were built and compared between five transcript variants. The amino acid alignments was also done with other plant's CHS (Figure 4.7). Genomic DNA extracted from leaves of *B. rotunda* was used for multiple sequence alignment between gDNA and cDNA fragment (refer to Appendix G). The *BrCHS* isolated from *B. rotunda* plants contained one intron and two exons with the presence of single intron was inserted in the first Cysteine (Cys) codon and was consistent with the result of previous studies such as in *Curcuma longa* (Wannapinpong et al. 2013) and *Gingko biloba* (Pang et al. 2005)

Based on the amino acid alignments, there were several conserved regions which were identified: (a) Cys^{164}-His^{303}-Asn^{336} as a catalytic triad for CHS-like enzymes, (b) a highly conserved Pro in the GFGP^{375}G loop which was a unique residue for CHS superfamily enzymes, and (c) active sites residues for CHS. These residues represented the domains for active sites and are important to serve as a scaffold during cyclization reaction. There were also a few variations detected among *BrCHS*s in cyclization pocket residues, which were Thr132Ser in *BrCHS5* transcript variant, Thr194Arg and Gly256Ala in *BrCHS3* transcript variant.

```
#BrCHS1       MAK----VQE IRQRQRAEGP AAILAIGTAT PTNVVYQADY ADYYFRITKS EHLTELKEKF KRMCDKSMIR KRYMHVTEEI
#BrCHS2       ...----... .......... .......... .......... .......... .......... .......... ..........
#BrCHS3       ...----... ..L....... .......K.. .......... ......V... .......... .......... .....L....
#BrCHS4       ...----... ..L...S... .......... .......... .......... .......... .......... .....L....
#BrCHS5       ...----... ..L...S... .......... .......... .......... .......... .......... .....L....
#ClPKS9       ...---L.T. ..KS...... ..V....... .P........ P.......R. ...V...... .......... ..H.YL....
#CaCHS        .T.---L.T. ..RS...... ..V......N .P........ P.......R. .......... .......... ..H.YL....
#MsCHS2       .VS----.S. ..KA...... .T.......N .A.C.E.ST. P.F..K..N. ..K....... Q........K R...YL....
#MaPKSIII3    ...----LA. ...S.....S .TV....... .V..L..... P.......... .......... .........H .....IN...
#OsCHS        ..-AAVT.E. V.RA...... .TV....... .A.C..... P.......... ..MV...... .......Q.. .....L....
#ZmCHS        ..GATVT.D. V.KG...T.. .TV....... .A.C..... P.......... D...D..... .......... .....L...F

#BrCHS1       LKENPNMCAY MEPSLDERQD IVVVEVPKLG KEAAAKAIKE WGQPKSKITH LIVCTTSGVD MPGADYQITK LLGLRPSVNR
#BrCHS2       .......... .......... .......... .......... .......... ..F....... .......... ..........
#BrCHS3       .......S.. .......... .......... .......... .......... ..F....... .......... ..........
#BrCHS4       .......... ......V... .......... .......... .......... ..F....... ...V...... ..........
#BrCHS5       .......... .......... .L........ .......... .......... ..F..S.... .......... ..........
#ClPKS9       .R...K.... ..A...A... .......R.. ....V..... .......... .VF....... .......L.. ..........
#CaCHS        .R...K.... ..A...A... .......R.. ....V..... .......... .VF......E .......L.. ..........
#MsCHS2       ......V.E. .A....A... M......R.. ....V..... .......... .......... .......L.. ......Y.K.
#MaPKSIII3    ......V... .A....A... .......... ....V..... ....E..... .VF....... .......L.. ..........
#OsCHS        .Q........ .A....A... .......... .A..Q..... ....R.R... .VF....... .......LA. M.....N...
#ZmCHS        .S...S.... .A....A... V..T...... .A..Q..... ......R... .VF....... .......L.. A.........
```

Figure 4.7: Amino acid alignment of BrCHSs and other plant's CHS

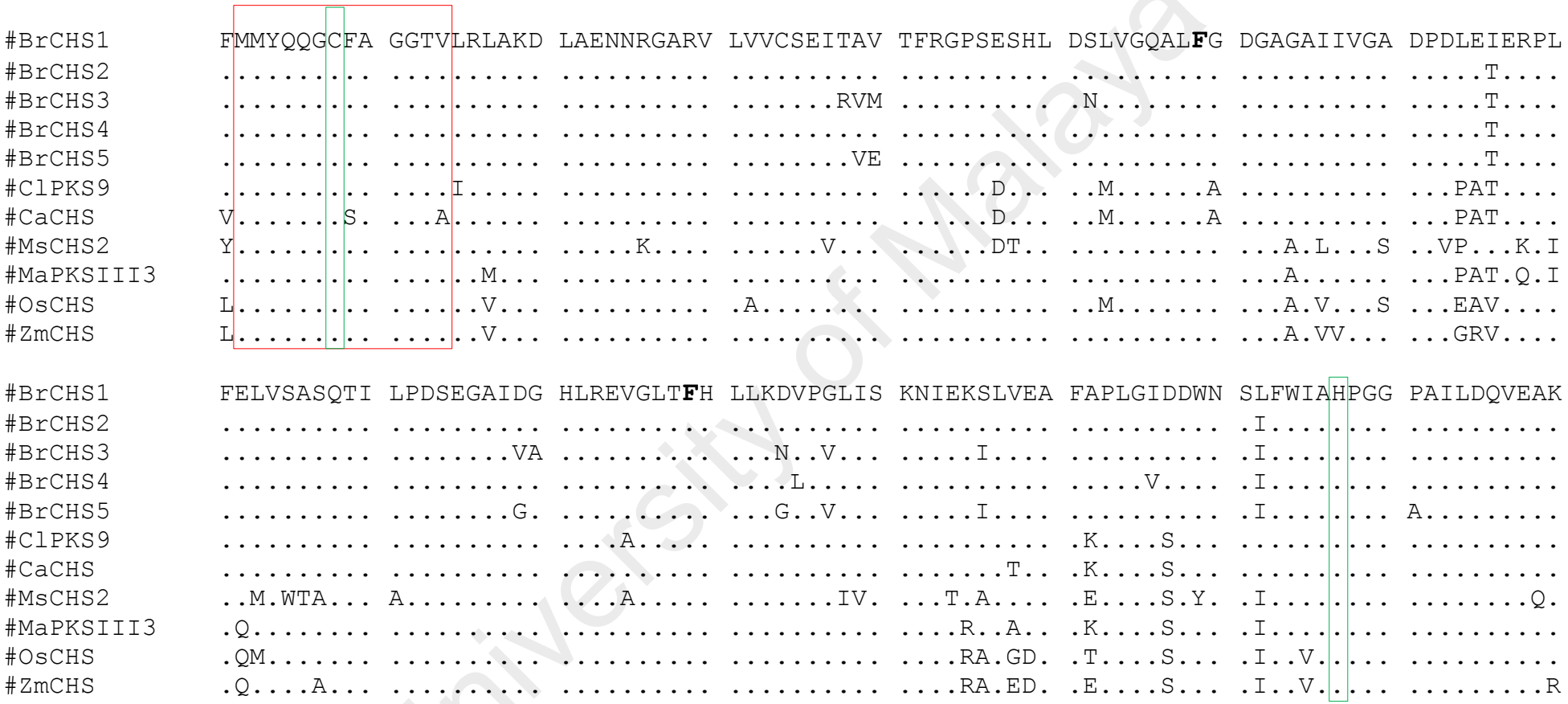


Figure 4.7, continued.

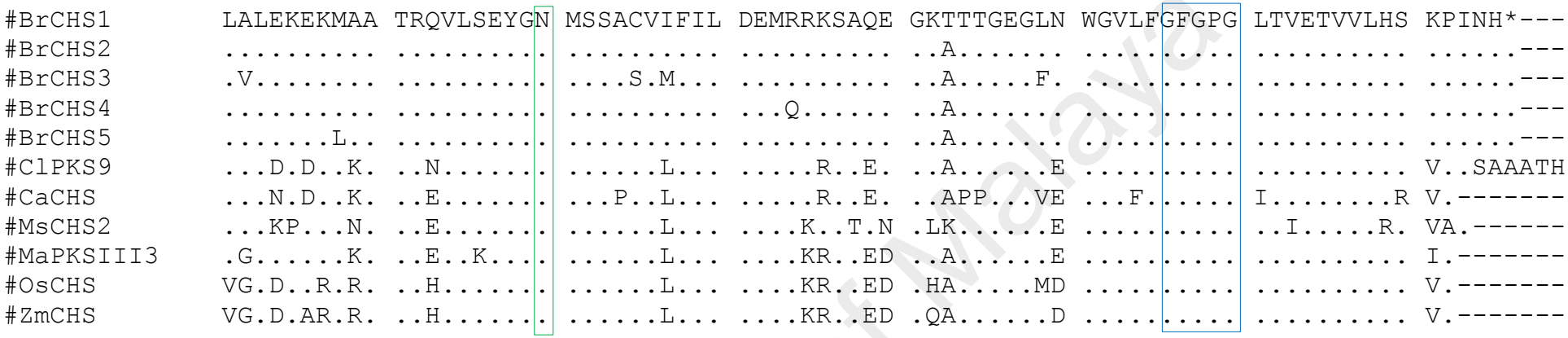


Figure 4.7, continued.

Alignment of CHS deduced amino acid sequences from *B. rotunda* (leaf, rhizome, flower and suspension cultures), *Cucurma longa* (ClPKS9; JN017186.1), *Curcuma alismatifolia* (CaCHS; GU140082.1), *Musa acuminata* (MaPKSIII3; GU724609.1), *Oryza sativa* (OsCHS; AB000801.2), *Zea mays* (ZmCHS; NM_001155550.1) and *Medicago sativa* (MsCHS2; L02902.1). The amino acids of the catalytic residues Cys^{164}-His^{303}-Asn^{336} triad were marked by green boxes. Meanwhile, the blue box represents Pro^{375} and $GFGP^{375}G$ loop and the red box represents active site residues for CHS. Bold sequence represents the Phe^{215} and Phe^{265}.

Sequence variation between five of *BrCHS* transcript variants shows a high percentage of similarity ranging from 97.0% to 99.1% between nucleotides and 94.3% to 99.0% between amino acids as shown in the Table 4.1. Both nucleotides and amino acids alignments showed that *BrCHS1* and *BrCHS2* transcript variants shared the highest similarity with only 10 nucleotides mismatch, which were causing differences in 4 amino acids in the coding DNA sequence (CDS). Meanwhile, *BrCHS3* and *BrCHS4* transcript variants that were differently grouped under the same plant sources; suspension cultures, showed that there were differences in 22 amino acids in the CDS. Both genes represent the lowest percentage amino acids similarity together with *BrCHS1* and *BrCHS3* transcript variants.

Table 4.1: The percentage similarity of nucleotides and amino acids among *BrCHS*s

	BrCHS1 (%)		BrCHS2 (%)		BrCHS3 (%)		BrCHS4 (%)	
	Nuc.	A.a.	Nuc.	A.a.	Nuc.	A.a.	Nuc.	A.a.
BrCHS2 (%)	99.1	99.0						
BrCHS3 (%)	97.0	94.3	97.4	95.4				
BrCHS4 (%)	98.0	96.9	98.3	98.0	96.6	94.4		
BrCHS5 (%)	97.6	95.6	98.0	96.7	97.0	95.4	97.8	96.2

Nuc.: Nucleotides, A.a.: Amino acids

The amino acid alignments of BrCHS clones were subjected to phylogenetic analysis to verify the BrCHS groups. According to phylogenetic analysis, all of *BrCHS*s were branched differently as shown in Figure 4.8. These transcript variants were predominantly expressed in different parts of the plants; therefore, they were further characterized as five different *BrCHS*s. This observation thus verified the previous preliminary group of five *BrCHS*s based on their amino acid sequences alignment (refer to Figure 4.6). Based on the cluster results, one *BrCHS* was predominantly transcribed in leaf (*BrCHS1*), rhizome (*BrCHS2*) and flower (*BrCHS5*) while two *BrCHS*s were predominantly transcribed in suspension cultures (*BrCHS3* and *BrCHS 4*).

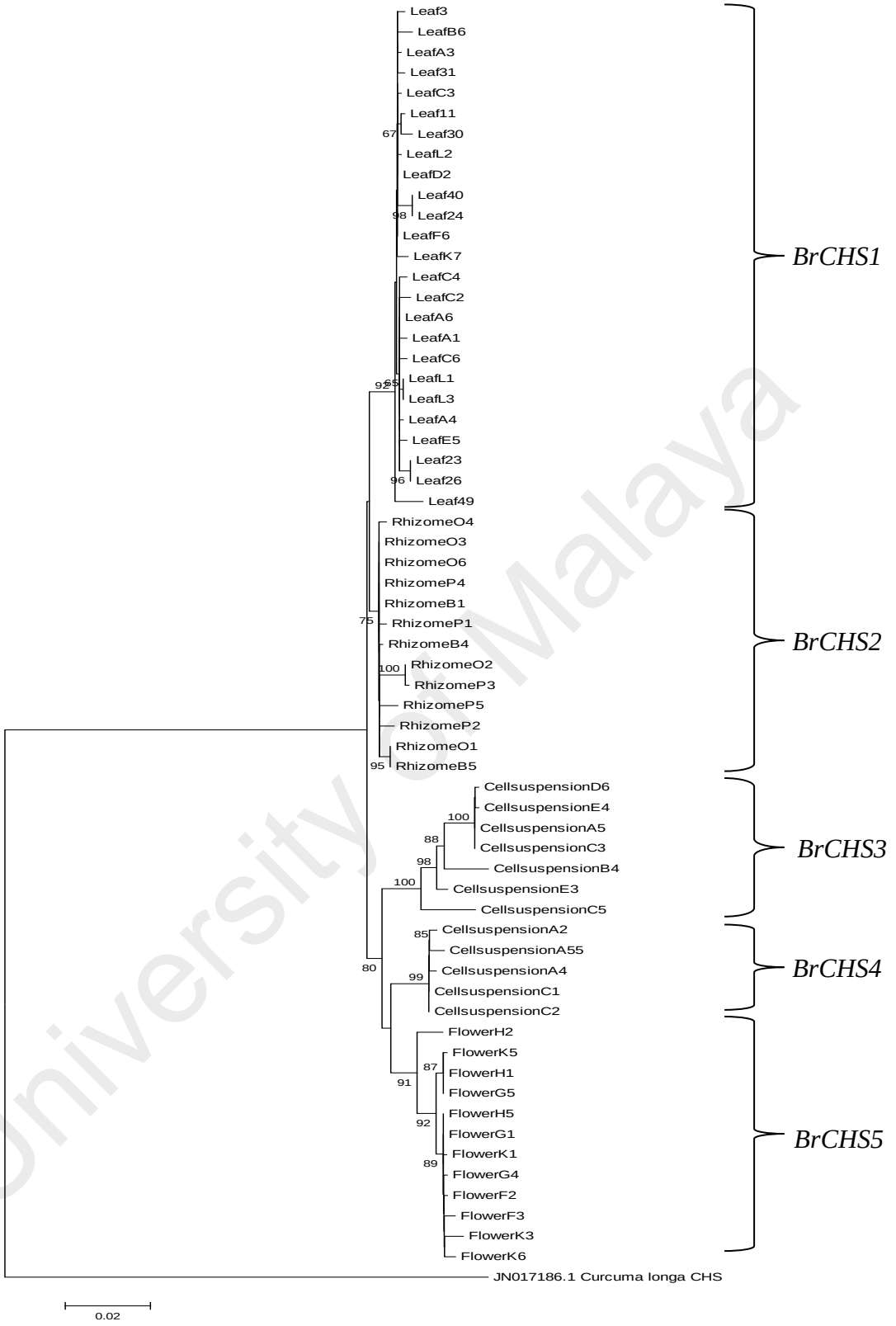


Figure 4.8: Evolutionary relationships of taxa

The phylogenetic tree was inferred using the Neighbor-Joining method based on deduced amino acid sequence. CHS from *Curcuma longa* (JN017186.1) was used as an outgroup. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown above the branches.

In addition, the representative of *BrCHS*s were also aligned with *CHS* of some other plants which were extracted from GenBank database. Amino acids from five *BrCHS*s from *B. rotunda* were clustered together with other monocots in Zingiberales Order and most closely related to Curcuma genus (*C. longa* and *C. alismatifolia*) as shown in Figure 4.9. Comparative analysis also revealed that the deduced amino acid sequences of *BrCHS* was highly homologous to *CHS* from other monocotyledonous plant species. For example, *M. acuminata* (87-89%), *O. sativa* (81-82%), *C. longa* (86-89%), *C. alismatifolia* (84-88%), *Z. mays* (79-82%) and *M. sativa* (79-82%).

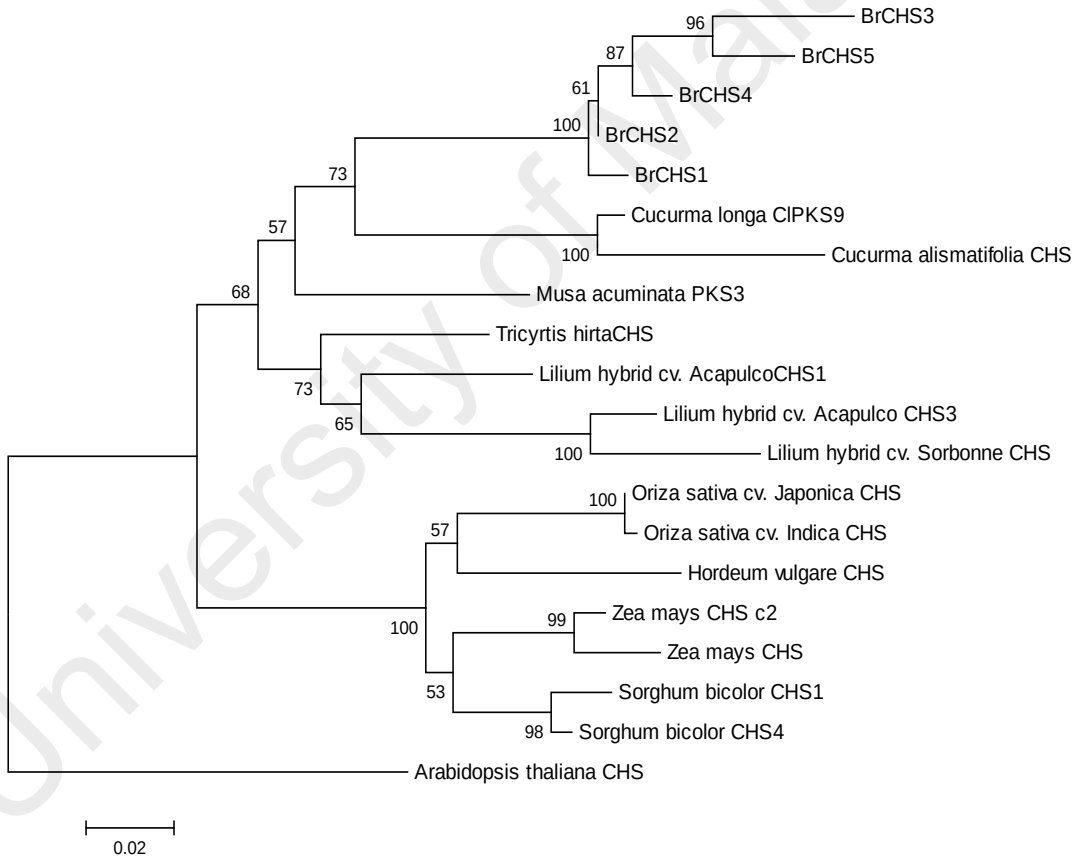


Figure 4.9: Phylogenetic tree among BrCHS and other plant's CHS

The phylogenetic tree was inferred using the Neighbor-Joining method. Alignment of deduced amino acid sequences was performed among five BrCHSs from *B. rotunda* plants (*BrCHS1, BrCHS2, BrCHS3, BrCHS4* and *BrCHS5*) and CHS in various monocot plants; *Cucurma longa* (ClPKS9; JN017186.1), *Curcuma alismatifolia* (CaCHS; GU140082.1), *Musa acuminata* (MaPKSIII3; GU724609.1), *Oryza sativa* (OsCHS; AB000801.2), *Zea mays* (ZmCHS; NM_001155550.1) and *Medicago sativa* (MsCHS2; L02902.1). The numbers above the branches refer to the bootstrap values.

4.5 Determining copy number by Southern blot analysis

Since there are four to five sequences identified which suggest the presence of four to five copies of BrCHSs; southern blot analysis was performed to corroborate the result. To achieve this, high molecular weight of DNA was successfully extracted from leaves (L) and *BrCHS* was amplified and further purified by gel extraction and purification prior to probe preparation (Figure 4.10).

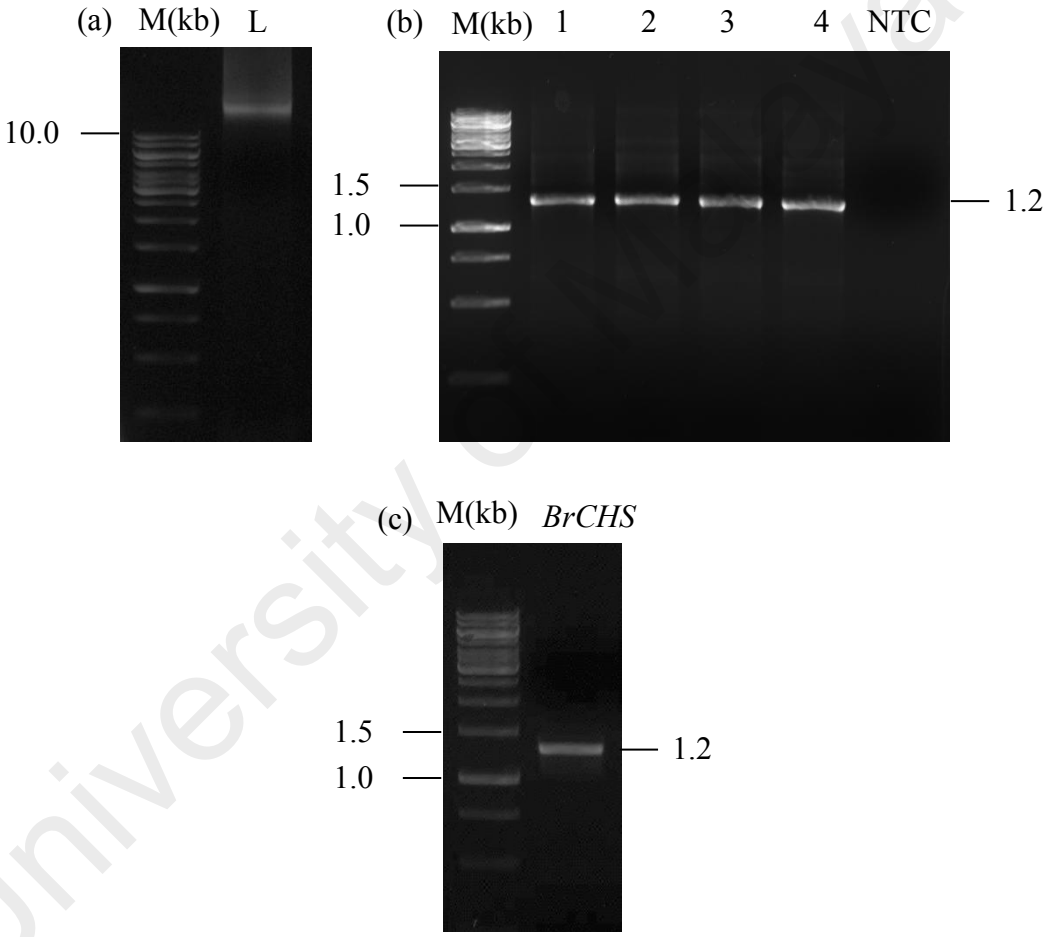


Figure 4.10: DNA extraction from leaves, amplification and purification of *BrCHS*

Southern blot analysis was prepared using (a) total genomic DNA extracted from leaves (L). Preparation of probe was set with (b) amplified *BrCHS* and (c) purified *BrCHS* fragment. Lane M: 1kb DNA ladder. NTC: Non-Template Control.

Purified *BrCHS* was DIG-labelled and subsequently subjected to probe efficiency test. A series of *BrCHS* probe concentrations was compared with DIG-labelled probe as positive control and reference (Figure 4.11). This BrCHS DIG-labelled probe was used for southern blot analysis. This analysis revealed that four to five bands can be clearly observed after genomic DNA digestion with *Dra*I and *Eco*RV, respectively (Figure 4.12).

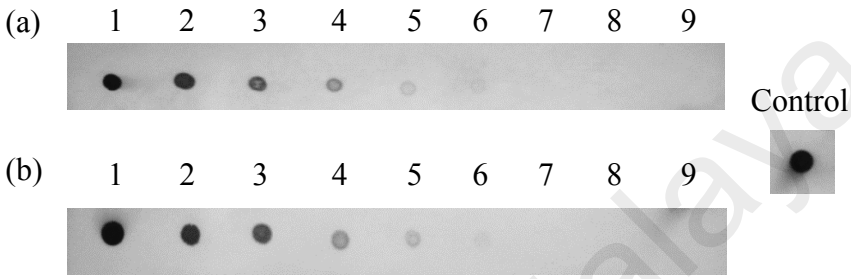


Figure 4.11: Serial dilution of DIG-labelled probe as reference for *BrCHS*-labelled probe

A total of nine serial dilution with different concentration (1 ng/µL – 0.01 pg/µL) showed from lanes 1 (high intensity) to 9 (low intensity) for (a) DIG-labelled probe and (b) *BrCHS*-labelled probe. The concentration BrCHS-labelled probe required for Southern blot analysis was calculated according to the standard concentration of DIG-labelled probe.

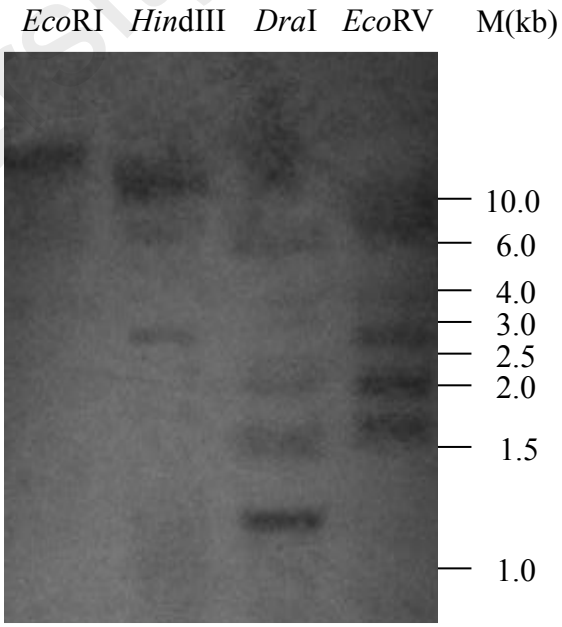


Figure 4.12: Southern blot analysis

Copy number was determined by digesting the DNA with four restriction enzymes, *Eco*RI, *Hin*dIII, *Dra*I and *Eco*RV. Based on the bands observed, there were at least four to five copies of *BrCHS*s presented in the plant genome.

4.6 Cloning of *BrCHS2* in *E. coli* (pCAMBIA1304 vector system)

4.6.1 Restriction endonuclease digestion of plasmid (pGEMT-BrCHS2)

Now, the second part of this study involves overexpression of *BrCHS2* in suspension cultures and its effect on flavonoid biosynthesis and related plant pathways. *BrCHS2* was overexpressed and the effect it had on flavonoid accumulation was investigated. For this purpose, the *BrCHS2* transcript variant which was predominantly expressed in rhizome was selected to be cloned into *B. rotunda* suspension cultures. Figure 4.13a shows the results after digestion with restriction endonucleases, *Nco*I and *Spe*I. The digested gene was purified prior to cloning with pCAMBIA1304 expression vector (Figure 4.13b).

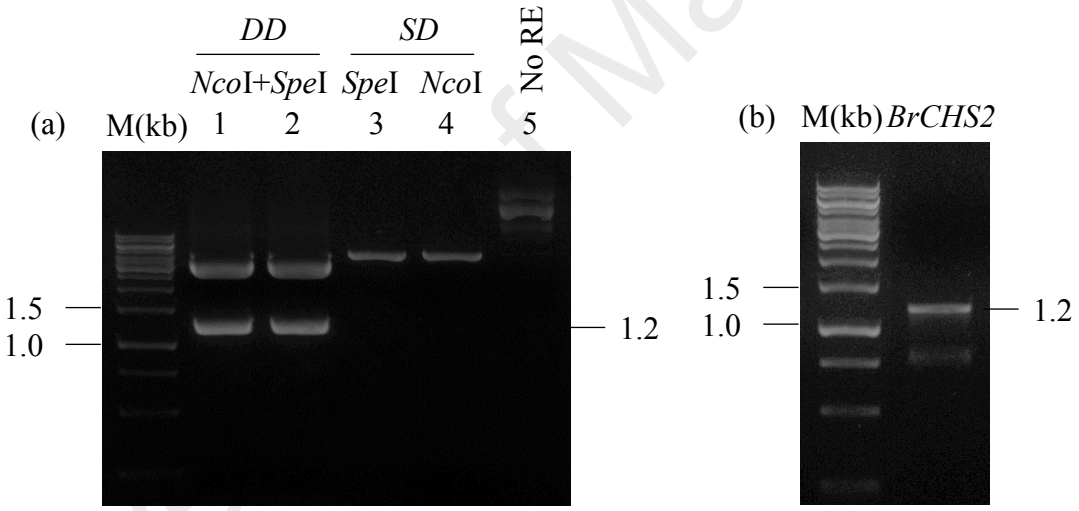


Figure 4.13: Plasmid DNA digestion and purified *BrCHS2* transcript variant

For plasmid DNA digestion, (a) Lane 1 and 2: Digested recombinant plasmid using double digestion *Nco*I and *Spe*I, Lane 3 and 4: Digested recombinant plasmid using single digestion *Nco*I or *Spe*I, Lane 5: Undigested recombinant plasmid. Digested fragment produced 1.2kb bands, M: 1kb DNA ladder (b) Purified *BrCHS2* transcript variant. DD: Double digestion, SD: Single digestion, M: 1kb DNA ladder.

4.6.2 Screening of *BrCHS* insert in pCAMBIA1304 vector system and sequencing analysis

After ligating with pCAMBIA1304 vector and transformed into *E. coli*, positive colonies harbouring a DNA insert with 1.2 kb of molecular weight were screened by PCR (Figure 4.14).

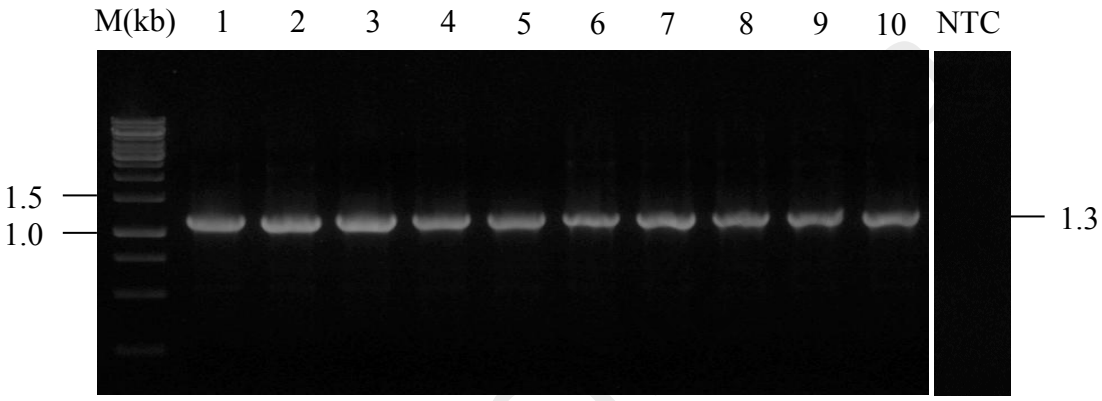


Figure 4.14: Positive colonies containing *BrCHS2* transcript variant in pCAMBIA1304 vector

Colony PCR analysis of pCAMBIA1304 harbouring *BrCHS2* transcript variant. Lane M: 1kb DNA Ladder. Lanes 1-10: Positive colonies with *pCAMBIA1304::BrCHS2* insert. NTC: Non-Template Control.

Prior to sequencing analysis of recombinant construct, *pCAMBIA1304::BrCHS2* was transformed in *Agrobacterium*. Colonies harbouring the desired gene were successfully screened for positive clones (Figure 4.15), amplified using vector-gene primers and purified as shown in Figure 4.16(a) and (b). Colonies with 1.3 kb size of band was selected and subjected to DNA sequencing.

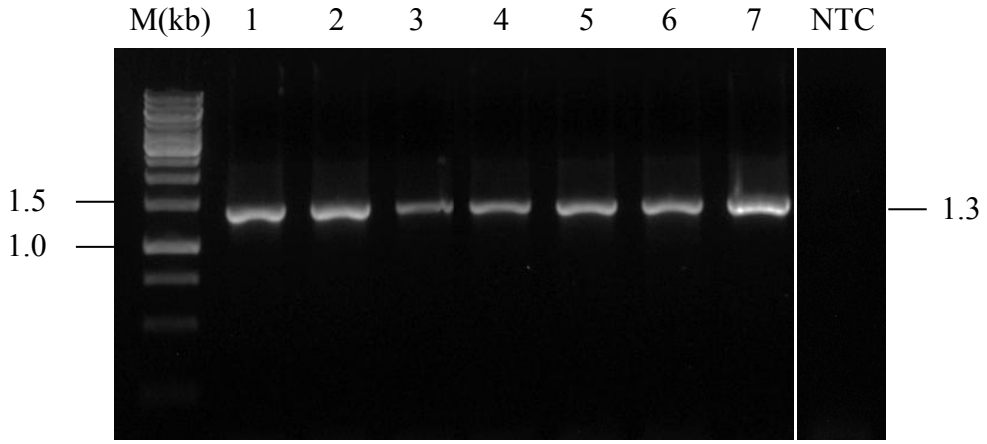


Figure 4.15: Positive colonies in *Agrobacterium* strain LBA4404

Successful amplification of recombinant *BrCHS2* from *Agrobacterium* colonies. Lane M: 1kb DNA ladder. Lanes 1-7: Colony PCR of recombinant *BrCHS2*. NTC: Non-Template Control.

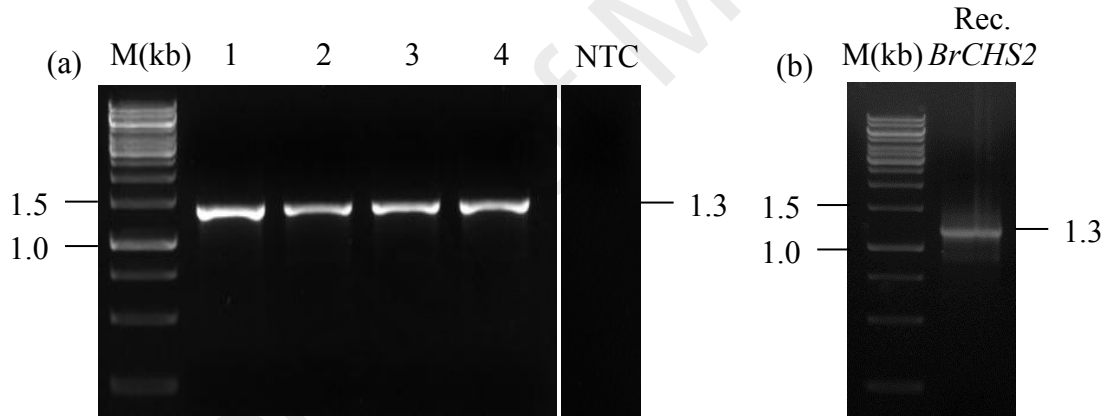


Figure 4.16: Amplified and purified of recombinant *BrCHS2*

Recombinant *BrCHS2* were successfully amplified as shown in (a) Lanes 1-4 and NTC: Non-Template Control. (b) Purified recombinant *BrCHS2*. Lane M: 1kb DNA ladder.

4.6.3 Sequence analysis of the recombinant construct

Recombinant *BrCHS2* that was previously purified was selected and sequenced in order to verify the boundary between 35S promoter, *BrCHS2* transcript variant and reporter gene; *mgfp*5 (Figure 4.17). The partial sequencing analysis showed that *BrCHS2* was confirmed and verified to be used for transformation and the insertion of

pCAMBIA1304::BrCHS2 revealed in-frame fusions together with the reporter gene (*mgfp*5) showed correct gene orientation.

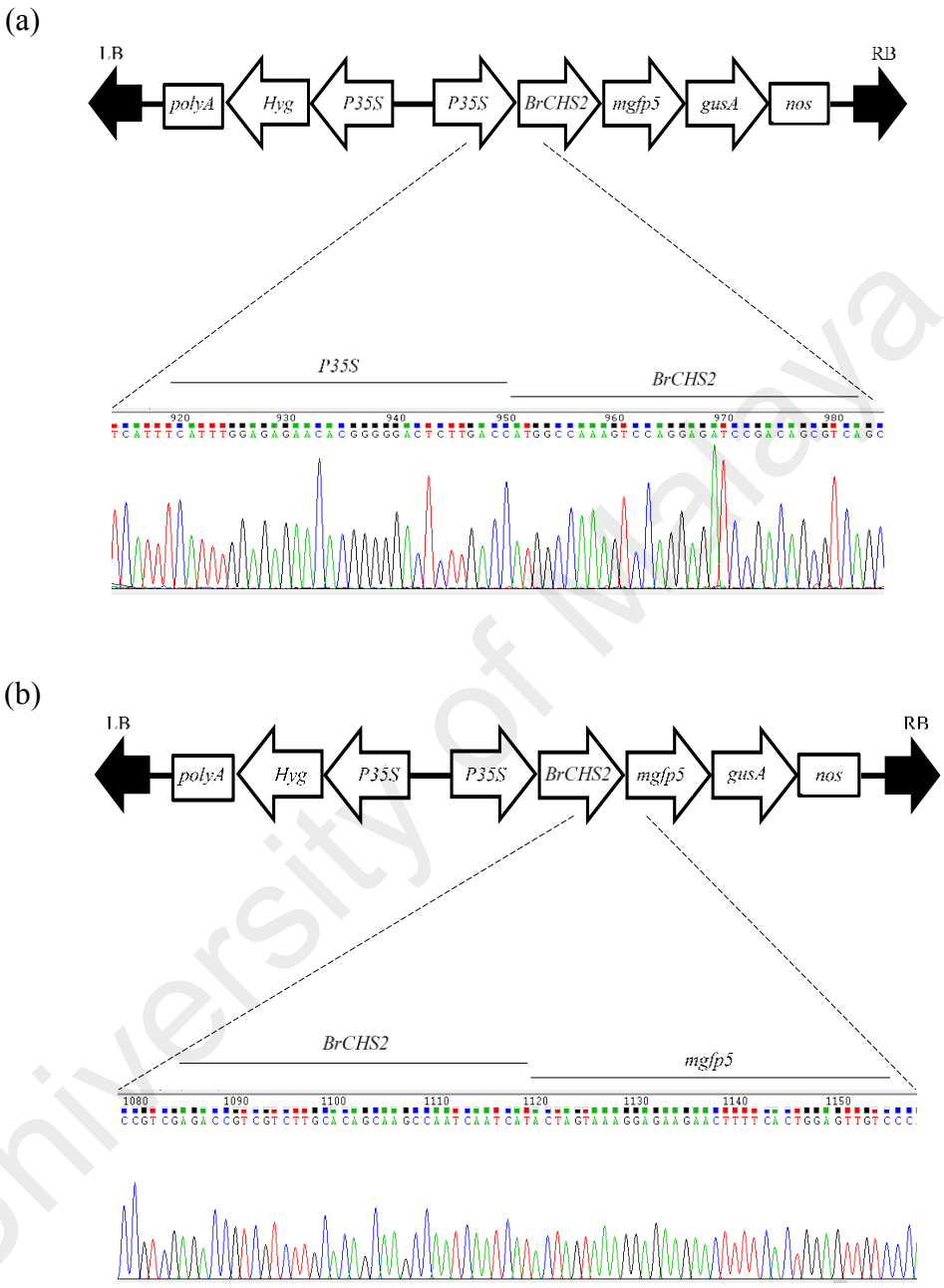


Figure 4.17: Partial DNA sequencing of *pCAMBIA1304::BrCHS2* construct

Partial DNA sequencing showed (a) Fusion of *P35S* promoter and *BrCHS2* gene and (b) Fusion of *BrCHS2* transcript variant with a reporter gene (*mgfp5).* RB: Right border, LB: left border, polyA: polyA-tail (polyadenylation), nos: nopaline synthase, Hyg: hygromycin, P35S: promoter 35S, mgfp5: modified green fluorescent protein 5, gusA: glucuronidase

4.6.4 Morphology of transgenic and wild-type callus

Suspension cultures were transformed using *Agrobacterium*-mediated method and cells were then grown on MS semi-solid media supplemented with Hygromycin antibiotic. Three representatives of transgenic callus and one wild-type were found to have similar yellow colour observed after several months of subculture with no changes in their callus morphology (Figure 4.18).

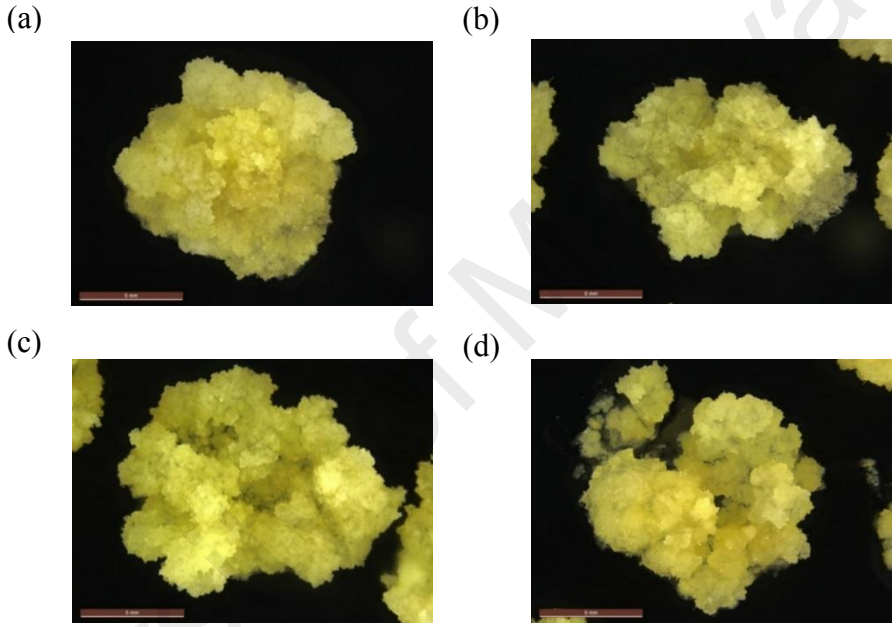


Figure 4.18: Morphology and colour of transgenic and wild-type callus

There were no differences in terms of the colour and morphology between (a) wild-type and three other transgenic callus (b, c and d). Callus were cultured on MS media supplemented with Hygromycin antibiotic.

4.6.5 GUS assay on transgenic callus and suspension cultures

GUS assay was chosen because one of the reporter genes, *gusA* was presence in the pCAMBIA1304 construct by referring to Jefferson et al. (1987). The introduction of *BrCHS* produced a fusion protein of BrCHS2 with the reporter. After three months of hygromycin selection, ten resistant callus that showed positive GUS coloration were

selected and transferred to liquid media for two months subcultures as shown in Figure 4.19.

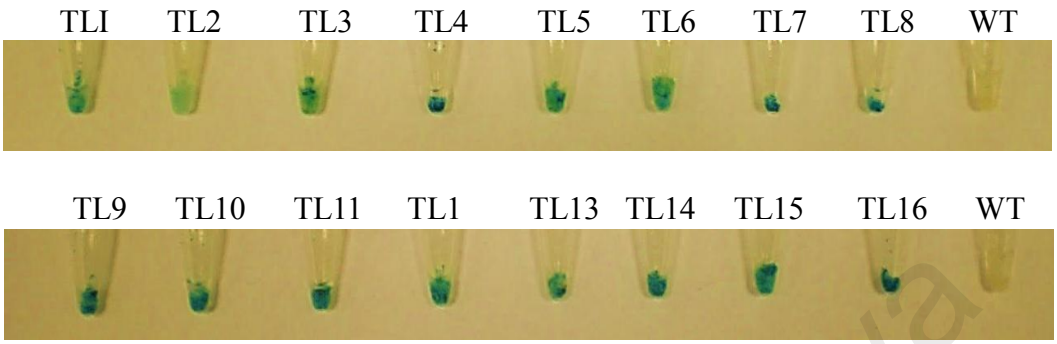


Figure 4.19: GUS assay on resistant callus

GUS assay on resistant callus (TL1-16) after three months of subculture. Colourless was observed in wild-type callus (WT).

4.6.6 Verification of transgenic suspension cultures

Stable line of ten suspension cultures was established for two months to get sufficient amount of cell cultures. A GUS assay was applied on the harvested cultures and GUS assay showed that recombinant *BrCHS2* was stably expressed in all ten transgenic lines compared to wild-type (Figure 4.20).

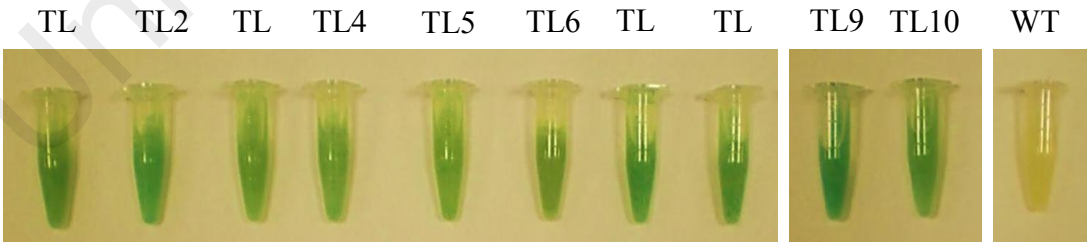


Figure 4.20: GUS assay on two month-old suspension cultures

After harvested, two-month-old suspension cultures were assayed with GUS and showed blue coloration in all ten transgenic lines (TL1-10) compared to wild-type (WT).

4.6.7 Multiplex PCR detecting recombinant BrCHS2, hptII and EF1α genes

Successful transformation of recombinant *BrCHS2* was confirmed by multiplex PCR analysis in ten transgenic suspension cultures. In each line, the recombinant *BrCHS2*, *hptII* and *EF1α* were successfully amplified with 1.3 kb, 0.7 kb and 0.2 kb size respectively (Figure 4.21). Selection of transgenic suspension cultures was based on the presence of *hptII* which detoxified the aminocyclitol antibiotic hygromycin B. Meanwhile, *EF1α* was a housekeeping gene that involved in basic functions needed for the sustenance of the cell and constitutively expressed in plants.

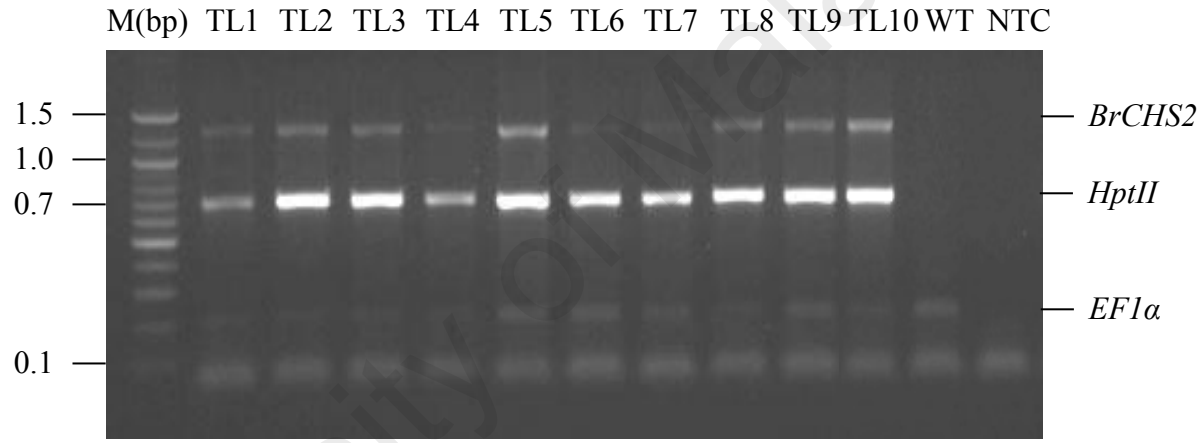


Figure 4.21: Multiplex PCR analysis

Multiplex PCR analysis shows all ten transgenic suspension cultures harbouring *BrCHS2* transcript variant and *HptII*. No amplified products were observed except for *EF1α* in wild-type (WT). Lane M: 100bp ladder, Lanes TL1-10: Ten transgenic suspension cultures, NTC: Non-Template Control.

4.7 Expression level of BrCHS in transgenic lines

4.7.1 Testing primer combination for RT-qPCR analysis

After two months of subculture, all ten transgenic suspension cultures were harvested and subjected to gene expression study. Before performing the analysis, two sets of primers were designed to amplify specific region; conserved region of *BrCHS* and *EF1α* gene (housekeeping gene). Different combinations of primer concentrations were

optimized to obtain high primer efficiency in RT-qPCR. The consistency and efficiency of mean Ct value were checked and verified during analysis. According to the optimization result, the best primer concentration for both primers (forward and reverse) was 100 nM as it has a consistent mean Ct value at cycle 23.486 (*BrCHS*) and 16.879 (*EF1α*), as shown in Figure 4.22 a) and b). This best primer combination indicated abundant of target nucleic acid successfully detected in real-time and have a strong reaction at the earlier cycle to detect a real signal from sample (exceed threshold level).

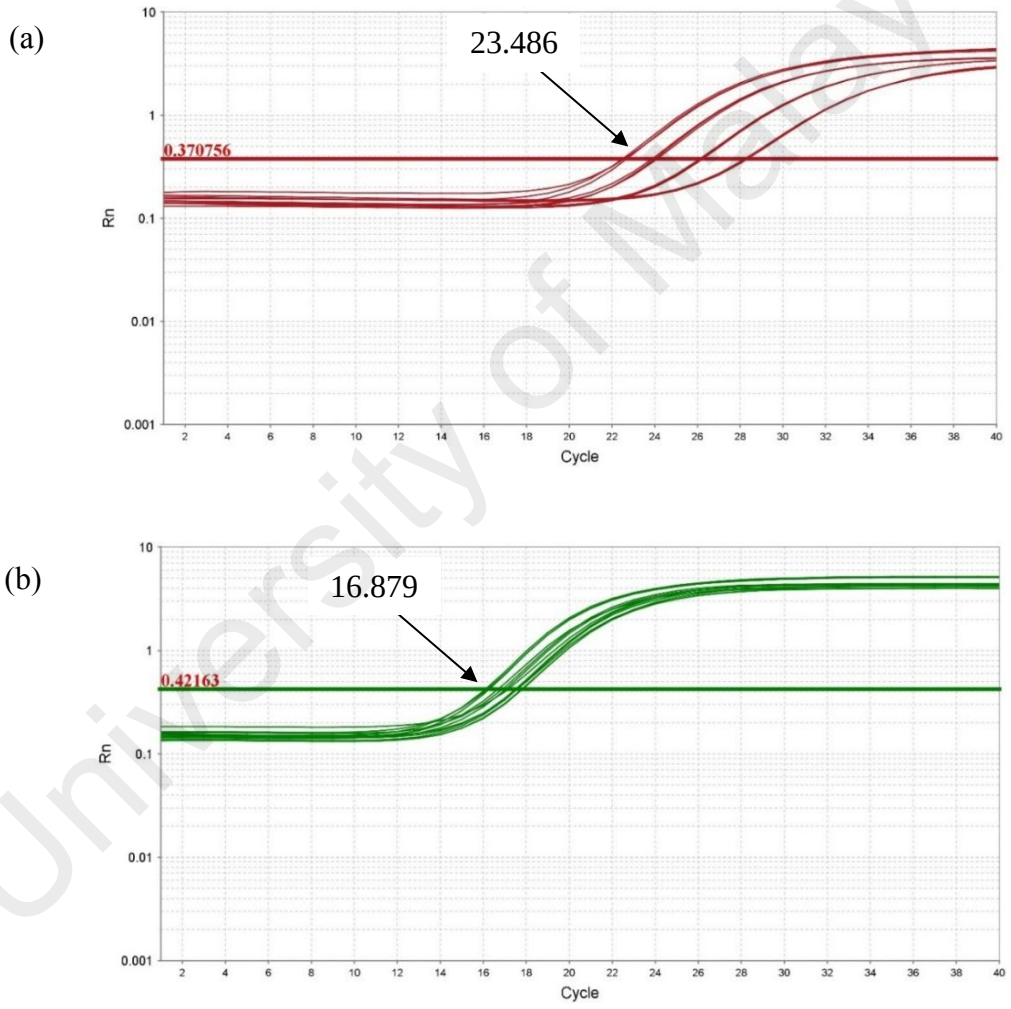


Figure 4.22: Amplification plot shows Ct value of different combination of primer concentration (Rn versus Cycle)

The above figures represent (a) Amplification plot for *BrCHS* and (b) Amplification plot of *EF1α* genes. Among four different combination of primer concentration, primers with 100nM forward (F) and 100nM reverse (R) were selected for expression level experiment of ten transgenic suspension cultures.

4.7.2 RT-qPCR on transgenic suspension cultures

Prior to RT-qPCR analysis, total RNA was successfully extracted from ten transgenic *B. rotunda* suspension cultures (Figure 4.23) including wild-type. Eight out of ten transgenic suspension cultures; TL1 (9.62-fold), TL2 (9.63-fold), TL3 (8.14-fold), TL5 (8.10-fold), TL6 (5.61-fold), TL7 (5.35-fold), TL9 (27.49-fold) and TL10 (19.74-fold) showed higher expression level of *BrCHS* level than that of wild-type suspension culture at significant level of $P<0.01$ and $P<0.001$ (Figure 4.24). TL4 and TL8 expressed 2.76-fold and 1.93-fold higher than wild-type.

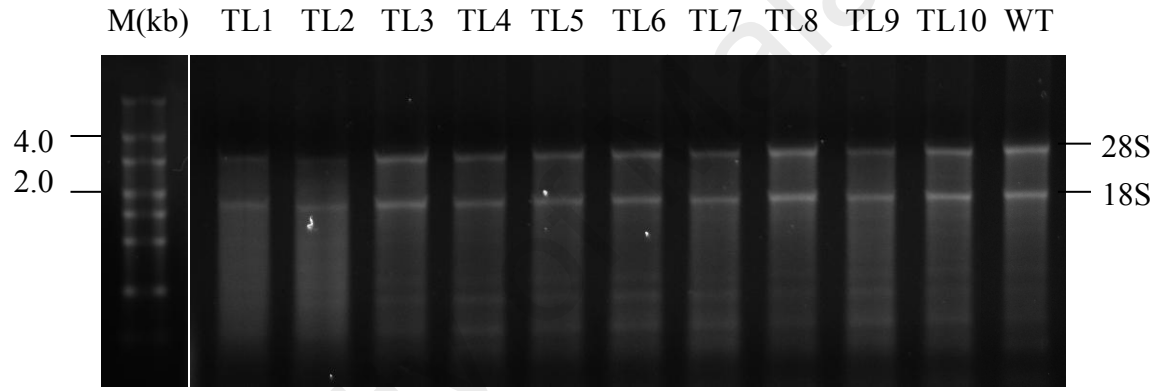


Figure 4.23: Total RNA extracted from two month-old suspension cultures

Total RNA was successfully extracted from suspension cultures. Lane M: RNA Ladder, Lanes TL1-10: Ten transgenic suspension cultures, WT: Wild-type suspension culture

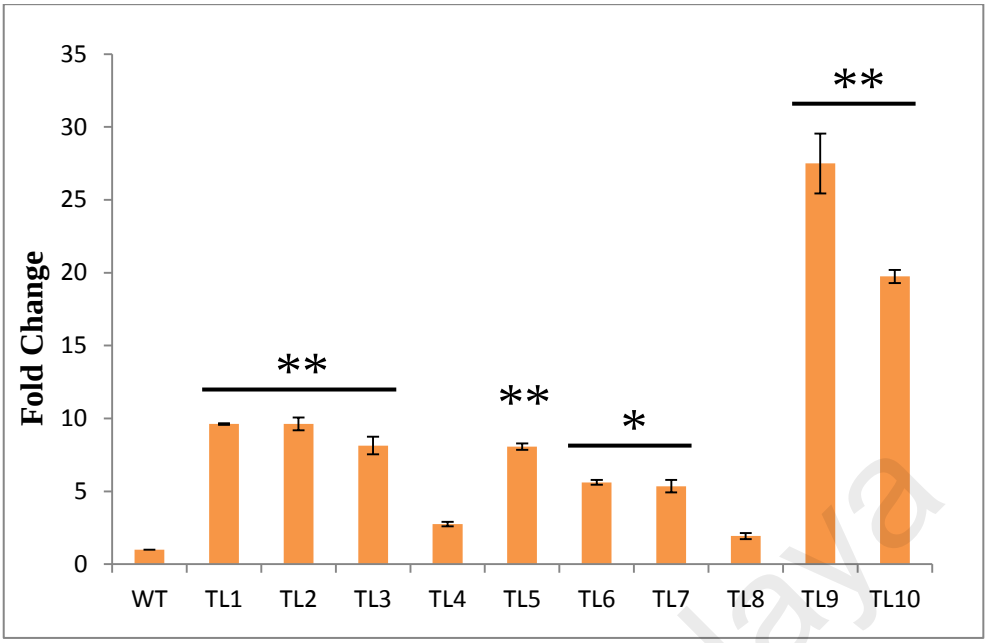


Figure 4.24: Expression level of *BrCHS* in ten transgenic lines

Ten transgenic lines (TL1-10) showed different expression levels (2 to 27-fold change) of *BrCHS* compared to wild-type (WT). The expression was normalized with elongation factor gene, *EF1α.* * Significant at P<0.01, ** Significant at P<0.001

4.8 CHS enzymatic assay

Referring to KEGG database, CHS is shown to be involved in multiple biosynthesis pathways. Different CHSs are required to react towards different substrates toproduce different compounds as shown in Figure 4.25. For example, one CHS works on Coumaroyl-CoA to produce naringenin while the others react with Cinnamoyl-CoA, Caffeoyl-CoA and Feruloyl-CoA substrate. In this study, the Cinnamoyl-CoA was chosen because it was reported to produce pinocembrin and was expected to direct the production of cyclohexenyl chalcone derivative compounds (CCDs) such as panduratin and hydroxypanduratin A that have several pharmaceutical interests.

Therefore, in order to see the functionality of BrCHS2, six independent transgenic lines (TL1, TL2, TL3, TL5, TL9 and TL10) that showed higher expression level (based on gene expression level result, refer to Figure 4.24), were selected and subjected to

Bradford assay for identification of total protein concentration (Table 4.2). The six reference compounds which include naringenin, naringenin chalcone, pinocembrin, pinocembrin chalcone, pinostrobin and pinostrobin chalcone were used to validate the CHS functionality. The retention times for all compounds identified in HPLC chromatogram were listed in Table 4.3 and referred prior to CHS enzymatic assay.

Table 4.2: Mean protein concentration according to Bradford assay standard

Suspension cultures	**Protein concentration (mean, µg/µL)**	**Standard Error (SE)**
Wild-type	8.20±0.1	2.75E-01
TL1	10.60±0.1	1.01E+00
TL2	9.43±0.1	1.46E+00
TL3	5.31±0.1	1.89E-01
TL5	7.47±0.1	3.23E-01
TL9	4.00±0.1	6.77E-01
TL10	6.70±0.1	7.18E-01

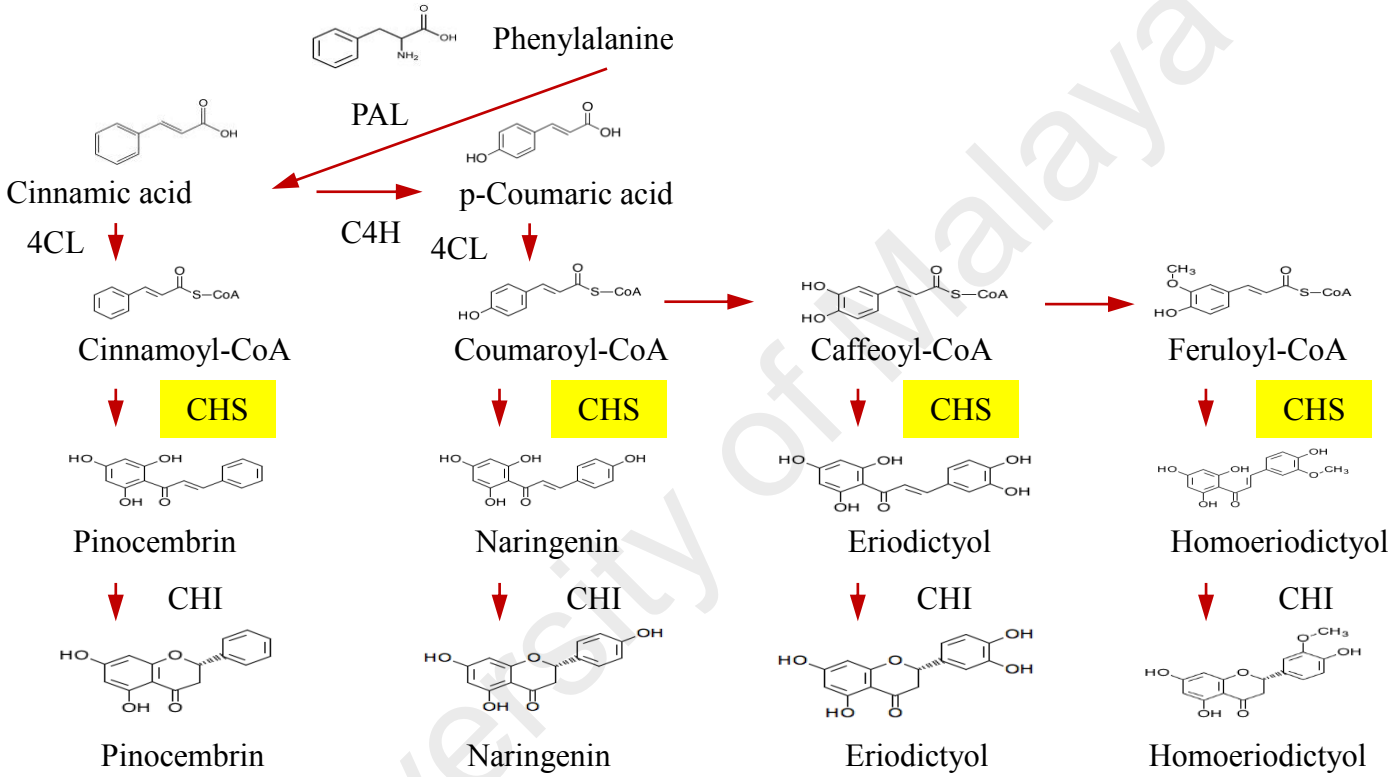


Figure 4.25: Multiple biosynthesis pathway of CHS

Different CHS enzyme reacts with different set of substrates to produce various flavonoid compounds. These bioactive compounds include pinocembrin, naringenin, eriodictyol and homoeriodictyol that have pharmaceutical values. Abbreviations: PAL, phenylalanine ammonia lyase; C4H, cinnamate 4-hydroxylase; 4CL, 4-coumaroyl:coenzyme A ligase

Table 4.3: Retention time of substrates and six reference compounds

Reference compounds	Retention time (min)
Naringenin	6.531±0.10
Naringenin Chalcone	6.571±0.10
Pinocembrin	10.304±0.10
Pinocembrin chalcone	8.913±0.10
Pinostrobin	12.107±0.10
Pinostrobin chalcone	10.727±0.10
Malonyl-CoA (substrates)	2.231±0.10
Cinnamoyl-CoA (substrates)	5.068±0.10

As expected, one of the compound, pinocembrin chalcone (intermediate compound of pinocembrin) was detected in all six transgenic lines. There was a significant increase in the amount of compounds in all six transgenic lines compared to that of wild-type at significant level of $P<0.05$ when using Cinnamoyl-CoA and Malonyl-CoA as substrates. Pinocembrin chalcone labelled as C1 in Figure 4.26, was eluted at a retention time of 8.90 min. Another abundant compound labelled as C2 was eluted at a retention time of 11.70 min and was detected in both wild-type and transgenic lines. This compound was polyethyleneglycol (PEG), possibly derived from Triton X-100, and after was confirmed by liquid chromatography-mass spectrometry (LC-MS) (refer to Appendix H). In addition, there were two other unidentified compounds labelled as C3 and C4 which were eluted at a retention time of 12.40 min and 14.00 min, respectively.

(a)

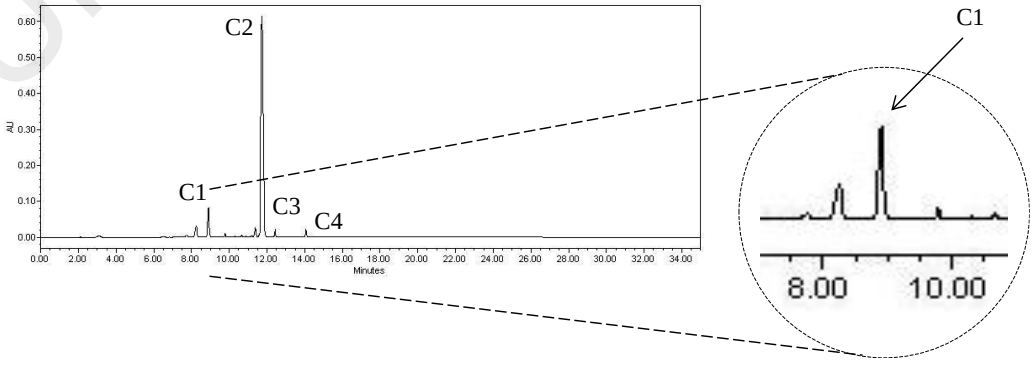
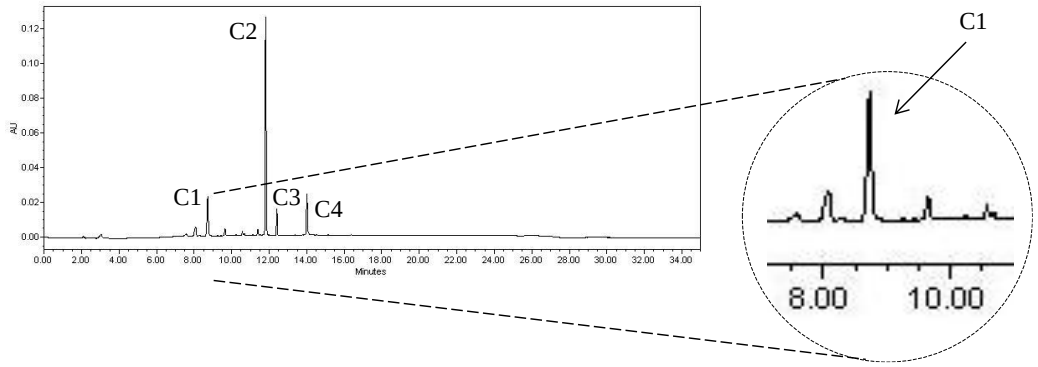
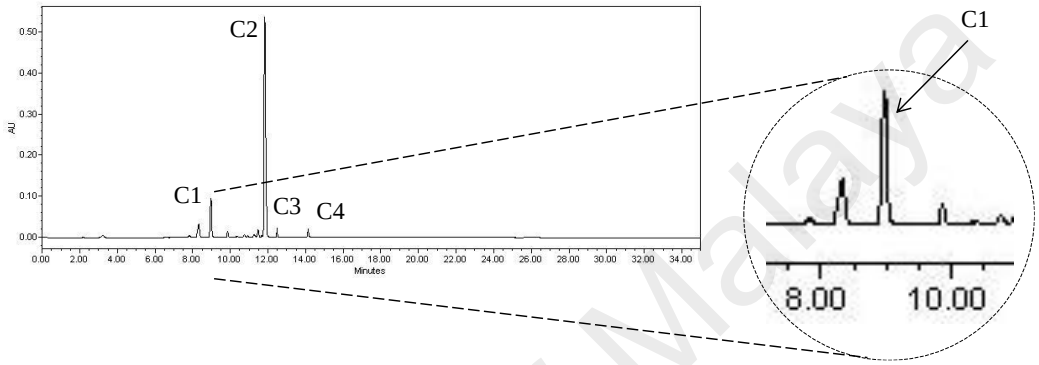


Figure 4.26: HPLC chromatogram detecting pinocembrin chalcone (C1) compound

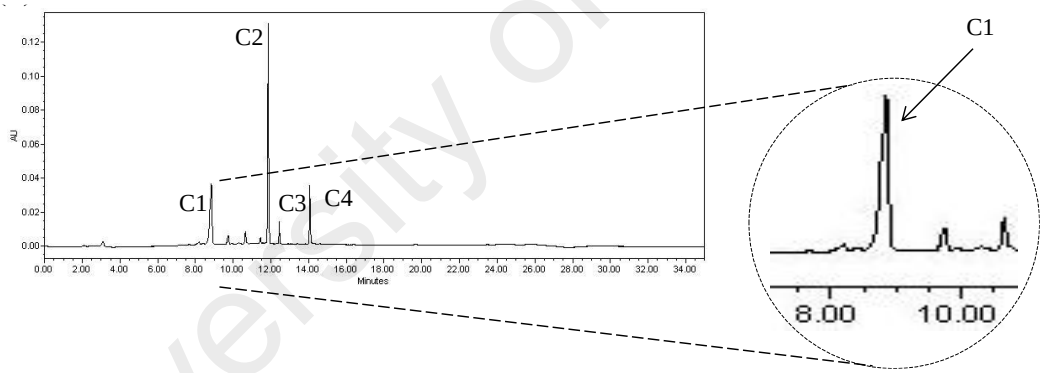
(b)



(c)



(d)



(e)

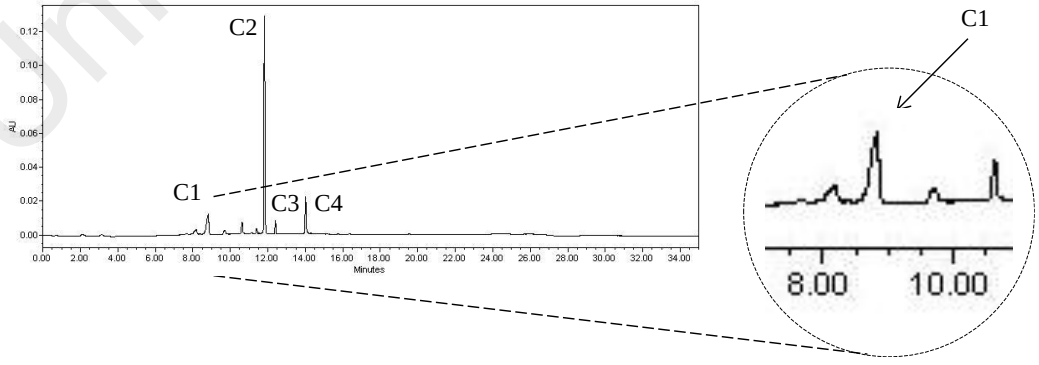
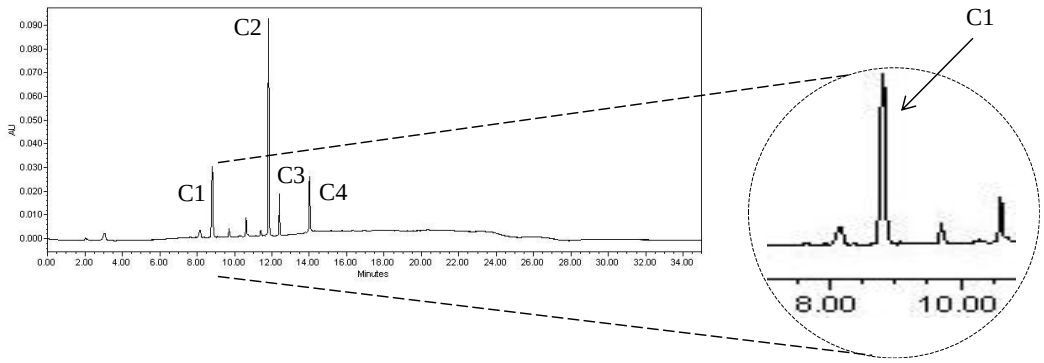


Figure 4.26, continued.

(f)



(g)

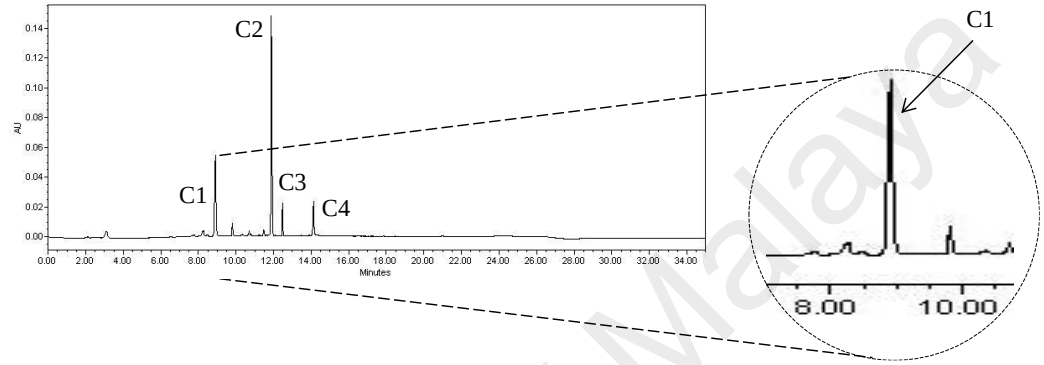


Figure 4.26, continued.

After CHS enzymatic assay, total compounds were extracted from wild-type and six transgenic lines and separated in HPLC: (a) Wild-type, (b) TL1, (c) TL2, (d) TL3, (e) TL5, (f) TL9 and (g) TL10.

Table 4.4 summarize the list of four major compounds produced after CHS enzymatic assay and ethyl acetate partitioning in percentage. Compared to the wild-type suspension cultures, pinocembrin chalcone (C1) that was produced in all six transgenic lines had increased significantly at $P<0.001$, except for TL2 while by-product of PEG (C2) had decreased in all transgenic lines. The highest yield of pinocembrin chalcone was detected in TL3 followed by TL10 and TL9. The yield was calculated by per gram of total crude protein.

Table 4.4: The amount of four major compounds detected in HPLC analysis

Compound	**C1**		**C2**		**C3**		**C4**	
Retention time (min)	8.90±0.10		11.70±0.10		12.40±0.10		14.00±0.10	
Suspension cultures	Amount (mg/g)	S.E.	Amount (mg/g)	S.E.	Amount (mg/g)	S.E.	Amount (mg/g)	S.E.
WT	64.67	6.67E-02	818.17	1.58E+00	9.80	5.77E-01	12.03	3.71E-01
TL1	154.43[a,b,c]	7.42E-01	521.37 [a,b,c]	3.97E+00	53.27 [a,b,c]	1.03E+00	105.50 [a,b,c]	1.47E+00
TL2	112.20 [c]	1.00E-01	715.23 [a,b,c]	1.43E+00	16.20	1.15E-01	17.40	5.77E-02
TL3	275.63 [a,b,c]	7.69E+00	446.33 [a,b,c]	9.34E+00	36.27 [a,b,c]	1.17E+00	90.03 [a,b,c]	1.30E+01
TL5	131.83 [a,b,c]	2.50E+00	574.43 [a,b,c]	7.79E+00	29.23 [b,c]	1.44E+00	96.77 [a,b,c]	3.41E+00
TL9	207.30 [a,b,c]	5.47E+00	418.23 [a,b,c]	7.46E+00	75.50 [a,b,c]	6.01E+00	112.33 [a,b,c]	5.82E+00
TL10	231.03 [a,b,c]	1.83E+01	447.00 [a,b,c]	4.03E+01	49.77 [a,b,c]	2.98E+00	100.30 [a,b,c]	1.67E+01

[a] Significant at P<0.001, [b] Significant at P<0.01, [c] Significant at P<0.05

4.9 RNA-Seq profiling

4.9.1 *De novo* transcriptome assembly

Based on the previous results, high gene expression level and high yield of pinocembrin chalcone in transgenic suspension cultures were successfully obtained, and their profile RNAs on its effect of *BrCHS2* overexpression was further discovered through RNA-Seq analysis. Whole transcriptome sequencing of two months old wild-type and transgenic suspension cultures were carried out using Illumina Hiseq™2000 (Illumina Inc, USA) platform at Beijing Genome Institute (BGI), Shenzhen, China.

A total of 31.5 million sequences reads corresponding to 1.2Gb of raw sequence reads using system were successfully generated. After a stringent data cleaning and quality check, the sequencing process showed an average GC content of 49.19% and average reads quality of 97.95% Q20 bases as shown in Table 4.5.

After clustered by TGI Clustering Tools (TGICL), a total of 102,195 unigenes with a total length of 103,043,426nt were generated. All-unigenes sequences from six samples with the mean size of assembled unigenes were 1008 bp. The dominance of this longer unigenes suggested that the quality of our assembly was high as it was also represented by the value of N50 (1646 bp) as shown in Table 4.6.

Table 4.5: Summary of reads assembly generated by Trinity from control and transgenic *B. rotunda* suspension cultures

Samples	Total Raw Reads	Total Clean Reads	Total Clean Nucleotides (nt)	Q20 percentage	GC percentage
WT-R1	71,160,282	64,400,870	5,796,078,300	97.94%	49.58%
WT-R2	66,868,696	61,111,682	5,500,051,380	97.93%	50.23%
WT-R3	54,766,806	49,816,986	4,483,528,740	97.80%	49.50%
TL-R1	56,634,676	52,543,920	4,728,952,800	97.96%	48.83%
TL-R2	69,256,744	64,557,498	5,810,174,820	97.99%	48.29%
TL-R3	63,821,100	57,834,176	5,205,075,840	98.06%	48.68%

Table 4.6: Summary of total number of contigs and unigenes from control and transgenic *B. rotunda* suspension culture

	Sample	**Total Number**	**Total Length (nt)**	**Mean Length (nt)**	**N50**
Contig	WT-R1	142,208	48,314,260	340	554
	WT-R2	130,497	45,497,655	349	577
	WT-R3	133,480	43,536,894	326	548
	TL-R1	151,276	48,249,930	319	523
	TL-R2	159,131	51,902,397	326	545
	TL-R3	146,220	48,581,277	332	524
Unigene	WT-R1	87,451	58,809,010	672	1217
	WT-R2	81,443	53,380,869	655	1152
	WT-R3	72,559	50,422,783	695	1237
	TL-R1	80,996	57,116,019	705	1265
	TL-R2	88,436	64,849,473	733	1348
	TL-R3	92,055	58,084,783	631	1146
	All	102,195	103,043,426	1008	1646

All assembled unigenes were more than 300 bp in size and the maximum length of unigenes was 101,630 bp. Among these unigenes, 61,838 unigenes (60.85%) were in the length range of 300 to 1000, 25,301 unigenes (24.89%) were in the length ranging from 1000 to 2000 and 14,536 unigenes (14.30%) were longer than 2000 bp (Figure 4.27). Compared with previous *B. rotunda* transcriptome databases (Md-Mustafa et al. 2014), the average length and the N50 values (1646bp) in our study were much longer. In addition, nearly half of these unigenes were more than 1000bp in size. All of the data indicate that the integrity of unigenes was relatively good.

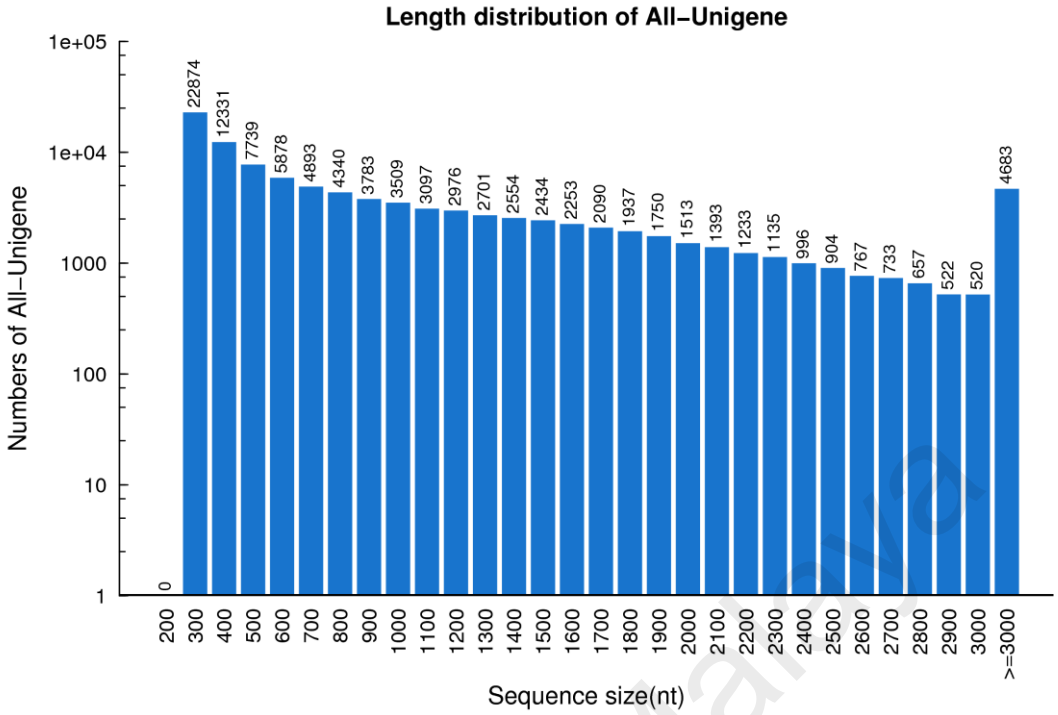


Figure 4.27: Histogram of the unigenes length distribution

The X-axis indicates the length range of unigenes. The Y-axis denotes the number of unigenes in every range of length.

4.9.2 Functional annotation against database

Most of the unigenes were successfully annotated using the Nr database (63.9%), followed by Nt (49.6%), Swiss-Prot (44.1%), KEGG (40.8%), GO (46.1%) and COG 26.7%) (Table 4.7). Therefore, out of 102,195 All-unigenes, only 67,251 unigenes (65.8%) were successfully annotated, which showed similarity to known proteins in all databases. The remainder which account of 34.2 % were unannotated.

Table 4.7: Functional annotation of the *B. rotunda* suspension cultures in public databases

Annotated database	No. of unigene hits	Percentage
Nr	65296	63.9%
Nt	50704	49.6%
Swiss-Prot	45050	44.1%
KEGG	41749	40.8%
GO	47111	46.1%
COG	27330	26.7%
Total annotated unigenes	67251	65.85

In addition, E-value distribution of the top hits in the Nr database revealed that 58.10% of the mapped sequences showed significant homology (less than 1.0E-^{45}) (Figure 4.28a), and 54.30% and 14.90% of the sequences with similarities greater than 60% and 80%, respectively, were found (Figure 4.28b). Besides that, based on BLASTX results, 65,296 unigenes were matched with homologous proteins in the Nr protein databases with 13,255 (20.3%) unigenes showed significant homology with sequences of *Vitis Vinifera*, and 7,966 (12.2%) and 4,505 (6.9%) unigenes had high similarities with the sequences of Japanese rice and *Zea mays*, respectively (Figure 4.28c).

(a)

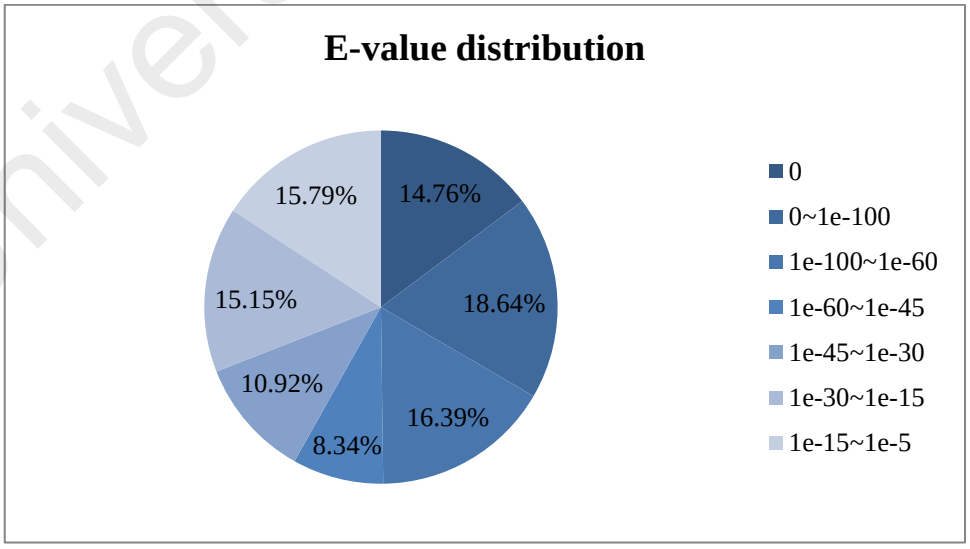
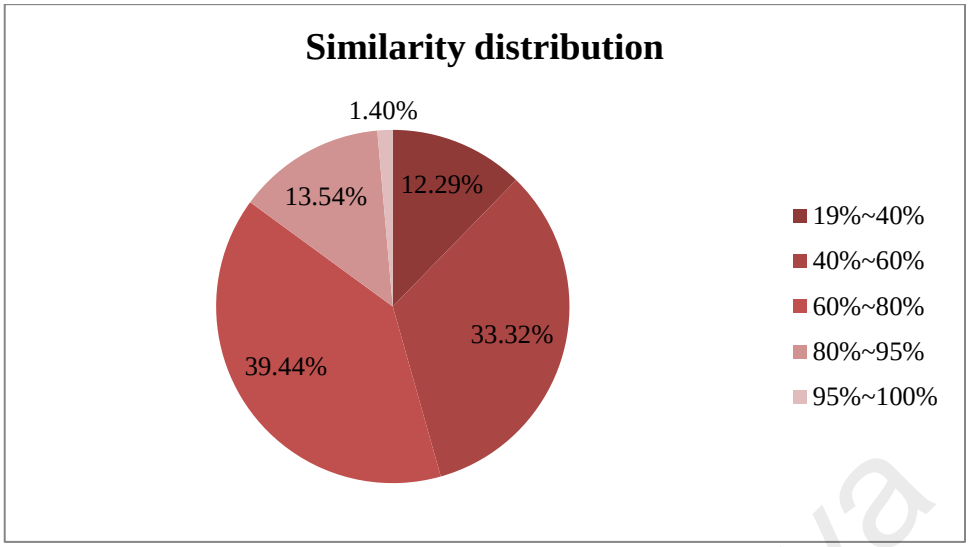


Figure 4.28: Characteristics of similarity search of unigenes against Nr databases

(b)



(c)

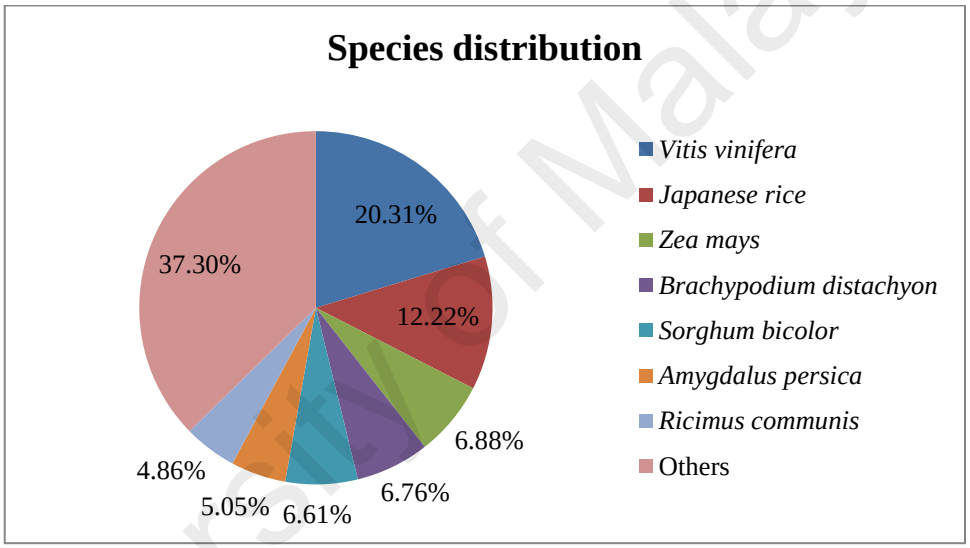


Figure 4.28, continued.

(a) The E-value distribution of the result of Nr annotation. (b) The similarity in the result distribution of Nr annotation. (c) The species distribution of the result of Nr annotation where 65,296 BLASTX-hit unigenes were calculated.

4.9.3 Functional classification by GO and COG

For evaluating the function of assembled unigenes, the annotated unigenes were mapped using Gene Ontology (GO) database, which offered an updated controlled vocabulary to describe comprehensively the properties of genes and their products in any

organism. GO has three ontologies: molecular function, cellular component and biological process.

In total, 47,111 (46.1%) unigenes were searched for GO classes and distributed among the three categories with 16,493 unigenes assigned to cellular components (Figure 4.29a), 10,040 unigenes assigned to biological process (Figure 4.29b) and 7,451 unigenes assigned to molecular function (Figure 4.29c). Fewer unigenes were classified under molecular function, where those unigenes were mainly categorized into "binding" and "catalytic activity". In addition, one unigene may be assigned into several different GO-terms.

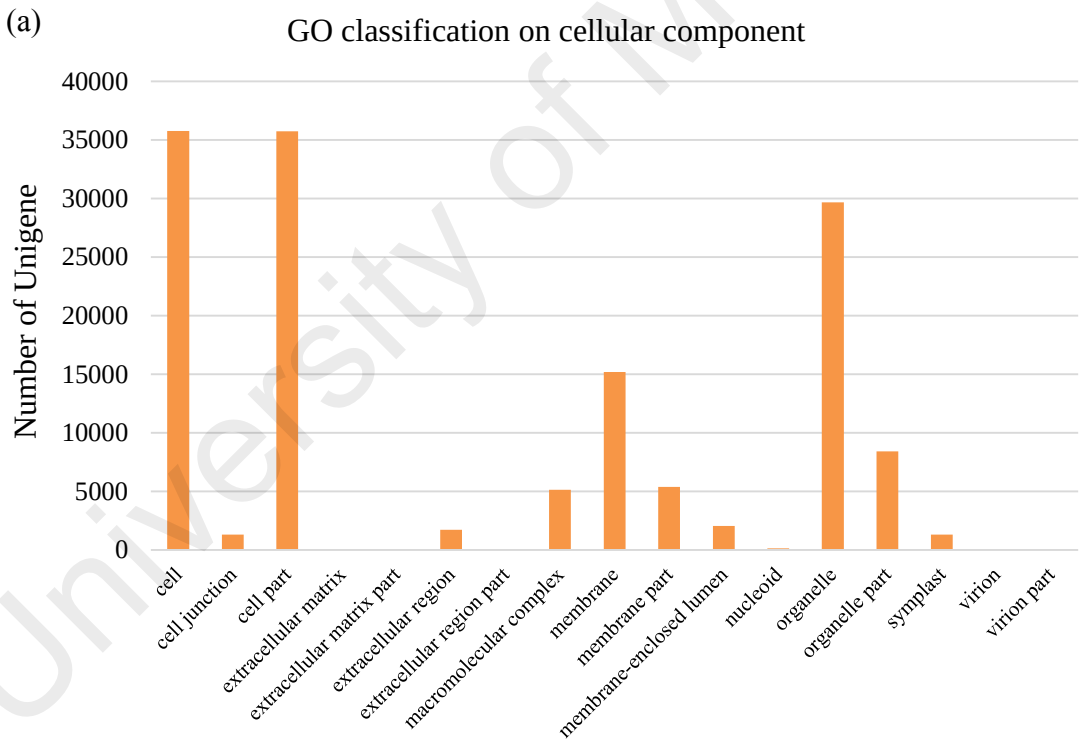
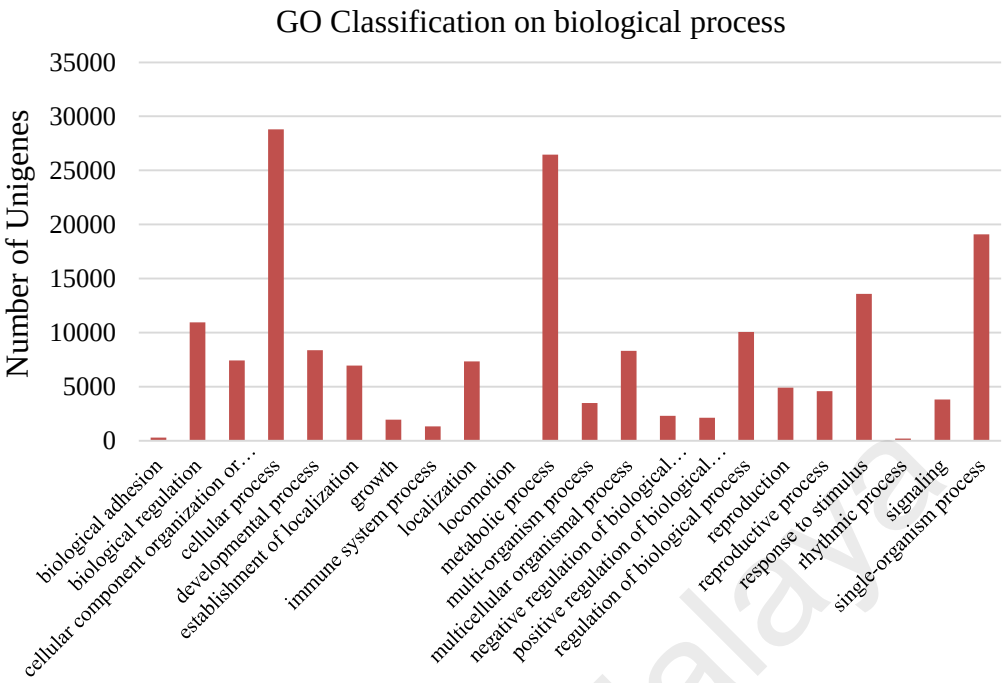


Figure 4.29: Histogram of GO classifications of assembled *B. rotunda* unigenes

(b)



(c)

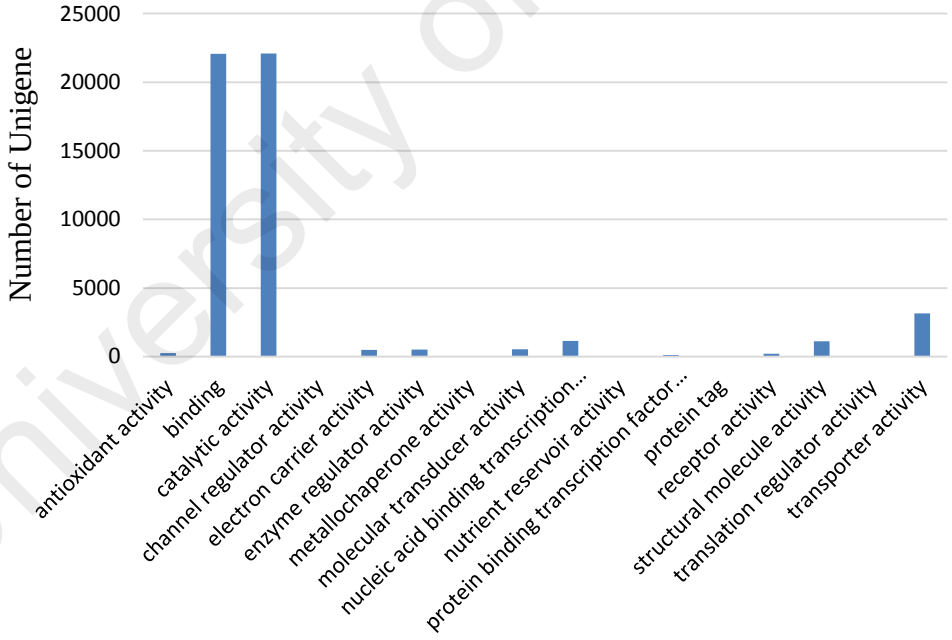


Figure 4.29, continued.

Number of unigenes annotated under a) cellular component, b) biological process and c) molecular function. GO function is shown in X-axis. The right Y-axis shows the number of genes, which have the GO function, and the left Y-axis shows the percentage of unigenes.

Next, further classification of orthologous gene product was evaluated by Clusters of Orthologous Groups (COG) database. In this COG database, every protein was expected to be evolved from ancestral protein. This database was constructed based on complete coding proteins together with their complete genome as well as based on evolutionary relationships of bacteria, algae and eukaryotes. All unigenes were mapped and aligned according to the COG database to predict their potential functions.

In total, 27,330 (26.7%) unigenes were successfully assigned to 25 COG classifications, leading to a total of 62,306 unigenes assigned in all categories. Among the 25 COG categories, most of the unigenes were represented under the largest group of the cluster for General function prediction (9,426, 15.1%) followed by transcription (7,022, 11%), replication, recombination and repair (5,600, 9%) and signal transduction mechanisms (4,773, 8%) (Figure 4.30).

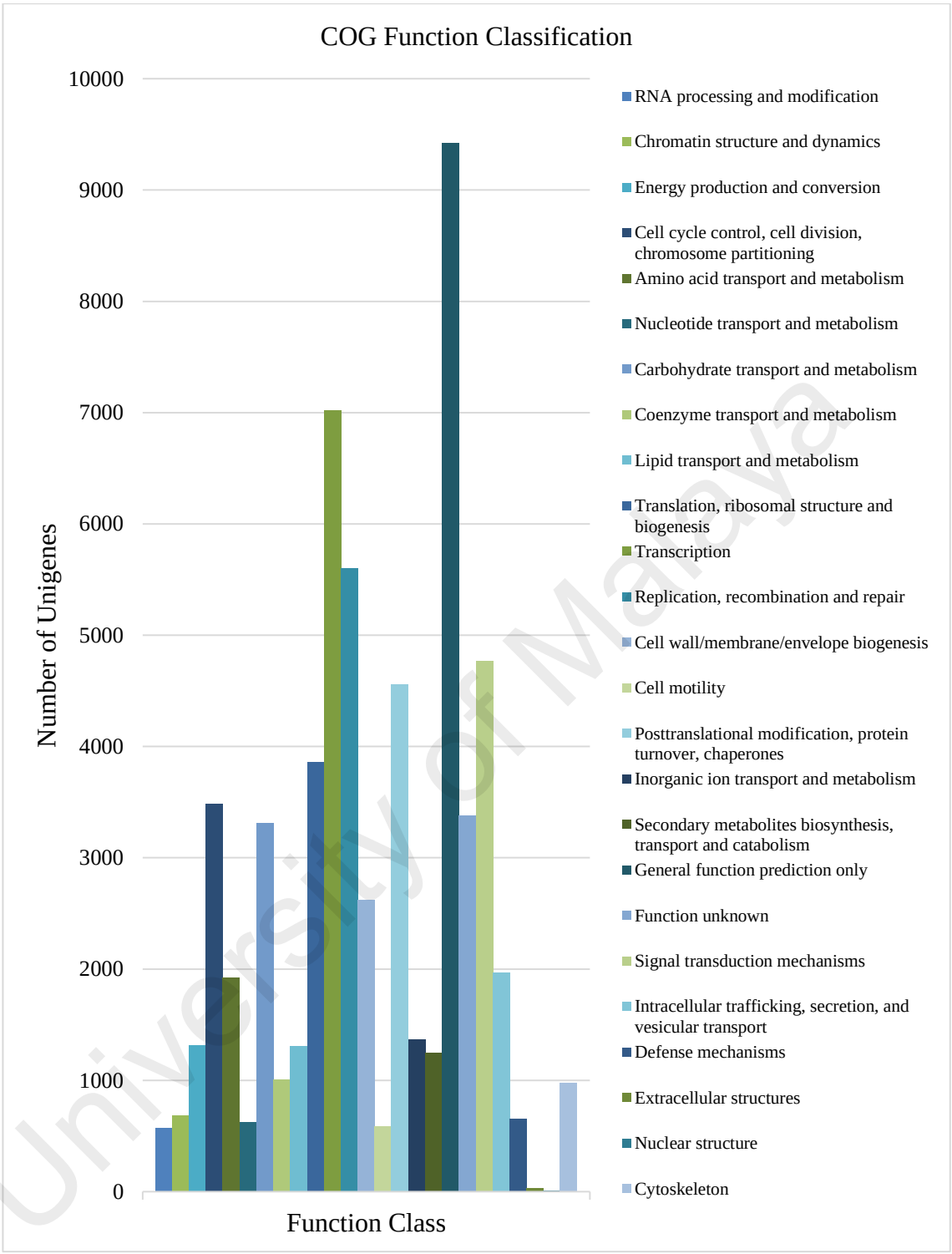


Figure 4.30: Histogram of COG function classification of *B. rotunda* unigenes

4.10 Analysis of differentially expressed (DE) genes

The gene expression levels were conducted using Fragment per kb per Million Reads (FPKM) method to identify the up-regulated and down-regulated genes in the wild-type and transgenic suspension cultures. Differential expression analysis was calculated using Poisson distribution equation, having the threshold of FDR ≤ 0.001 and value of $|\log_2 \text{Ratio}| \geq 1$. In this study, we sequenced triplicates of wild-type and transgenic suspension cultures and classified them into three groups: Group 1 (WT-R1 vs TL-R1), Group 2 (WT-R2 vs TL-R2) and Group 3 (WT-R3 vs TL-R3). Figure 4.31 shows the distribution of DE genes in these three groups.

(a)

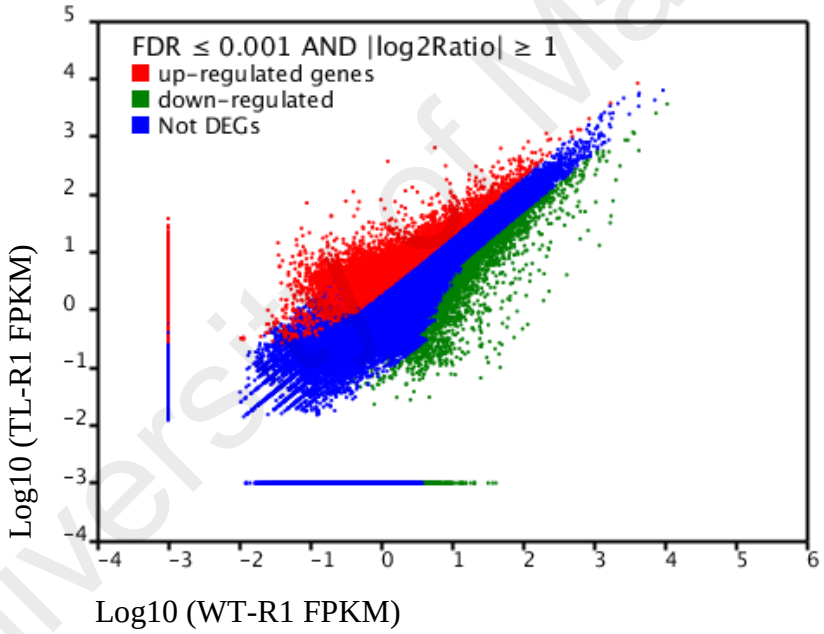
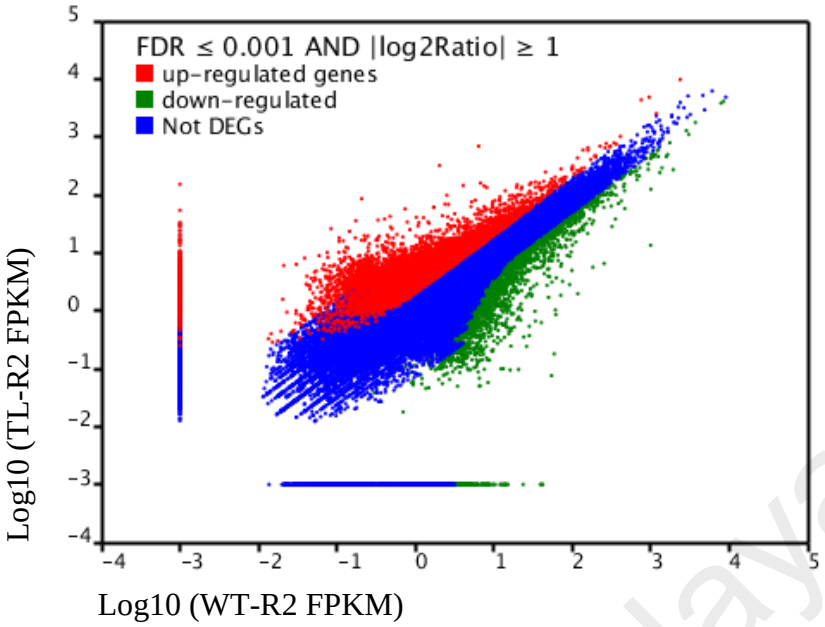


Figure 4.31: The scatter of expressed genes in wild-type (WT) and transgenic (TL) suspension cultures

(b)



(c)

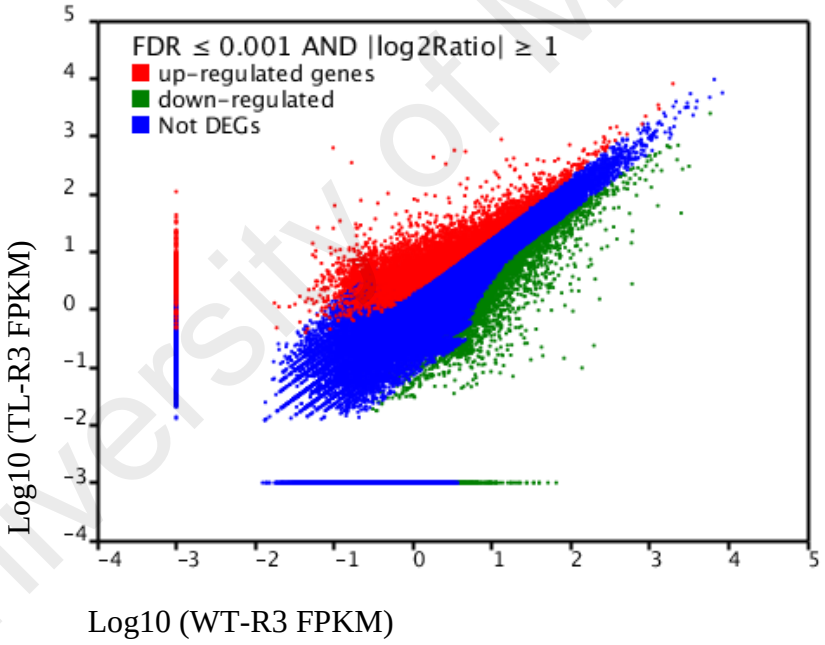
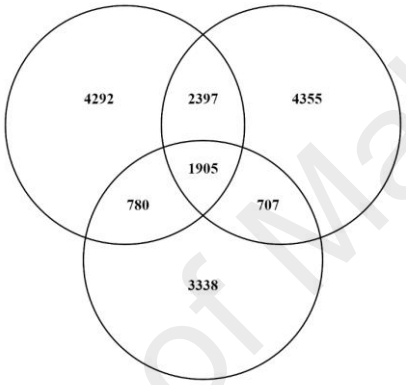


Figure 4.31, continued.

Distribution of up-regulated and down-regulated genes was denoted by red and green spots, respectively, while undifferentiated expressed genes were denoted as blue spots. The horizontal coordinates were the expression level of wild-type samples, and the vertical coordinates were the expression level of transgenic samples as represented in a) Group 1: WT-R1 vs TL-R1, b) Group 2: WT-R2 vs TL-R2, and c) Group 3: WT-R3 vs TL-R3.

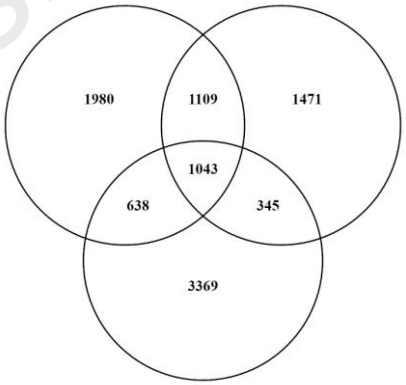
Overall, an inter-sectioned among the three groups showed 3148 unigenes which were significantly expressed, consisted of 1905 genes that were up-regulated and 1043 genes were markedly down-regulated, sharing among them (Figure 4.32). DE genes associated with biosynthesis of secondary metabolites were prepared to gain a better understanding of differential expression pattern. Consequently, we sought to analyse the DE genes to identify candidate genes involved in flavonoid biosynthesis.

(a) WT-R1 vs TL-R1 WT-R2 vs TL-R2



WT-R3 vs TL-R3

(b) WT-R1 vs TL-R1 WT-R2 vs TL-R2



WT-R3 vs TL-R3

Figure 4.32: Venn diagram analysis on DE genes in three groups

An inter-sectioned of WT vs TL showed the presence of (a) 1905 up-regulated genes and (b) 1043 down-regulated genes shared by the three groups.

4.11 Differentially expressed (DE) gene in the flavonoid biosynthetic pathway

Looking at the distribution of differentially expressed (DE) genes, the overall picture showed that there were relatively higher numbers of upregulated genes in the transgenic suspension cultures compared to wild type as the result of *BrCHS* overexpression. To further analyse the possible function of unigenes showing differential expression between the two samples (WT vs TL) in the three groups, we assessed their KEGG analysis. After filtering the DE genes according to the parameter stated in previously mentioned method (refer to Section 3.14.1), an inter-sectioned between three groups of WT vs TL demonstrated 305 up-regulated genes and 181 down-regulated genes, functionally annotated against KEGG database. This work further focused on elucidating potential unigenes being regulated in flavonoid biosynthesis pathways and related plant pathways (flavone and flavonol, isoflavonoid pathway) as highlighted in the Table 4.8.

Comparative analysis showed that in flavonoid biosynthesis pathway, there were 5 up-regulated unigenes and the rest were down-regulated unigenes where most of them coded for downstream enzymes as shown in Table 4.8. The similar findings also found in its related pathways that showed most of the unigenes were more down-regulated. CL3398.Contig3_All was the recombinant *BrCHS* and found to be up-regulated at ranging from 5.16 to 6.30 fold higher compared to wild-type. In the present study, the levels of pinocembrin chalcone in TL suspension cultures and this data supports the correlation between the biosynthesis of active compounds and its related gene expression. The amino acid sequence alignments between CL3398.Contig3_All and *BrCHS2* transcript variant is listed in Appendix K.

4.11.1 Discovery of CYP450s co-ordinately expressed with key flavonoid biosynthetic genes and DE genes in plant pathogen pathway

The discovery of cytochrome P450s (CYP450s) are co-ordinately expressed with variety of biosynthetic genes which catalyse monooxygenation/hydroxylation reactions involved at the molecular level of biochemical pathways (Nelson et al. 2004). According to the results, a total of 10 unigenes encoding CYP450s were related to flavonoid biosynthesis namely CYP71, CYP75 and CYP99 family. Most of these transcripts were down-regulated and played an important role in flavonoid biosynthesis.

In this study, the focus was on CYP71 and CYP75 family (CYP71A1, CYP71D9, CYP75A1) as this family involves the flavonoid biosynthesis. CYP99A3 participated in momilactone phytoalexins biosynthesis that acts as a multifunctional diterpene oxidase as reported in *oryza sativa* (Wang et al. 2011). CYP71 and CYP75 are the two transcripts encode flavonoid 6-hydroxylase (F6H) and flavonoid 3′,5′- hydroxylase (F3'5'H), respectively. They belongs to the cytochrome P450 monooxygenase system, is the only enzyme that catalyses the 5′- position hydroxylation of the B-ring.

Apart from flavonoid and its related downstream pathways, overexpression of BrCHSs caused changes in the expression level of genes involved in plant pathogen interaction. During stress conditions, a number of disease resistance genes were expressed as part of its plant defense, besides being part of the plant developmental program. Among these genes, *CHS* was induced in plants under different form of stresses such as UV, wounding and in the presence of herbivory or microbial pathogens (Dao et al. 2010).

There were 69 up-regulated genes and 29 down-regulated genes together with several transcription factors as listed in Table 4.9. This finding proved that CHS can be induced during environmental stress and result in the increasing amount of defense metabolites. Based on the dataset, disease resistance protein RPM1, RPS2, WRKY transcription

factors, other proteins such as Interleukins and Calcium-binding protein CML were identified to be up-regulated.

Table 4.8: Summary of significant DE genes in the biosynthesis of flavonoids

Downstream pathway	Enzyme name	Up/Down	Unigene ID
Flavonoid	Shikimate O-hydroxycinnamoyltransferase (HCT)	Up Down	CL4317.Contig3_All Unigene7406_All
	6'-deoxychalcone synthase (DOCS)	Up	CL9892.Contig2_All
	Chalcone synthase (CHS)	Up Down Down Down Down	CL3398.Contig3_All CL3398.Contig1_All CL3398.Contig2_All CL7278.Contig3_All Unigene8648_All
	Chalcone isomerase (CHI)	Up Down	Unigene27073_All Unigene28503_Al
	Isoflavone reductase (IFR)	Down	CL263.Contig1_All
	Flavonol synthase (FLS)	Down	Unigene31817_All
	Naringenin 3-dioxygenase (F3H)	Down	Unigene5234_All
	Cytochrome P450: CYP75A1 (flavonoid 3',5'-hydroxylase) CYP71A1 (unknown) CYP99A3 (9beta-pimara-7,15-diene oxidase)	 Down Down Up	 CL1414.Contig2_All Unigene21075_All CL5989.Contig2_All
	trans-cinnamate 4-monooxygenase	Down Down	Unigene21841_All CL12450.Contig1_All
Flavone and Flavonol	Flavonol 3-O-methyltransferase	Down Down	Unigene32384_All CL12706.Contig3_All
	UDP-glucosyl transferase 73C	Down Down Down	Unigene22825_All Unigene28219_All Unigene19067_All
	trans-resveratrol di-O-methyltransferase	Up	CL937.Contig2_All

Table 4.8, continued.

	Cytochrome P450: CYP99A3(9beta-pimara-7,15-diene oxidase) CYP75A1 (flavonoid 3',5'-hydroxylase)	 Up Down Down	 CL5989.Contig2_All Unigene21075_All CL1414.Contig2_All
Isoflavonoid	2-hydroxyisoflavanone dehydratase	Down	CL8947.Contig1_All
	Cytochrome P450: CYP99A3 (9beta-pimara-7,15-diene oxidase)	 Down Down Up	 CL1010.Contig2_All CL1010.Contig4_Al CL5989.Contig2_All
	CYP71D9 (flavonoid 6-hydroxylase)	Down	CL5075.Contig2_All

Table 4.9: List of DE genes presented in plant pathogen interaction

DE gene	Up	Unigene ID	Down	Unigene ID
Disease resistance protein RPM1	6	CL5287.Contig3_All Unigene14207_All Unigene7777_All Unigene17034_All CL2941.Contig5_All Unigene34015_All	5	CL5287.Contig6_All Unigene16345_All Unigene21008_All CL4287.Contig2_All Unigene22335_All
Disease resistance protein RPS2	4	CL3143.Contig3_All CL709.Contig2_All CL709.Contig7_All Unigene51825_All	-	-
DNA-directed RNA polymerases I, II, and III subunit RPABC1	7	Unigene16450_All Unigene16514_All Unigene17646_All Unigene14145_All Unigene35720_All Unigene4066_All CL13309.Contig6_All	-	-
Interleukin-1 receptor-associated kinase 4	7	CL10548.Contig2_All CL2833.Contig3_All Unigene221_All Unigene31072_All Unigene25328_All Unigene611_All	7	Unigene4773_All Unigene8806_All Unigene6530_All Unigene22535_All Unigene19040_All CL10627.Contig2_All CL2617.Contig4_All

Table 4.9, continued.

Calcium-binding protein CML	5	CL10585.Contig1_All CL11807.Contig2_All CL8005.Contig2_All CL8005.Contig4_All CL11677.Contig2_All	3	CL9800.Contig2_All CL9312.Contig2_All Unigene33440_All
WRKY transcription factor 33	5	CL136.Contig4_All CL5017.Contig1_All Unigene14319_All Unigene27331_All Unigene35147_All	2	Unigene1079_All Unigene6579_All
WRKY transcription factor 22	2	Unigene6251_All CL7243.Contig3_All	-	-
WRKY transcription factor 25	1	Unigene42849_All	-	-
Phytochrome-interacting factor 3	2	CL3941.Contig12_All CL7729.Contig1_All	-	-
Mitogen-activated protein kinase	2	CL9812.Contig3_All Unigene34416_All	1	Unigene15104_All
brassinosteroid insensitive 1-associated receptor kinase 1	4	Unigene219_All CL3824.Contig3_All Unigene966_All CL13600.Contig2_All	-	-
Serine/threonine-protein kinase PBS1	1	CL6693.Contig1_All	1	CL1839.Contig1_All
Calcium-dependent protein kinase	1	Unigene22404_All	1	CL9177.Contig3_All
Hypothetical protein kinase	2	Unigene25246_All CL12543.Contig2_All	-	-
Aprataxin	-	-	1	Unigene35303_All

Table 4.9, continued.

Guanine nucleotide-exchange factor	1	Unigene34874_All	-	-
Jasmonate ZIM domain-containing protein	3	Unigene6904_All CL1493.Contig1_All CL3795.Contig1_All	1	CL10950.Contig1_All
aarF domain-containing kinase	-	-	1	CL2781.Contig3_All
ubiquitin-conjugating enzyme E2 H	1	Unigene22822_All	-	-
Extracellular signal-regulated kinase 1/2	2	Unigene39056_All CL288.Contig2_All	-	-
Calmodulin	2	CL10917.Contig3_All CL6279.Contig2_All	1	CL2045.Contig6_All
somatic embryogenesis receptor kinase 1	-	-	1	CL4768.Contig2_All
Leucine-rich repeat (LRR) protein	3	CL6159.Contig1_All CL6159.Contig2_All CL251.Contig2_All	-	-
LRR receptor-like serine/threonine-protein kinase FLS2	7	CL1177.Contig19_All CL12505.Contig5_All CL1266.Contig2_All CL1266.Contig5_All CL838.Contig3_All CL9006.Contig3_All Unigene38460_All	3	Unigene8684_All Unigene31154_All Unigene40018_All
Transcription factor MYC2	2	CL9569.Contig2_All Unigene18886_All	1	CL12800.Contig1_All
Total	**69**		**29**	

4.12 Experimental validation

For validation of unigenes presented in RNA-Seq database, five unigenes were randomly selected and showed general agreement with their transcript expression as determined by RNA-Seq in all three comparison groups which indicated the reliability of the transcriptome profiling data (Figure 4.33). Unigene 28275_All (avg. fold-change value: 1.742) and CL4317.Contig3_All (avg. fold-change value: 1.738) represent up-regulated DE genes while CL1414.Contig2_All (avg. fold-change value: -2.530), Unigene 5234_All (avg. fold-change value: -3.062) and CL12450.Contig1_All (avg. fold-change value: -4.070) represented down-regulated DE genes in transcriptome data. Despite these unigenes showed some discrepancies values in fold change for up- and down-regulation in all comparison groups, however, when compared to RNA-Seq results, the same expression patterns were observed in all selected unigenes. Different platforms (RT-qPCR and RNA-Seq) use different formulas for fold change calculation causing the discrepancy of fold-change values.

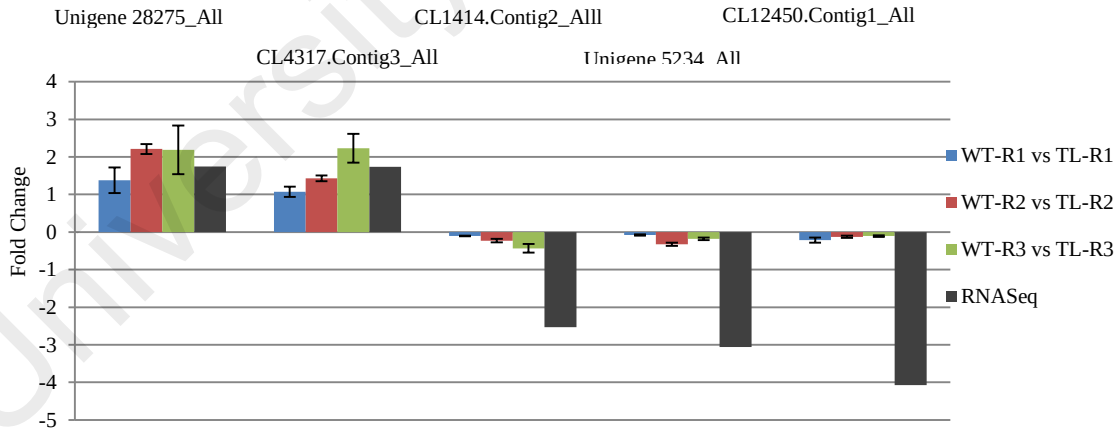


Figure 4.33: Expression pattern validation of selected unigenes by RT-qPCR

Expression pattern validation of selected unigenes by RT-qPCR in three groups is shown. Fold-change in mRNA transcript levels of five randomly selected unigenes represented by blue, red and green bars. Black bar showed average fold-change value for each unigene that was calculated according to RNA-Seq analysis. X-axis showed fold changes in transcript abundance of unigenes. Standard error bar represents relative expression level determined by RT-qPCR using $2^{-\Delta\Delta CT}$ method of three experimental replicates

CHAPTER 5: DISCUSSION

Biosynthesis of the high-value of natural bioactive compounds has been an important target for metabolic engineering owing to their role in insect resistance, flower pigmentation, UV protection and the health beneficial properties it provides (Dao et al. 2011; Jenkins et al. 2001; Kirana et al. 2007; Li et al. 2016; Pattaratanawadee et al. 2006). Despite their potential, the availability of these compounds is limited due to various spatial and temporal cues. Continuous supply of bioactive compounds from *B. rotunda* plant is getting crucial because propagation by vegetative technique using a rhizome segment is time-consuming and susceptible to rhizome soft rot disease, leaf spot and soil-borne pathogens such as *Coleotrichum* species (Yusuf et al. 2013). This would results in non-economically viable towards market demand. Hence, in-vitro propagation method with morphogenetic potential and growth capacity from suspension cultures of *B. rotunda* was successfully developed to obtain optimized sources of bioactive compounds.

Previously, in-vitro propagation method with morphogenetic potential and growth capacity from suspension cultures of *B. rotunda* were explored as an alternative source of flavonoid compounds, namely alpinetin, pinocembrin, cardamonin, pinostrobin and panduratin A (Wong et al. 2013; Yusuf et al. 2013). With several advantages of utilizing suspension cultures as plant host to produce bioactive compounds; this further provides the idea of enhancing the accumulation of targeted compounds in suspension cultures through metabolic engineering. Therefore, this study was carried out with the aim of assessing the effect of CHS overexpression on the flavonoid biosynthesis and plant pathways in suspension cultures of *B. rotunda.*

5.1 Isolated BrCHS showed a typical CHS protein

In this study, the key enzyme of the flavonoid biosynthetic pathway, Chalcone synthase (CHS) was evaluated which directed the variety production of metabolites. In an effort to investigate the effect of the CHS overexpression implicated on the flavonoid biosynthesis and related plant pathways, the study was firstly designed to characterize more than 50 *BrCHS* clones from four different parts of *B. rotunda* plants. BrCHS sequence analysis and comparison with other plant CHS revealed that the open reading frame for BrCHS was 1176bp in length and putatively encoded a polypeptide of 392 amino acids.

BrCHS protein comprised of an active amino-acid residues that are highly conserved and reside at the second exon. Between the active site cavities, there are two conserved "gatekeeper" phenylalanine's (Phe-215 and Phe-265) connected with CoA-binding tunnel. The existence of these two "gatekeeper", facilitates decarboxylation of malonyl-CoA and involved in the condensation reactions of CHS (Abe et al. 2003; Jez et al. 2002).

According to Neighbour Joining tree, five *BrCHS*s were predominantly expressed and differently clustered in the phylogenetic analysis. This gene characterization constituted multiple copies, supported by a variety of copy number that has also been reported in several plant species. For example, a total of seven *CHS* have been isolated in a genomic library of *Sorghum bicolor* (Lo et al. 2002) and nine *CHS* were found in the genome of *Glycine max* (Matsumura et al. 2005) as well as in morning glories (*Ipomoea* spp.) (Durbin et al. 2000). Our result also corroborated with southern blot analysis where at least four to five clear bands were observed after treated with *Dra*I and *Eco*RV indicating that there were more than one copy of *BrCHS* which existed in the *B. rotunda* plant genome.

As mentioned by Harker et al. (1990), members of multigene family are referred to be differentially expressed in different ways, for example, during plant development or in response to environmental cues. Some of these multigene are constitutively expressed while others are transcriptionally induced by environmental factors such as elicitor, precursor, UV light and pathogen attack (Cai et al. 2011; Gueven & Knorr, 2011; Zhao et al. 2010).

Despite the identification of more than one *BrCHS* copies in *B. rotunda* plants, this occurrence might offer new possibility that all copies belong to multigene family and might elucidate evolutionary changes underlying *BrCHS* evolution in *B. rotunda*. Further evaluation using physical map such as molecular linkage mapping and in-depth analysis of phylogeny can be drawn to show the distribution of *BrCHS* on different chromosomes of *B. rotunda* (Matsumura et al. 2005).

Since there were multiple copies of *BrCHS*s detected in *B. rotunda* plants, this study found that homology analysis among five representatives of *BrCHS*s revealed high degree of similarity between 97.0% to 99.1% and 94.3% to 99.0% at nucleotide and amino acid level, respectively (refer to Table 4.1). This finding was supported by earlier studies, for example, 92.3% to 100% of *CHS* sequence similarity was reported in *Triticum aestivum* (Trojan et al. 2014) and *Gerbera hybrida* which shared 80.0% to 88.0% of CHS amino acid sequence identity (Deng et al. 2014). Meanwhile, there were eight copies of *CHS* members detected in *Glycine max*; however, only *CHS7* and *CHS8* had contributed to most of *CHS* transcripts and shared 97.0% similarity at the nucleotide level (Tuteja et al. 2004).

The occurrence of high similarity among BrCHS members indicates a strong conservation of the structure and active site residues that are important during the synthesis of plant metabolites. Additionally, BrCHS has exhibited high similarity in

sequences (84%–90%) with CHSs from other plant species, also suggesting that the CHS family is conserved during the evolutionary process and thus being a typical CHS protein.

5.2 *BrCHS* expressed specifically in four parts of *B. rotunda* plant

*BrCHS*s have formed separate clusters and showed tissue-specific expression pattern, designated as *BrCHS1 (isolated from leaves), BrCHS2 (isolated from rhizomes), BrCHS3 and BrCHS4 (isolated from suspension cultures)* and *BrCHS5 (isolated from flowers).* Even though there was no confirmation on *BrCHS* expression level in every parts of *B. rotunda* plants using RT-qPCR approach, our results were in accordance with previous study. For example, spatial regulation of the *CHS* in different tissues was previously detected especially in the roots of *Polygonum minus* ranging approximately from 10 to 15-fold higher compared in the leaves and stems (Roslan, et al. 2013). High level of gene expression in roots produced wide range of key signalling compounds associated with plant-microbe interactions into the root-soil interphase (Steinkellner et al. 2007).

A possible explanation for tissue-specific expression may be due to differences in the structural sequence diversity of the cis-acting element (at 5' untranslated region) within all five BrCHS promoters which controlled the gene regulation at the promoter level. Previous study reported that the *CHS* promoter has a complex series of regulator cis-acting elements localized upstream of the transcriptional start (Hernandez-Garcia & Finer, 2014). According to report by Faktor and colleagues (1997), a 39 bp motif was identified which described the two adjacent motifs, the G-Box (CACGTG) and the H-Box (CCTACC) that were often found near a TATA Box in the promoter of *CHS15* in *Phaseolus vulgaris*. The role of these motifs was for specific expression in flowers and roots of *Phaseolus vulgaris* and their role was described as tissue-specific regulatory element.

Thus, it might be possible that the expression patterns among five *BrCHS*s are due to the sequence variation and regulatory position that have been mapped between the two adjacent motifs or possessed a complex structure of the motif(s) in the promoter region. Hence, taking *BrCHS2* transcript variant as an example, the complex interplay between different cis-elements in *BrCHS2* promoter may be crucial for its expression specifically in *B. rotunda's* rhizome compared to other parts of the plant.

Tuteja et al. (2004) presumed that sequence diversity between *CHS* members in different tissues is more closely associated with specialization of tissue-specific expression rather than to enzymatic specific function. However, one of the limitations with their presumption was that there was no validation of enzymatic reaction provided in order to claim that the diversification of CHS expression profiles were not associated with CHS enzymatic function. The CHS diversity study would have been more interesting if the catalytic properties of each of the corresponding enzymes of the studied genes (*CHS1-CHS8*) that were found active in four tissues, namely roots, young leaves, cotyledons, and seed coats in the *Glycine max* could be included. Nevertheless, in this study, enzymatic functional activity was developed for *BrCHS2* gene that is predominantly expressed in rhizomes as described in the next subtopic.

Thus, to develop a full picture of *BrCHS* expression pattern and differentiate each of these *BrCHS* transcript variants, additional studies on tissue specific expression pattern and functional dissection of *B. rotunda's CHS* promoter are needed to delineate the function of each transcript variant. Our study on differentiating CHS expression profiles in specific tissue of *B. rotunda* plants cannot be achieved since there is a high sequence similarity among the coding region of *BrCHS* transcript variants. Therefore, it is imperative to obtain quantitative data using a more sensitive and powerful technique by

designing gene specific primers and probes for TaqMan RT-qPCR or by evaluating the *BrCHS* sequence variation at the intergenic region (5' untranslated region).

5.3 BrCHS2 shows its function by producing targeted metabolite compound

Previous study has successfully analysed at proteome level under the addition of elicitor such as phenylalanine to identify chalcone derivatives involved in flavonoid and phenylpropanoid biosynthesis pathways in suspension cultures of *B. rotunda* (Tan et al. 2012). Besides the idea of adding elicitor as one of the efficient treatment to induce *CHS* expression, overexpression of *CHS* is one of the ways to synthesize targeted compound in plants. Therefore, to stimulate higher accumulation of the targeted compound, this present study was designed by adding more copies of *BrCHS* into suspension cultures.

Since genes related to protein modification appeared to be expressed at higher levels in the rhizome than in the leaf or root for both turmeric and ginger (Koo et al. 2013) and associated with medicinal interest (Trakoontivakorn et al. 2001; Kirana et al. 2007; Shindo et al. 2006), with these references, therefore, *BrCHS2* gene (isolated from rhizomes) was the best choice for overexpression study. This could help to explore its role as one of the functional gene in the rhizomes of *B. rotunda* plants.

The functional diversity of CHS-superfamily type III PKSs is caused by the utilization of different CoA-linked starter substrates ranging from aliphatic-CoA to aromatic-CoA, from small acetyl-CoA to bulky N-methylanthraniloyl-CoA substrates, and from polar malonyl-CoA to nonpolar n-hexanoyl-CoA substrates (Abe & Morita, 2010). Since *BrCHS*s exhibited variable expression patterns, it can be suggested that each of the gene might catalyse similar and/or different substrates in the same plant biochemical pathway.

In this study, the biosynthesis of subsequent bioactive compound such as pinocembrin was expected to be produced via reaction of a starter substrate; Cinnamoyl-CoA. This

starter substrate was chosen by referring to the previously proposed pathways that showed a series of deamination, isomerization and cyclisation of pinocembrin chalcone to pinocembrin which was one of the favourable and expected bioactive compound to be detected (Hwang et al. 2003; Jiang et al. 2005).

Pinocembrin chalcone acted as an intermediate precursor for subsequent flavonoids formation. It was detected in all transgenic lines along with significantly variable amount after the cell pellet was streamed by nitrogen gas and extracted with ethyl acetate (refer to Table 4.4). In this study, by comparing transgenic lines overexpressing *BrCHS2* genes with wild-type as a control has shown that the accumulation of the pinocembrin chalcone (C1) was significantly higher (ranging from 112.20 mg/g to 275.63 mg/g) whereas only 64.67 mg/g of pinocembrin chalcone accumulated in wild-type.

The yield of secondary metabolites via suspension cultures remains a bottleneck due to unstable and unpredictable variability in accumulation. Trace amount of pinocembrin was detected in all suspension cultures. Trakoontivakorn et al. (2001) claimed that pinocembrin chalcone can be easily converted to pinocembrin in methanol solution. But for this study, even though methanol solution was used as a solvent to extract flavonoid, separation of pinocembrin compound was still undetectable using HPLC analysis. Despite maintaining their transgenic character, very low levels of pinostrobin and pinostrobin chalcone together with undetectable naringenin and naringenin chalcone were also reported in the suspension cultures.

A possible explanation might be due to long-term subculture that consequently affect the age of the cell lines. Since the age of the *B. rotunda* suspension cultures were initially maintained and sub cultured for more than one year, these undifferentiated cultures might result in partially or totally loss of production capacity. As reported by Qu et al. (2005)

four cell line suspensions of *Vitis viniferea* L. showed a gradual loss of anthocyanin biosynthetic capacity during weekly subculturing over 8 months.

Similar result was also observed in the cell lines of *Catharanthus roseus*, in which it gradually lost the ability to produce terpenoid indole alkaloids (TIAs) over 30 months period with significantly decreased activity of transgene-encoded enzymes of the TIA pathway that were initially capable of accumulating relatively large amounts of alkaloids (Whitmer et al. 2003). Thus, extended culturing could induce instability of metabolite accumulation, which results in decrease productivity and changes in chemical components in the producing cell lines.

The presence of pinocembrin was nearly undetectable, suggesting that its intermediate compound; pinocembrin chalcone was isomerized and cyclised to the other favourable compounds; C3 and C4 (unidentified compounds, refer to Figure 4.27). These unpredictable compounds might also be derived from by-products after each deamination series of Cinnamoyl-CoA with the enzyme. For compound C3 and C4, their retention time did not match the standards of the six compound retention time and cannot be analysed using mass spectrophotometry. Abundant of samples with high concentration were needed in order to determine the mass spectrum value for both compounds.

Besides being involved in a specific pathway, these unpredictable compounds might also derived from different series of complex pathway that is driven by more than one gene (Kumar & Gupta, 2008). As an example, an attempt has been made to manipulate a first committed enzyme in carotenoid biosynthesis; phytoene synthase (Ye et al. 2000). Phytoene production was increased but none of carotenoid content was accumulated in rice endosperm. However, when *lycopene β-cyclase* gene (*LCY-B*) and *Erwinia phytoene desaturase* gene (*crtI*) were added in the entire pathway of carotenoid, the rice accumulated β-carotene.

Thus, this occurrence explained that complex and redundant biosynthetic pathways coupled with incomplete knowledge of their regulation which could lead to the production of unpredictable compounds from a targeted pathway obtained via metabolic engineering strategy. To note, none of the potential compounds such as alpinetin, cardamonin, panduratin A and 4- hydroxypanduratin A, have been identified due to the lack of suitable standards. The presence of panduratin A and 4-hydroxypanduratin A in *B. rotunda* were previously reported to be very low and the synthesis was likely to involve many proteins and complex biosynthesis pathways (Yadnya-Putra et al. 2014; Yusuf et al. 2013).

The present study has also raised the possibility that variation in the expression level of *BrCHS* correlated with the variable amount of pinocembrin chalcone (C1) in TL3 and TL9 suspension cultures. Based on the results, both, TL3 and TL9 have shown different overexpression effects with 8.1- and 27.5-fold change higher of *BrCHS* expression, respectively, compared to wild-type. The massive expression in TL9 has yielded 207.30 mg/g of C1, that was lower compared to TL3 (275.63 mg/g of C1) even though TL3 has three times lower amount of transcript level compared to TL9.

Different level in transgene expression might be influenced by the surrounding DNA at the site of transgene integration (Kole et al. 2010). An integrated transgene is controlled under the influence of local regulatory element such as enhancer that interacts with regulatory elements in the transformation construct. As reported in *Arabidopsis*, the T-DNA was examined to preferentially integrate "between" genes, in 5'-gene regulatory regions, polyadenylation site regions, A+T rich regions (Gelvin, 1998) and gene-rich regions; telomeric and subtelomeric regions (Kole et al. 2010) to reflect the final expression level. Different expression level affects the yield of BrCHS enzyme and subsequently the amount of metabolites produced in the samples.

Previously, transformation of gene into monocotyledonous plant was not amenable with *A. tumefaciens* compared to dicotyledonous plants that were well-established (Sood et al. 2011). Transformation of monocots plants using *Agrobacterium*-mediated method gained popularity when a large number of transgenic monocots were successfully produced and the function of the inserted gene being extensively studied (Éva et al. 2008; Hiei et al. 2014; Reyes et al. 2010; Zhang et al. 2013). Based on the results on positive GUS coloration (refer to Figure 4.21), a stable expression of the recombinant *BrCHS2* gene with the fused reporter gene, *mgfp5::gusA* was successfully established and proved that the recombinant *BrCHS2* gene was stably integrated into the *B. rotunda* plant genome.

Therefore, with the successful and efficient method of using *Agrobacterium* as a way for gene transfer in several monocots plants, transferring gene in *B. rotunda* plant genome would not be a problem because of better understanding of the monocots transformation process and availability of the vectors for efficient gene transfer (Komori et al. 2008). *Agrobacterium*-mediated transformation has become a viable and useful tool for the genetic engineering in monocots *B. rotunda* which would later drive the synthesis of important metabolites compounds in the flavonoid biosynthesis pathway.

5.4 Importance of having good quality of RNA for RNA-Seq analysis

Transcriptomic analysis is aiming to elucidate potential genes and explore specialized pathways of secondary metabolites biosynthesis in non-model medicinal plants. This method allows low cost of performing sequence analysis for every nucleotide of the transcript. In this study, RNA-Seq platform was utilized to analyse gene regulatory network in suspension cultures overexpressed with *BrCHS2* gene. The accuracy and validity of RNA-Seq datasets were determined by employing and analysing six samples (three biological samples of transgenic lines; namely TLs and three biological samples of

wild-type lines; namely WTs). To achieve this, quality control at different stages of analysis was assessed and optimal methods for transcript quantification, normalization, and differential expression analysis was applied to ensure both reproducibility and reliability of the results.

On top of that, one of the key elements for successful analysis is using intact RNA. If the initial samples were partially or completely degraded, it might distort the amount of reads produced from the transcript. Minimal decrease in RNA quality due to thawing steps of frozen tissues could cause significant changes in gene expression patterns and this degradation can be assessed by using RNA Integrity Number (RIN) during sample screening (Schroeder et al. 2006). To note, all six RNA samples used in this study showed RIN values ranging from 7.5 to 9.5, indicated high integrity of RNA. These qualified samples have led to 90% and 95% successful rate of RNA-Seq analysis for each library preparation.

Secondly, assessment of sequencing platform accuracy is gaining importance to result in sequencing data quality and low costs for validation experiments. To achieve this, Phred quality score (Q score) was used to measure base calling accuracy (Ewing & Green, 1998). In this analysis, Q20 score represents the probability of a base call error and more than 90% (Q20) bases was obtained for each sample. With this value, failed reaction and low-quality reads that could interfere the assembly process can be directly identified and trimmed to reflect a high accuracy of sequencing run.

Thirdly, in determining how good the transcript assembly was, basic statistics such as the N50 was computed once the assembly process was complete. N50 is a way of calculating an average sequence length from a set of sequences, giving the length of the smallest contig that contains the fewest (largest) contigs whose combined length represents at least 50% of the assembly (Miller et al. 2010). Based on the lengths of the

assembled contigs, at least half of all the assembled bases in transcript contigs were of at least 1646 bases in length. The observed N50 value (1,646 bases) was significantly higher than those obtained in other plant transcriptome sequencing projects. Previous studies have reported N50 value of 485 bases in *Hevea brasiliensis* (Xia et al. 2011), 765 bases in *Ipomoea batatas* (Zhangying et al. 2010), 948 bases in *Acacia auriculiformis* (Wong et al. 2011), 1192 bases in *Cicer arietinum* L. (Garg et al. 2011) and 1515 bases in *Cucurma longa* (Annadurai et al. 2013). Another report in the same plant study, *B. rotunda*, showed N50 value of 804 bases (Md-Mustafa et al. 2014). Ideally, the higher the N50 value, the better the assembly is.

Another way to assess the quality of transcript assembly is by examining distribution of length of contigs. The presence dataset showed all assembled unigenes were more than 300bp in size. Among these assembled unigenes, 61,838 unigenes (60.85%) were in the length ranging from 300bp to 1000bp, 25,301 unigenes (24.89%) were in the length ranging from 1000bp to 2000bp and 14,536 unigenes (14.30%) were longer than 2000bp. In addition, nearly half of these unigenes were more than 1000bp in size and indicated that the integrity of unigenes was relatively good. Compared with previous RNA-Seq profile on suspension cultures of *B. rotunda*, the distribution of length of contigs in this study was much longer and was enough to be analysed (Md-Mustafa et al. 2014).

Ultimately, the goal of most RNA sequencing study is to estimate the variation in gene expression levels and to explore specialized pathways of secondary metabolites biosynthesis in non-model plants. Therefore, for this analysis, the effects of having good RNA quality on measurement of relative variation in gene expression for both samples were considered. According to the data report, it is confident that the data analysis which resulted in high quality of long reads that could be further utilized for downstream application and future references.

5.5 RNA-Seq profiling of CHS overexpressed suspension cultures of *B. rotunda*

A comprehensive coverage of the number of sequenced transcriptomes of selected plant species could elucidate the biosynthesis of targeted metabolites. RNA-Seq has been applied to sequence hundreds of close related species of non-model plants (Johnson et al. 2012; Rastogi et al. 2014; Swarbreck et al. 2011). Xiao et al. (2013) previously established a data-mining framework of 75 non-model plants by employing RNA-Seq to explore specialized pathways of metabolite biosynthesis.

To comprehensively examine the regulation of genes associated with flavonoids accumulation, RNA-Seq analysis of transgenic and wild-type suspension cultures were performed. High quality and integrity of total RNA was successfully extracted from suspension cultures to develop cDNA libraries. About 3 Gbp of clean reads and a total of 102,195 unigenes were generated, of which 67,251 (65.8%) were successfully annotated based on BLAST searches against public databases, suggesting the presence of their relatively conserved functions.

However, the remaining unigenes (34.2% of the assembled unigenes) failed to generate significant homology to existing genes. The absence of homology could be due to the absence or limited homologous sequences in the public databases and there was a large proportion (34.45%) of unigenes that were shorter than 500bp, some of which were too short to allow statistically meaningful matches. This elucidated that there were more potential unigenes which remained unannotated and might fill-in the gaps in the regulation networks among differentially regulated genes generated in this study.

A further analysis of the transcriptomic datasets of *B. rotunda* suspension cultures harbouring recombinant *BrCHS2* which have significantly higher levels of pinocembrin chalcone (C1) was further evaluated. In order to increase the sequencing depth, a paired-end library sequencing strategy was applied and all seven public databases were selected

for gene annotation comparisons to acquire comprehensive functional information. Multiple databases of three groups of WT vs TL, namely Group 1 (WT-R1 vs TL-R1), Group 2 (WT-R2 vs TL-R2) and Group 3 (WT-R3 vs TL-R3) were inter-sectioned to obtain a comprehensive classified and regulated transcript by referring to KEGG database.

In this study, differentially expressed (DE) genes were analysed and potential candidate genes that were responsible for the metabolites production were identified to clarify the molecular mechanism of flavonoids biosynthesis. As expected, the recombinant of *BrCHS2* (CL3398.Contig3_All) was found to be up-regulated in flavonoid biosynthesis pathway. *BrCHS2* transcript variant were directly involved in flavonoid biosynthesis (specifically pinocembrin compound) since it supports the significantly higher amount of pinocembrin chalcone than that of wild-type. However, endogenous *BrCHS*s were found to be down-regulated and this finding was supported by the previous prediction on the introduction of transgene that caused unintended modifications in transcription and translation of native genes in the host plants (Kanobe et al. 2013).

An inter-sectioned between three groups of wild-type (WT) and transgenic (TL) produced 3,148 differentially expressed genes, of which 1,905 were up-regulated and 1,043 were down-regulated in the suspension culture library. Looking at the distribution of differentially expressed genes, the overall picture showed that there were relatively higher numbers of upregulated genes in the transgenic suspension cultures compared to wild type as the result of *BrCHS* overexpression. However, further comparative analysis against the KEGG database highlighted the fact that for flavonoid biosynthesis and related pathways, there were actually more down-regulated genes. All of the reported down-regulated genes related to the flavonoid biosynthesis includes Isoflavone reductase

(IFR), Flavonol synthase (FLS), Flavonoid/Flavanone 3-hydroxylase (F3'H), flavonoid 3',5'-hydroxylase (F3'5'H) and flavonoid 6-hydroxylase (F6H).

Due to the abundance of down-regulated transcripts, overexpression of *BrCHS2* might direct and switched on the expression of other genes in the other additional branching pathways as proposed by previous researchers (Md-Mustafa et al. 2014; Yadnya-Putra et al. 2014). Reports on the suggestion and proposed plant pathway by Tan et al. (2015) showed possible branching pathways that could actually emerged after the production of pinocembrin chalcone. Presumably, it was predicted that the accumulation of higher pinocembrin chalcone could led to the production of other types of metabolites in the other plant pathway. This was supported by recent studies on transcriptomic analysis in phenylalanine-treated suspension cultures of *B. rotunda* which claimed that there was a possibility of new plant biosynthesis pathway to occur, branching from the pinocembrin chalcone (Md-Mustafa et al. 2014; Tan et al. 2012).

The production of these compounds from new branching pathway can be inferred since overexpression of *BrCHS*2 could give some insight on the unidentified compound detected during previous enzymatic analysis with Cinnamoyl-CoA (refer to Appendix L). However, enzymes directly involved in the unknown pathway(s) remained to be elucidated until a reference pathway is available.

Potential pathways involve in accumulating several bioactive compounds that look more like a complex metabolic grid than a linear pathway. The existence of yet uncharacterized plant biosynthesis pathways could be speculated to occur and may produce other novel bioactive compounds. This enable future effort to produce new and valuable flavonoids derived from pinocembrin chalcone compound.

5.6 Plant cytochrome family (CYP71 and CYP75) encoded the production of flavonoids

Cytochrome P450 monooxygenases (P450s) catalyse a wide variety of monooxygenation reactions in primary and secondary metabolism in plants that involved in catalysing the oxidation of organic substrates, participating in a variety of biochemical pathways and playing roles in plant physiology (Annadurai et al. 2013; Dixon & Pasinetti, 2010). The existence of cytochrome P450 monooxygenases (P450s)-related metabolons has been demonstrated in the phenylpropanoid, flavonoid, cyanogenic glucoside, and other biosynthetic pathways (Ralston & Yu, 2006). P450 enzymes are highly diverse and presence abundantly with the highest proliferation in plants as well as requirement of molecular oxygen and NADPH as cofactors for the modification of basic flavonoid skeletons (Nelson et al. 2004; Werck-Reichhart & Feyereisen, 2000).

Typical enzymes belonging to cytochrome P450-dependent hydroxylases; flavonoid 3′-hydroxylase (F3′H) and flavonoid 3′,5′-hydroxylase (F3′5′H), are responsible for the formation of hydroxylation pattern by introducing hydroxyl groups at the 3′- or the 3′- and 5′-position of the flavonoid B-ring, respectively (Seitz et al. 2007; Wang et al. 2014). The reaction involved by which the flavanone (naringenin) and dihydroflavonol (dihydrokaempferol) is regarded as the main substrates of F3′H and are converted into the 3′,4′-hydroxylated eriodictyol and dihydroquercetin, respectively. Meanwhile, both compounds (naringenin and dihydrokaempferol) are also regarded as main substrates of F3′5′H and are converted into 3′,4′,5′-hydroxylated pentahydroxyflavanone (dihydrotricetin) and dihydromyricetin, respectively. The figure below depicts the reaction between these two P450 enzymes.

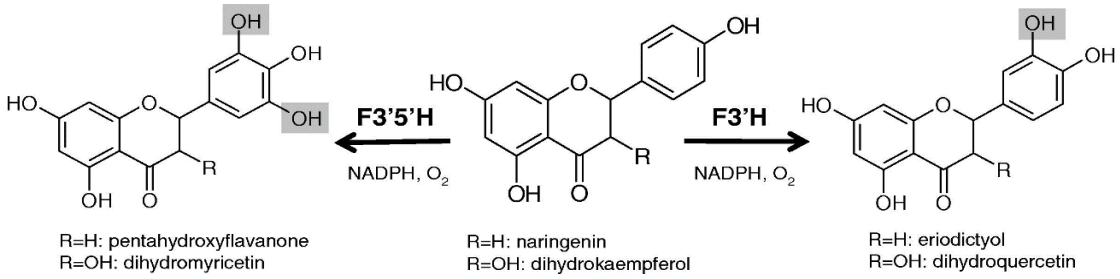


Figure 5.1: The reaction of two cytochrome (P450s)

Conversion of metabolites compounds from naringenin and dihydrokaempferol by the reaction of F3'5'H and F3'H to produce respective compounds.

In this study, two family of P450s have been identified; CYP71 and CYP75 and have been designated as CYP71: CYP71D9 and CYP71A subfamily as well as CYP75: CYP75A1 subfamily. All of these enzymes were down-regulated in their respective pathway. The CYP71 family comprises of CYP71A subfamily with 28 members including 17 annotations obtained from the *Arabidopsis thaliana* genome. The first member of this subfamily derived from ripening fruits of *Persea americana* (avocado) (Bozak et al. 1992). However, no specific enzymatic activity has been demonstrated for CYP71A subfamily and remains unknown.

Next subfamily of CYP71 is CYP71D which contains 22 members from 10 different plant species. CYP71D9 has been isolated from soybean (*Glycine max* L.) cells as a flavonoid 6-hydroxylase (F6H). A cDNA representing an elicitor induced P450 (CYP71D9) with flavonoid 6-hydroxylase (F6H) activity involved in the production of isoflavonoids with 6,7-dihydroxylation of the A-ring of flavonoid and codes for a protein with a novel hydroxylase activity (Latunde-dada et al. 2001).

Flavonoid 3′-hydroxylase (F3′H) was also identified in flavonoid biosynthesis pathway. However, it was not annotated as one of cytochrome P450 in the dataset. Based

on KEGG pathway, this enzyme controls the hydroxylation of dihydrokaempferol to dihydroquercetin and of naringenin to eriodictyol. Previously, a cDNA clone that corresponds to the *Ht1* locus of petunia was isolated and reported to control the same activity by the action of flavonoid 3′-hydroxylase (F3′H) (Brugliera et al. 1999).

Other than this family, CYP75 also contributes to the yield of flavonoid metabolites. The F3'5'H belongs to the CYP75 superfamily of P450 enzymes with vast majority of characterized F3′5′H that belongs to the CYP75A subfamily (Seitz et al. 2006). Previously, recombinant CYP75A encoding F3′5′H enzymes isolated from *Catharanthus roseus* (Kaltenbach et al. 1999) and *Solanum lycopersicum* (Olsen et al. 2010) showed broad substrate specificity in vitro, producing 3′,5′-hydroxylated flavonoids, including flavones.

The differential expression patterns of these cytochrome P450s in WT and TL indicate of its contribution to the biosynthesis of flavonoid. Further isolation of these three Cytochrome P450 enzymes; flavonoid 3′-hydroxylase (F3′H), flavonoid 3′,5′-hydroxylase (F3′5′H) and flavonoid 6-hydroxylase (F6H) could help in identifying more potential bioactive compounds as a result of this plant cytochrome reaction.

5.7 CHS function in plant resistance

Due to the overexpression of *BrCHS*s, unigenes that were responsible for plant defense mechanism have also increased as a response towards stress, and thus obtained important function on plant resistance. During stress conditions, a number of disease resistance genes are expressed as part of its plant defense, besides being part of the plant developmental program. *CHS* is also induced in plants under different form of stresses such as UV, wounding and in the presence of herbivory or microbial pathogens (Dao et al. 2011). This induction results in the production of compounds that have several activities of antimicrobial (phytoalexins), insecticidal and antioxidant and also compound

that quench UV light. In the present study, transcriptomic analysis has examples and pointed out that CHS-accumulating lines induced a set of defense related genes as well as transcription factors in the plant pathogen mechanism.

Disease resistance protein RPM1 and RPS2 have shown abundant expression in this particular pathway with up-regulated of DE genes for each of the protein. As reported in the *Arabidopsis*, RPM1 conferred resistance against *Pseudomonas syringae* expressing either the AvrRpm1 or the AvrB type III effector protein (Tornero et al. 2002). Meanwhile, RPS2 specifically recognizes *Pseudomonas syringae* pv. tomato strains expressing the *avrRpt2* gene and initiates defense responses to bacteria carrying avrRpt2 in *Arabidopsis* plant (Axtell & Staskawicz, 2003). Thus, in response to microbial/pathogen attack, antimicrobial such as phytoalexins are produced and derived from different classes of metabolites such as flavonoids, isoflavonoids, stilbenoids, sesquiterpenoids, steroids and alkaloids (Angelova et al. 2010; Cho & Lee, 2015; Ibraheem et al. 2010; Li et al. 2015; Malacarne et al. 2011; Sobolev et al. 2011).

In addition, there has been an interest in finding defense metabolites that could potentially serve as potential drug candidate for inhibition of pathogenic activities particularly in Zingiberaceae species such as *Zingiber officinale* Rosc., *Curcuma longa* Linn., *Alpinia galangal* Stuntz., and *Boesenbergia pandurata* Schltr. (Pattaratanawadee et al. 2006). According to their results, ethanolic extract from *Boesenbergia pandurata* Schltr. showed stronger inhibitory activity against *L. monocytogenes*, *B. cereus*, and *S. aureus* with MIC value of 0.2–0.4% (v/v), while having MIC value of 8–10% (v/v) against *E. coli* and *S. Typhimurium*. Thus, for future references, with the advancement of RNA-Seq technology, a more comprehensive research on plant infection towards different types of pathogens could reveal candidate

of defense genes involved in various mechanisms of action against pathogens attack in *B. rotunda*.

To sum up, RNA-Seq technology is suitable for gene expression profiling in non-model organisms that lack genomic sequence. Potential genes responsible for flavonoid biosynthesis coupled with identification of candidate cytochrome P450s and defense related genes were mined from transcriptomic data and found to be differentially expressed and involved in flavonoid accumulation. It is confident that the transcriptome database is a valuable addition to the publicly available *B. rotunda* genomic information. Our work generated a large set of cDNA sequences that represent the flavonoid biosynthesis pathways of known plant genes and potential new genes in the genus *Boesenbergia*.

CHAPTER 6: SIGNIFICANCE OF THE STUDY

The findings of this study will redound to the benefit of society considering that fingerroot is one of underrated but essential herb as it is useful in preventing and treating several diseases. Much attention is drawn towards the regular use of fingerroot in an attempt to benefit from their medicinal, therapeutic and culinary attributes. Despite the criticism received by herbal medicines among mainstream medical professionals, it is wise to remember that some well-known medicines that we use today have come from plants. Since there are a lot of research which have been done on fingerroot herbs in fighting dengue disease, the consumption of its extract may offer great advantage.

In addition, the finding of this study on how plants produce pharmacologically active compounds will offer the researchers a chance to further exploit the rich of biochemistry found in these plants by genetic modification. In addition, plant engineering could then be utilized to produce larger quantity of bioactive compounds using large-scale method. This study could help in exploring the means for how natural products are created, empowering efforts to discover new drug candidates and increasing the efficiency of the existing ones. The investigation on antidengue properties exhibited by panduratin A and hydroxypandutin A could give a promising solution towards dengue diseases. It can be achieved by focusing more on how to improve their production in vitro.

The idea on overexpressing desired gene such as *BrCHS2* gene (isolated from rhizomes) could give a clue to exploit the synthesis of both compounds. Collaborative efforts between a broad spectrum of expertise should greatly contribute to understanding how the plant biosynthesizes them and how the plant genes control the desired compound in a particular plant pathway.

CHAPTER 7: CONCLUSION

Referring back to the main hypothesis in this study, it is now possible to state that adding copy of *BrCHS* into suspension cultures of *B. rotunda* could enhance the flavonoid production and cause several genes to be differentially expressed in flavonoid and related plant pathways. Overall, there were at least five *BrCHS*s that were successfully characterized and detected in different parts of *B. rotunda* plants.

The study on enzymatic functionality of overexpressed CHS has shown that pinocembrin chalcone was successfully produced in higher amount in transgenic lines compared to wild-type. This depicted the BrCHS with abundance of targeted compound can be synthesized in that particular pathway by utilizing the rhizome's *CHS* sequence. By targeting to increase specific metabolite production in a particular pathway, this study provides a framework for the exploration of potential unigenes regulated in *B. rotunda* plants.

The limitation of this study is the lack of information on unavailable pathway in KEGG databases as other unigenes could appear as promising targets for overproduction of flavonoids in *B. rotunda.* Additional of reference pathway should be available in order to obtain a comprehensive metabolite pathway overview. Transcriptome data obtained in this work could enlarge the plant database as a reference for other Zingiberceae family members. This will provide new insight to predict and understand the mechanism involved in the synthesis of specific compound with potent medicinal properties which will ultimately be useful for future drug analysis.

Future research could also be conducted to determine the effect on silencing the *CHS* itself, adding more copies of the other downstream or potential genes involve in the

pathway that may lead to flavonoid production. More information on changes in gene expression level would help us to establish a greater degree of accuracy on this matter.

REFERENCES

Abe, I., & Morita, H. (2010). Structure and function of the chalcone synthase superfamily of plant type III polyketide synthases. *Natural Product Reports*, *27*(6), 809–838.

Abe, I., Sano, Y., Takahashi, Y., & Noguchi, H. (2003). Site-directed mutagenesis of benzalacetone synthase: The role of PHE 215 in plant type III polyketide synthases. *Journal of Biological Chemistry*, *278*(27), 25218–25226.

Angelova, S., Buchheim, M., Frowitter, D., Schierhorn, A., & Roos, W. (2010). Overproduction of alkaloid phytoalexins in california poppy cells is associated with the co-expression of biosynthetic and stress-protective enzymes. *Molecular Plant*, *3*(5), 927–939.

Annadurai, R. S., Neethiraj, R., Jayakumar, V., Damodaran, A. C., Rao, N., Katta, M. A. V. S. K. … Mugasimangalam, R. C. (2013). *De novo* transcriptome assembly (NGS) of *Curcuma longa* L . rhizome reveals novel transcripts related to anticancer and antimalarial terpenoids. *PLOS ONE*, *8*(2), 1–11.

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.

Axtell, M. J., & Staskawicz, B. J. (2003). Initiation of RPS2-specified disease resistance in Arabidopsis is coupled to the AvrRpt2-directed elimination of RIN4. *Cell*, *112*(3), 369–377.

Beerhues, L. (1996). Benzophenone synthase from cultured cells of *Centaurium erythraea*. *FEBS Letters*, *383*(3), 264–266.

Bhamarapravati, 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*, 157–163.

Bozak, K. R., O'keefe, D. P., & Christoffersen, R. E. (1992). Expression of a ripening-related avocado (*Persea americana*) cytochrome P450 in yeast. *Plant Physiology*, *100*(4), 1976–1981.

Brugliera, F., Barri-rewell, G., Holton, T. A., & Mason, J. G. (1999). Isolation and characterization of a favonoid 3-hydroxylase cDNA clone corresponding to the Ht1 locus of *Petunia hybrida*. *The Plant Journal*, *19*(4), 441–451.

Cai, Z., Riedel, H., Saw, N. M. M. T., Mewis, I., Reineke, K., Knorr, D., & Smetanska, I. (2011). Effects of elicitors and high hydrostatic pressure on secondary metabolism of *Vitis vinifera* suspension culture. *Process Biochemistry*, *46*(7), 1411–1416.

Carpentier, S. C., Witters, E., Laukens, K., Deckers, P., Swennen, R., & Panis, B. (2005). Preparation of protein extracts from recalcitrant plant tissues: An evaluation of

different methods for two-dimensional gel electrophoresis analysis. *Proteomics*, *5*(10), 2497–2507.

Chan, K.-L., Ho, C.-L., Namasivayam, P., & Napis, S. (2007). A simple and rapid method for RNA isolation from plant tissues with high phenolic compounds and polysaccharides. *Protocol Exchange*, 1. Retrieved from http://www.nature.com/protocolexchange/protocols/208

Chee, C. F., Abdullah, I., Buckle, M. J. C., & Rahman, N. A. (2010). An efficient synthesis of (±)-panduratin A and (±)-isopanduratin A, inhibitors of dengue-2 viral activity. *Tetrahedron Letters*, *51*, 495–498.

Cheenpracha, S., Karalai, C., Ponglimanont, C., Subhadhirasakul, S., & Tewtrakul, S. (2006). Anti-HIV-1 protease activity of compounds from *Boesenbergia pandurata*. *Bioorganic and Medicinal Chemistry*, *14*(6), 1710–1714.

Cho, M. H., & Lee, S. W. (2015). Phenolic phytoalexins in rice: Biological functions and Biosynthesis. *International Journal of Molecular Sciences*, *16*(12), 29120–29133.

Christensen, A. B., Gregersen, P. L., Schröder, J., & Collinge, D. B. (1998). A chalcone synthase with an unusual substrate preference is expressed in barley leaves in response to UV light and pathogen attack. *Plant Molecular Biology*, *37*(5), 849–857.

Ciulla, D., Giannoukos, G., Earl, A., Feldgarden, M., Gevers, D., Levin, J. … Birren, B. (2010). Evaluation of bacterial ribosomal RNA (rRNA) depletion methods for sequencing microbial community transcriptomes. *Genome Biology*, *11*(Suppl 1), 2142.

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

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A. … Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biology*, *17*(1), 13.

Dao, T. T. H., Linthorst, H. J. M., & Verpoorte, R. (2011). Chalcone synthase and its functions in plant resistance. *Phytochemistry Reviews*, *10*(3), 397–412.

DellaPenna, D. (2001). Plant metabolic engineering. *Plant Physiology*, *125*, 160–163.

Deng, X., Bashandy, H., Ainasoja, M., Kontturi, J., Pietiäinen, M., Laitinen, R. A. E. … Teeri, T. H. (2014). Functional diversification of duplicated chalcone synthase genes in anthocyanin biosynthesis of *Gerbera hybrida*. *New Phytologist*, *201*, 1469–1483.

Der, J. P., Barker, M. S., Wickett, N. J., DePamphilis, C. W., & Wolf, P. G. (2011). *De novo* characterization of the gametophyte transcriptome in bracken fern, *Pteridium aquilinum*. *BMC Genomics*, *12*(1), 99.

Dixon, R. A., & Pasinetti, G. M. (2010). Flavonoids and isoflavonoids: From plant biology to agriculture and neuroscience. *Plant Physiology*, *154*, 453–457.

Dukowic-Schulze, S., Harris, A., Li, J., Sundararajan, A., Mudge, J., Retzel, E. F. … Chen, C. (2014). Comparative transcriptomics of early meiosis in arabidopsis and maize. *Journal of Genetics and Genomics*, *41*(3), 139–152.

Durbin, M. L., McCaig, B., & Clegg, M. T. (2000). Molecular evolution of the chalcone synthase multigene family in the morning glory genome. *Plant Molecular Biology*, *42*(1), 79–92.

Éva, C., Csóti, I., & Tamás, L. (2008). Agrobacterium-mediated barley transformation. *Acta Biologica Szegediensis*, *52*(1), 49–51.

Ewing, B., & Green, P. (1998). Base-calling of automated sequencer traces using Phred. II. Error probabilities. *Genome Research*, *8*, 186–194.

Faktor, O., Loake, G., Dixon, R. A., & Lamb, C. J. (1997). The G-box and H-box in a 39 bp region of a french bean chalcone synthase promoter constitute a tissue-specific regulatory element. *Plant Journal*, *11*(5), 1105–1113.

Falcone Ferreyra, M. L., Rius, S. P., & Casati, P. (2012). Flavonoids: Biosynthesis, biological functions, and biotechnological applications. *Frontiers in Plant Science*, *3*, 1–22.

Feinbaum, R. L., & Ausubel, F. M. (1992). Transcriptional regulation of the *Arabidopsis thaliana* chalcone synthase gene. *Molecular and Cellular Biology*, *8*(5), 1985–1992.

Ferrer, J.-L., Austin, M. B., Stewart, C., & Noel, J. P. (2008). Structure and function of enzymes involved in the biosynthesis of phenylpropanoids. *Plant Physiology and Biochemistry*, *46*(3), 356–70.

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 Biology*, *6*(8), 775–784.

Franken, P., Niesbach-Klösgen, U., Weydemann, U., Maréchal-Drouard, L., Saedler, H., & Wienand, U. (1991). The duplicated chalcone synthase genes C2 and Whp (white pollen) of *Zea mays* are independently regulated; evidence for translational control of Whp expression by the anthocyanin intensifying gene in. *The EMBO Journal*, *10*(9), 2605–2612.

Franssen, S., Shrestha, R., Brautigam, A., Bauer, E., & Weber, A. (2011). Comprehensive transcriptome analysis of the highly complex *Pisum sativum* genome using next generation sequencing. *BMC Genomics*, *12*(1), 227.

Garg, R. O., Patel, R. A. V. I. K., Tyagi, A. K. K., & Jain, M. U. (2011). *De novo* assembly of chickpea transcriptome using short reads for gene discovery and marker identification. *DNA Research*, *18*, 53–63.

Garg, R., Patel, R. K., Jhanwar, S., Priya, P., Bhattacharjee, A., Yadav, G. … Jain, M. (2011). Gene discovery and tissue-specific transcriptome analysis in chickpea with massively parallel pyrosequencing and web resource development. *Plant Physiology*, *156*(4), 1661–1678.

Gelvin, S. B. (1998). The introduction and expression of transgenes in plants. *Current Opinion in Biotechnology*, *9*(2), 227–232.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I. … Regev, A. (2013). Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature Biotechnology*, *29*(7), 644–652.

Gueven, A., & Knorr, D. (2011). Isoflavonoid production by soy plant callus suspension culture. *Journal of Food Engineering*, *103*(3), 237–243.

Guo, Y., Zhao, S., Sheng, Q., Guo, M., Lehmann, B., Pietenpol, J. … Shyr, Y. (2015). RNAseq by total RNA library identifies additional RNAs compared to Poly(A) RNA library. *BioMed Research International*, *2015*, 1–9.

Hah, N., Danko, C. G., Core, L., Waterfall, J. J., Siepel, A., Lis, J. T., & Kraus, W. L. (2012). A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. *Cell*, *145*(4), 622–634.

Halbwirth, H. (2010). The creation and physiological relevance of divergent hydroxylation patterns in the flavonoid pathway. *International Journal of Molecular Sciences*, *11*(2), 595–621.

Han, Y., Ding, T., Su, B., & Jiang, H. (2016). Genome-wide identification, characterization and expression analysis of the chalcone synthase family in maize. *International Journal of Molecular Sciences*, *17*(2).

Hao, D. C., Ge, G., Xiao, P., Zhang, Y., & Yang, L. (2011). The first insight into the tissue specific taxus transcriptome via illumina second generation sequencing. *PLOS ONE*, *6*(6).

Harker, C. L., Ellis, T. H., & Coen, E. S. (1990). Identification and genetic regulation of the chalcone synthase multigene family in pea. *The Plant Cell*, *2*(3), 185–94.

Hendrawati, O., Woerdenbag, H. J., Hille, J., & Kayser, O. (2012). Metabolic engineering of medicinal plants and microorganisms for the production of natural products. In *Pharmaceutical biotechnology: Drug Discovery and clinical applications.* (2nd ed., pp 491–526)

Hernandez-Garcia, C. M., & Finer, J. J. (2014). Identification and validation of promoters and cis-acting regulatory elements. *Plant Science*, *217–218*, 109–119.

Hiei, Y., Ishida, Y., & Komari, T. (2014). Progress of cereal transformation technology mediated by *Agrobacterium tumefaciens*. *Frontiers in Plant Science*, *5*, 1–11.

Hwang, E. Il, Hwang, E. Il, Ohnishi, Y., Ohnishi, Y., Horinouchi, S., & Horinouchi, S. (2003). Production of plant-specific flavanones by flavonones by *Escherichia coli* containing an artificial gene cluster. *Applied and Environmental Microbiology*, *69*(5), 2699–2706.

Ibraheem, F., Gaffoor, I., & Chopra, S. (2010). Flavonoid phytoalexin-dependent resistance to anthracnose leaf blight requires a functional *yellow seed1* in *Sorghum bicolor*. *Genetics*, *184*(4), 915–926.

Ito, M., Ichinose, Y., Kato, H., Shiraishi, T., & Yamada, T. (1997). Molecular evolution and functional relevance of the chalcone synthase genes of pea. *Molecular & General Genetics*, *255*, 28–37.

Jedinak, A., Farago, J., Psenakova, I., & Maliar, T. (2004). Approaches to flavonoid production in plant tissue cultures. *Biologia*, *59*(6), 697–710.

Jefferson, R. A., Kavanagh, T. A., & Bevan, M. W. (1987). GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *The EMBO Journal*, *6*(13), 3901–3907.

Jenkins, G. I., Long, J. C., Wade, H. K., Shenton, M. R., & Bibikova, T. N. (2001). UV and blue light signalling: pathways regulating chalcone synthase gene expression in Arabidopsis. *New Phytologist*, *151*(1), 121–131.

Jez, J. M., Bowman, M. E., & Noel, J. P. (2002). Expanding the biosynthetic repertoire of plant type III polyketide synthases by altering starter molecule specificity. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(8), 5319–5324.

Jez, J. M., Ferrer, J. L., Bowman, M. E., Dixon, R. A., & Noel, J. P. (2000). Dissection of malonyl-coenzyme A decarboxylation from polyketide formation in the reaction mechanism of a plant polyketide synthase. *Biochemistry*, *39*(5), 890–902.

Jiang, C., Schommer, C. K., Kim, S. Y., & Suh, D.-Y. (2006). Cloning and characterization of chalcone synthase from the moss, *Physcomitrella patens*. *Phytochemistry*, *67*(23), 2531–2540.

Jiang, H., Wood, K. V, & Morgan, J. A. (2005). Metabolic engineering of the phenylpropanoid pathway in *Saccharomyces cerevisiae*. *Applied and Environmental Microbiology*, *71*(6), 2962–2969.

Johnson, M. T. J., Carpenter, E. J., Tian, Z., Bruskiewich, R., Burris, J. N., Carrigan, C. T. … Wong, G. K. S. (2012). Evaluating methods for isolating total RNA and predicting the success of sequencing phylogenetically diverse plant transcriptomes. *PLOS ONE*, *7*(11), 1–12.

Junghanns, K. T., Kneusel, R. E., & Matern, D. G. U. (1998). Differential regulation and distribution of acridone synthase in *Ruta graveolens* in honour of Professor G. H. Neil Towers 75th birthday. *Phytochemistry*, *49*(2), 403–411.

Kaltenbach, M., Schröder, G., Schmelzer, E., Lutz, V., & Schröder, J. (1999). Flavonoid hydroxylase from Catharanthus roseus: cDNA, heterologous expression, enzyme properties and cell-type specific expression in plants. *Plant Journal*, *19*(2), 183–193.

Kanobe, M. N., Rodermel, S. R., Bailey, T., & Scott, M. P. (2013). Changes in endogenous gene transcript and protein levels in maize plants expressing the soybean ferritin transgene. *Frontiers in Plant Science*, *4*, 196.

Khayat, E., Duvdevani, A., Lahav, E., & Balesteros, B. A. (2004). Somaclonal variation in banana (*Musa acuminate* cv. Grande Naine): genetic mechanism, frequency and application as a tool for clonal selection. In S. M. Jain & R. Swennen (Eds.), Banana

improvement: cellular, molecular biology and induced mutation. *Science*, (pp 97-109).

Kirakosyan, A., & Kaufman, P. B. (2009). *Recent advances in plant biotechnology* (pp 15-33). Springer.

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*, 131–137.

Klingauf, P., Beuerle, T., Mellenthin, A., El-Moghazy, S. A. M., Boubakir, Z., & Beerhues, L. (2005). Biosynthesis of the hyperforin skeleton in *Hypericum calycinum* cell cultures. *Phytochemistry*, *66*(2), 139–145.

Koes, R. E., Spelt, C. E., & Mol, J. N. M. (1989). The chalcone synthase multigene family of *Petunia hybrida* (V30): differential, light-regulated expression during flower development and UV light induction. *Plant Molecular Biology*, *12*(2), 213–225.

Kole, C., Michler, C. H., Abbott, A. G., & Hall, T. C. (2010). *Transgenic crop plants*. (Vol 1). Berlin Heidelberg: Springer-Verlag

Komori, T., Takakura, Y., Ueld, J., Kato, N., Ishida, Y., & Hiei, Y. (2008). Binary vectors and super binary vectors. In methods in molecular biology (Vol. 343, pp 15–42). Humana Press Inc

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.

Koulman, A., Kubbinga, M. E., Batterman, S., Woerdenbag, H. J., Pras, N., Woolley, J. G., & Quax, W. J. (2003). A phytochemical study of lignans in whole plants and cell suspension cultures of *Anthriscus sylvestris*. *Planta Medica*, *69*, 733–738.

Kreuzaler, F., Ragg, H., Fautz, E., Kuhn, D. N., & Hahlbrock, K. (1983). UV-induction of chalcone synthase mRNA in cell suspension cultures of *Petroselinum hortense*. *Proceedings of the National Academy of Sciences of the United States of America*, *80*(9), 2591–3.

Kumar, J., & Gupta, P. K. (2008). Molecular approaches for improvement of medicinal and aromatic plants. *Plant Biotechnology Reports*, *2*(2), 93–112.

Łabaj, P. P., Leparc, G. G., Linggi, B. E., Markillie, L. M., Wiley, H. S., & Kreil, D. P. (2011). Characterization and improvement of RNA-seq precision in quantitative transcript expression profiling. *Bioinformatics*, *27*(13), 383–391.

Latest dengue information. (2017). Retrieved from http://idengue.remotesensing.gov.my/idengue/index.php

Latunde-dada, A. O., Cabello-hurtado, F., Czittrich, N., Didierjean, L., Schopfer, C., Hertkorn, N. … Ebel, J. (2001). Flavonoid 6-hydroxylase from soybean (*Glycine max* L.), a novel plant P-450 monooxygenase. *The Journal of Biological Chemistry*,

276(3), 1688–95.

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), 1–10.

Li, M., Cao, Y.-T., Ye, S.-R., Irshad, M., Pan, T.-F., & Qiu, D.-L. (2016). Isolation of CHS gene from *Brunfelsia acuminata* flowers and its regulation in anthocyanin biosysthesis. *Molecules*, *22*(1), 44.

Li, R., Tee, C.-S., Jiang, Y.-L., Jiang, X.-Y., Venkatesh, P. N., Sarojam, R., & Ye, J. (2015). A terpenoid phytoalexin plays a role in basal defense of *Nicotiana benthamiana* against potato virus X. *Scientific Reports*, *5*, 9682.

Li, Y., Wang, X., Chen, T., Yao, F., Li, C., Tang, Q. … Song, S. (2015). RNA-Seq based *de novo* transcriptome assembly and gene discovery of *Cistanche deserticola* fleshy stem. *PLOS ONE*, *10*(5), 1–19.

Liu, B., Falkenstein-Paul, H., Schmidt, W., & Beerhues, L. (2003). Benzophenone synthase and chalcone synthase from *Hypericum androsaemum* cell cultures: cDNA cloning, functional expression, and site-directed mutagenesis of two polyketide synthases. *Plant Journal*, *34*(6), 847–855.

Liu, C., Zhao, Y., & Wang, Y. (2006). Artemisinin: Current state and perspectives for biotechnological production of an antimalarial drug. *Applied Microbiology and Biotechnology*, *72*(1), 11–20.

Liu, Y., Zhou, J., & White, K. P. (2014). RNA-seq differential expression studies: More sequence or more replication? *Bioinformatics*, *30*(3), 301–304.

Lo, C., Coolbaugh, R. C., & Nicholson, R. L. (2002). Molecular characterization and in silico expression analysis of a chalcone synthase gene family in *Sorghum bicolor*. *Physiological and Molecular Plant Pathology*, *61*(3), 179–188.

Lu, T., Lu, G., Fan, D., Zhu, C., Li, W., Zhao, Q. … Han, B. (2010). Function annotation of the rice transcriptome at single-nucleotide resolution by RNA-seq. *Genome Research*, *20*(9), 1238–1249.

Malacarne, G., Vrhovsek, U., Zulini, L., Cestaro, A., Stefanini, M., Mattivi, F. … Moser, C. (2011). Resistance to Plasmopara viticola in a grapevine segregating population is associated with stilbenoid accumulation and with specific host transcriptional responses. *BMC Plant Biology*, *11*(1), 114.

Mandal, S. M., Chakraborty, D., & Dey, S. (2010). Phenolic acids act as signaling molecules in plant-microbe symbioses. *Plant Signaling & Behavior*, *5*(4), 359–68.

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

Martin, J., Zhu, W., Passalacqua, K. D., Bergman, N., & Borodovsky, M. (2010). *Bacillus anthracis* genome organization in light of whole transcriptome sequencing. *BMC Bioinformatics*, *11*(Suppl 3), 1–17.

Matsumura, H., Watanabe, S., Harada, K., Senda, M., Akada, S., Kawasaki, S. … Takahashi, R. (2005). Molecular linkage mapping and phylogeny of the chalcone synthase multigene family in soybean. *Theoretical and Applied Genetics*, *110*(7), 1203–1209.

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.

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

Mo, Y., Nagel, C., & Taylor, L. P. (1992). Biochemical complementation of chalcone synthase mutants defines a role for flavonols in functional pollen. *Proceedings of the National Academy of Sciences of the United States of America*, *89*(15), 7213–7217.

Morita, H., Abe, I., & Noguchi, H. (2010). Natural products structural diversity-I secondary metabolites: Organization and biosynthesis: Plant type III PKS. In Craig A. Townsend, & Yutaka Ebizuka (Eds.). *Comprehensive natural products II: Chemistry and biology*. (1st ed., pp 171-226). United Kingdom: Elsevier Ltd

Mouradov, A., & Spangenberg, G. (2014). Flavonoids: a metabolic network mediating plants adaptation to their real estate. *Frontiers in Plant Science*, *5*, 1–25.

Murashige, T. and F. S. (1962). A revised medium for rapid growth and bioassays with tobacco tissue cultures. *Physiologia Plantarum*, *15*, 473–497.

Nelson, D. R., Schuler, M. A., Paquette, S. M., Werck-Reichhart, D., & Bak, S. (2004). Comparative genomics of rice and Arabidopsis. Analysis of 727 cytochrome P450 genes and pseudogenes from a monocot and a dicot. *Plant Physiology*, *135*(2), 756–72.

Oberholzer, V., Durbin, M. L., & Clegg, M. T. (2000). Comparative genomics of chalcone synthase and MYB genes in the grass family. *Genes and Genetic Systems*, *75*(1), 1–16.

Okada, Y., & Ito, K. (2001). Cloning and analysis of valerophenone synthase gene expressed specifically in lupulin gland of hop (*Humulus lupulus* L.). *Bioscience, Biotechnology, and Biochemistry*, *65*(1), 150–155.

Olsen, K. M., Hehn, A., Jugdé, H., Slimestad, R., Larbat, R., Bourgaud, F., & Lillo, C. (2010). Identification and characterisation of CYP75A31, a new flavonoid 3'5'-hydroxylase, isolated from *Solanum lycopersicum*. *BMC Plant Biology*, *10*(21), 1–18.

Ozsolak, F., & Milos, P. M. (2011). RNA sequencing: advances, challenges and opportunities. *Nature Reviews. Genetics*, *12*(2), 87–98.

Pang, Y., Shen, G., Wu, W., Liu, X., Lin, J., Tan, F. … Tang, K. (2005). Characterization and expression of chalcone synthase gene from *Ginkgo biloba*. *Plant Science*,

168(6), 1525–1531.

Paniego, N. B., Zuurbier, K. W. M., Fung, S. Y., Van Der Heijden, R., Scheffer, J. J. C., & Verpoorte, R. (1999). Phlorisovalerophenone synthase, a novel polyketide synthase from hop (*Humulus lupulus* L.) cones. *European Journal of Biochemistry*, *262*(2), 612–616.

Pattaratanawadee, E., Thongson, C., Mahakarnchanakul, W., & Wanchaitanawong, P. (2006). Antimicrobial activity of spice extracts against pathogenic and spoilage microorganisms. *The Proceedings of the 44th Kasetsart University Annual Conference, Kasetsart, Subject: Agro-Industry, Economics, Business Administration*, *165*, 297–307.

Prasath, D., Karthika, R., Habeeba, N. T., Suraby, E. J., Rosana, O. B., Shaji, A. … Anandaraj, M. (2014). Comparison of the transcriptomes of ginger (*Zingiber officinale* Rosc.) and mango ginger (*Curcuma amada* Roxb.) in response to the bacterial wilt infection. *PLOS ONE*, *9*(6), 1–12.

Qu, J. G., Zhang, W., Yu, X. J., & Jin, M. F. (2005). Instability of anthocyanin accumulation in *Vitis vinifera* L. var. Gamay Freaux suspension cultures. *Biotechnology and Bioprocess Engineering*, *10*(2), 155–161.

Radhakrishnan, E. K., Varghese, R. T., & Vasudevan, S. E. (2010). Unusual intron in the second exon of a Type III polyketide synthase gene of *Alpinia calcarata* Rosc. *Genetics and Molecular Biology*, *33*(1), 141–145.

Rajesh, M., Jeyaraj, M., Sivanandhan, G., Subramanyam, K., Mariashibu, T. S., Mayavan, S. … Ganapathi, A. (2013). Agrobacterium-mediated transformation of the medicinal plant *Podophyllum hexandrum* Royle (syn. P. emodi Wall. ex Hook.f. & Thomas). *Plant Cell, Tissue and Organ Culture*, *114*(1), 71–82.

Ralston, L., & Yu, O. (2006). Metabolons involving plant cytochrome P450s. *Phytochemistry Reviews*, *5*(2–3), 459–472.

Ramachandra Rao, S., & Ravishankar, G. A. (2002). Plant cell cultures: Chemical factories of secondary metabolites. *Biotechnology Advances*, *20*(2), 101–153.

Rao, K., Hanuman, J., Alvarez, C., Stoy, M., Juchum, J., Davies, R., & Baxley, R. (1995). A new large-scale process for taxol and related taxanes from *Taxus brevifolia*. *Pharmaceutical Biotechnology*, *12*(7), 1003–1010.

Rasche, S., Herwartz, D., Schuster, F., Jablonka, N., Weber, A., Fischer, R., & Schillberg, S. (2016). More for less: Improving the biomass yield of a pear cell suspension culture by design of experiments. *Scientific Reports*, *6*, 23371.

Rastogi, S., Meena, S., Bhattacharya, A., Ghosh, S., Shukla, R. K., Sangwan, N. S. … Gupta, V. (2014). *De novo* sequencing and comparative analysis of holy and sweet basil transcriptomes. *BMC Genomics*, *15*(588), 1–22.

Reyes, F. C., Sun, B., Guo, H., Gruis, D. F., & Otegui, M. S. (2010). *Agrobacterium tumefaciens*-mediated transformation of maize endosperm as a tool to study endosperm cell biology. *Plant Physiology*, *153*, 624–631.

Ribas, A. F., Dechamp, E., Champion, A., Bertrand, B., Combes, M.-C., Verdeil, J.-L. … Etienne, H. (2011). Agrobacterium-mediated genetic transformation of *Coffea arabica* (L.) is greatly enhanced by using established embryogenic callus cultures. *BMC Plant Biology*, *11*(1), 92.

Roslan, N. D., Tan, C., Ismail, I., & Zainal, Z. (2013). cDNA cloning and expression analysis of the chalcone synthase gene (CHS) from *Polygonum minus*. *Australian Journal of Crop Science*, *7*(6), 777–783.

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.

Saito, K., Yonekura-Sakakibara, K., Nakabayashi, R., Higashi, Y., Yamazaki, M., Tohge, T., & Fernie, A. R. (2013). The flavonoid biosynthetic pathway in Arabidopsis: Structural and genetic diversity. *Plant Physiology and Biochemistry*, *72*, 21–34.

Sakarya, O., Breu, H., Radovich, M., Chen, Y., Wang, Y. N., Barbacioru, C. … Siddiqui, A. S. (2012). RNA-Seq mapping and detection of gene fusions with a suffix array algorithm. *PLOS Computational Biology*, *8*(4), 1–10.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). *Molecular cloning: A laboratory manual.* Cold Spring Harbor, NY: Cold Spring. Harbor Laboratory Press.

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

Seitz, C., Ameres, S., & Forkmann, G. (2007). Identification of the molecular basis for the functional difference between flavonoid 3'-hydroxylase and flavonoid 3',5'-hydroxylase. *FEBS Letters*, *581*, 3429–3434.

Seitz, C., Eder, C., Deiml, B., Kellner, S., Martens, S., & Forkmann, G. (2006). Cloning, functional identification and sequence analysis of flavonoid 3'-hydroxylase and flavonoid 3',5'-hydroxylase cDNAs reveals independent evolution of flavonoid 3',5'-hydroxylase in the Asteraceae family. *Plant Molecular Biology*, *61*, 365–381.

Shashank, K., & Abhay, K. (2013). Review article chemistry and biological activities of flavonoids: An overview. *The Scientific World Journal*, *4*(2), 32–48.

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.

Skae, F. M. (1902). Dengue fever in Penang. *The British Medical Journal*, *2*(2185), 1581–1582.

Snyder, M. (2010). The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science*, *320*(5881), 1344–1349.

Sobolev, V. S., Khan, S. I., Tabanca, N., Wedge, D. E., Manly, S. P., Coy, S. J. C. … Gloer, B. S. A. N. J. B. (2011). Biological activity of peanut (*Arachis hypogaea*)

phytoalexins and selected natural and synthetic stilbenoids. *Journal of Agricultural and Food Chemistry*, *59*(5), 1673–1682.

Sood, P., Bhattacharya, A., & Sood, A. (2011). Problems and possibilities of monocot transformation. *Biologia Plantarum*, *55*(1), 1–15.

Springob, K., Lukacin, R., Ernwein, C., Gröning, I., & Matern, U. (2000). Specificities of functionally expressed chalcone and acridone synthases from *Ruta graveolens*. *European Journal of Biochemistry*, *267*(22), 6552–6559.

Steinkellner, S., Lendzemo, V., Langer, I., Schweiger, P., Khaosaad, T., Toussaint, J. P., & Vierheilig, H. (2007). Flavonoids and strigolactones in root exudates as signals in symbiotic and pathogenic plant-fungus interactions. *Molecules*, *12*(7), 1290–1306.

Sun, W., Meng, X., Liang, L., Jiang, W., Huang, Y., He, J. … Wang, L. (2015). Molecular and biochemical analysis of chalcone synthase from *Freesia hybrid* in flavonoid biosynthetic pathway. *PLOS ONE*, *10*(3), 1–18.

Sun, X., Zhou, S., Meng, F., & Liu, S. (2012). *De novo* assembly and characterization of the garlic (*Allium sativum*) bud transcriptome by Illumina sequencing. *Plant Cell Reports*, *31*(10), 1823–1828.

Swarbreck, S. M., Lindquist, E. A., Ackerly, D. D., & Andersen, G. L. (2011). Analysis of leaf and root transcriptomes of soil-grown *Avena barbata* plants. *Plant and Cell Physiology*, *52*(2), 317–332.

Tabata, H. (2006). Production of paclitaxel and the related taxanes by cell suspension cultures of Taxus species. *Current Drug Targets*, *7*(4), 453–461.

Tai, D., Tian, J., Zhang, J., Song, T., & Yao, Y. (2014). A Malus crabapple chalcone synthase gene, McCHS, regulates red petal color and flavonoid biosynthesis. *PLOS ONE*, *9*(10), 1–13.

Tan, B. C., Tan, S. K., Wong, S. M., Ata, N., Rahman, N. A., & Khalid, N. (2015). Distribution of flavonoids and cyclohexenyl chalcone derivatives in conventional propagated and *in vitro*-derived field-grown *Boesenbergia rotunda* (L.) Mansf. *Evidence-Based Complementary and Alternative Medicine*, 1-7.

Tan, E. C., Karsani, S. A., Foo, G. T., Wong, S. M., Abdul Rahman, N., 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, H.-C., Tan, B.-C., Wong, S.-M., & Khalid, N. (2016). A medicinal ginger, *Boesenbergia rotunda*: From cell suspension cultures to protoplast derived callus. *Sains Malaysiana*, *45*(5), 795–802.

Tan, S. K., Pippen, 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 and Medicinal Chemistry Letters*, *16*(12), 3337–3340.

Tarazona, S., Garcıa-Alcalde, F., Dopazo, J., Ferrer, A., & Conesa, A. (2011). Differential expression in RNA-seq : A matter of depth. *Genome Research*, 1–20.

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., & Puripattanavong, Jindaporn Panphadung, T. (2003). HIV-1 protease inhibitory substances from the rhizomes of *Boesenbergia pandurata* Holtt . *Songklanakarin Journal of Science and Technology*, *25*(4), 503–508.

Tewtrakul, S., Subhadhirasakul, S., & Sopa, K. (2003). HIV-1 protease inhibitory effects of medicinal plants used as self medication by AIDS patients. *Songklanakarin Journal of Science and Technology*, 1–6.

The Plant List. (2016). Version 1.1. Retrieved from http://www.theplantlist.org/

Tornero, P., Chao, R., Luthin, W. N., Goff, S., & Dangl, J. L. (2002). Large-scale structure-function analysis of the Arabidopsis RPM1 disease resistance protein. *The Plant Cell Online*, *14*(2), 435–450.

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.

Treangen, T. J., & Salzberg, S. L. (2013). Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nature Reviews Genetics*, *13*(1), 36–46.

Trojan, V., Musilová, M., Vyhnánek, T., Klejdus, B., Hanáček, P., & Havel, L. (2014). Chalcone synthase expression and pigments deposition in wheat with purple and blue colored caryopsis. *Journal of Cereal Science*, *59*(1), 48–55.

Tuteja, J. H., Clough, S. J., Chan, W.-C., & Vodkin, L. O. (2004). Tissue-specific gene silencing mediated by a naturally occurring chalcone synthase gene cluster in Glycine max. *The Plant Cell*, *16*, 819–835.

Verpoorte, R., Contin, A., & Memelink, J. (2002). Biotechnology for the production of plant secondary metabolites. *Phytochemistry Reviews*, *1*(1), 13–25.

Wakasa, Y., Oono, Y., Yazawa, T., Hayashi, S., Ozawa, K., Handa, H. … Takaiwa, F. (2014). RNA sequencing-mediated transcriptome analysis of rice plants in endoplasmic reticulum stress conditions. *BMC Plant Biology*, *14*(1), 101.

Wang, H., Chung, P. J., Liu, J., Jang, I. C., Kean, M. J., Xu, J., & Chua, N. H. (2014). Genome-wide identification of long noncoding natural antisense transcripts and their responses to light in Arabidopsis. *Genome Research*, *24*, 444–453.

Wang, Q., Hillwig, M. L., & Peters, R. J. (2011). CYP99A3: Functional identification of a diterpene oxidase from the momilactone biosynthetic gene cluster in rice. *Plant Journal*, *65*(1), 87–95.

Wang, Y.-S., Xu, Y.-J., Gao, L.-P., Yu, O., Wang, X.-Z., He, X.-J. … Xia, T. (2014). Functional analysis of flavonoid 3',5'-hydroxylase from tea plant (*Camellia sinensis*): critical role in the accumulation of catechins. *BMC Plant Biology*, *14*, 347.

Wang, Y., Li, J., & Xia, R. (2010). Expression of chalcone synthase and chalcone isomerase genes and accumulation of corresponding flavonoids during fruit maturation of Guoqing No. 4 satsuma mandarin (*Citrus unshiu* Marcow). *Scientia Horticulturae*, *125*(2), 110–116.

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

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

Wanibuchi, K., Zhang, P., Abe, T., Morita, H., Kohno, T., Chen, G. … Abe, I. (2007). An acridone-producing novel multifunctional type III polyketide synthase from *Huperzia serrata*. *FEBS Journal*, *274*(4), 1073–1082.

Wannapinpong, S., Srikulnath, K., Thongpan, A., Choowongkomon, K., & Peyachoknagul, S. (2013). Molecular cloning and characterization of the CHS gene family in turmeric (*Curcuma longa* Linn.). *Journal of Plant Biochemistry and Biotechnology*, *24*(1), 25–33.

Werck-Reichhart, D., & Feyereisen, R. (2000). Cytochromes P450: A success story. *Genome Biology*, *1*(6), 1–9.

Whitmer, S., Canel, C., Van Der Heijden, R., & Verpoorte, R. (2003). Long-term instability of alkaloid production by stably transformed cell lines of *Catharanthus roseus*. *Plant Cell, Tissue and Organ Culture*, *74*(1), 73–80.

Winkel-Shirley, B. (2001a). Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiology*, *126*, 485–493.

Winkel-Shirley, B. (2001b). It takes a garden. How work on diverse plant species has contributed to an understanding of flavonoid metabolism. *Plant Physiology*, *127*(4), 1399–1404.

Wiriyaampaiwong, P., Thanonkeo, P., & Thanonkeo, S. (2010). Cloning and characterization of chalcone synthase gene from *Pueraria candollei* var. mirifica. *Journal of Biotechnology*, *150*(42), 478–478.

Witherup, K. M., Look, S. A., Stasko, M. W., Ghiorzi, T. J., & Muschik, G. M. (1990). Taxus Spp. needles contain amounts of taxol comparable to the bark of *Taxus Brevzfolza*: Analysis and isolation. *Journal of Natural Products*, *53*(5), 1249–1255.

Wong, M. M. L., Cannon, C. H., & Wickneswari, R. (2011). Identification of lignin genes and regulatory sequences involved in secondary cell wall formation in *Acacia auriculiformis* and *Acacia mangium* via *de novo* transcriptome sequencing. *BMC Genomics*, *12*(342), 1–16.

Wong, S. M., Salim, N., Harikrishna, J. A., & Khalid, N. (2013). Highly efficient plant regeneration via somatic embryogenesis from cell suspension cultures of *Boesenbergia rotunda*. *In Vitro Cellular and Developmental Biology - Plant*, *49*(6), 665–673.

Wu, Z.-G., Jiang, W., Mantri, N., Bao, X.-Q., Chen, S.-L., & Tao, Z.-M. (2015). Transciptome analysis reveals flavonoid biosynthesis regulation and simple sequence repeats in yam (*Dioscorea alata* L.) tubers. *BMC Genomics*, *16*(1), 346.

Xia, Z., Xu, H., Zhai, J., Li, D., Luo, H., He, C., & Huang, X. (2011). RNA-Seq analysis and *de novo* transcriptome assembly of *Hevea brasiliensis*. *Plant Molecular Biology*, *77*(3), 299–308.

Xiao, M., Zhang, Y., Chen, X., Lee, E. J., Barber, C. J. S., Chakrabarty, R. … Sensen, C. W. (2013). Transcriptome analysis based on next-generation sequencing of non-model plants producing specialized metabolites of biotechnological interest. *Journal of Biotechnology*, *166*(3), 122–134.

Yadnya-Putra, A. A. G. R., Chahyadi, A., & Elfahmi. (2014). Production of panduratin A, cardamomin and sitosterol using cell cultures of fingerroot (*Boesenbergia pandurata* (Roxb.) schlechter). *Biosciences Biotechnology Research Asia*, *11*(1), 43–52.

Yang, J., Huang, J., Gu, H., Zhong, Y., & Yang, Z. (2002). Duplication and adaptive evolution of the chalcone synthase genes of Dendranthema (Asteraceae). *Molecular Biology and Evolution*, *19*(10), 1752–1759.

Yarden, K. (2011). Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nature Methods*, *7*(12), 1009–1015.

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*, 293–297.

Ye, X., Al-Babili, S., Klöti, A., Zhang, J., Lucca, P., Beyer, P. … Uze, M. (2000). Engineering the Provitamin A (β-Carotene) Biosynthetic Pathway into (Carotenoid-Free) Rice Endosperm. *Science*, *287*(5451), 96–305.

Yue, W., Ming, Q.-L., Lin, B., Rahman, K., Zheng, C.-J., Han, T., & Qin, L.-P. (2014). Medicinal plant cell suspension cultures: pharmaceutical applications and high-yielding strategies for the desired secondary metabolites. *Critical Reviews in Biotechnology*, *8551*, 1–18.

Yusuf, N. A. (2011). *Biomass and selected flavonoids production in cell suspension cultures of Boesenbergia rotunda (L.) mansf.* Retrieved from http://studentsrepo.um.edu.my/6050/

Yusuf, N. A., Suffian, M., Annuar, M., & Khalid, N. (2013). Existence of bioactive

flavonoids in rhizomes and plant cell cultures of *Boesenbergia rotunda* (L.) Mansf. Kulturpfl. *Australian Journal of Crop Science*, *7*(6), 730–734.

Zaeoung, S., Plubrukarn, A., & Keawpradub, N. (2005). Cytotoxic and free radical scavenging activities of Zingiberaceous rhizomes. *Songklanakarin Journal Science and Technology*, *27*(4), 799–812.

Zhang, W. J., Dewey, R. E., Boss, W., Phillippy, B. Q., & Qu, R. (2013). Enhanced Agrobacterium-mediated transformation efficiencies in monocot cells is associated with attenuated defense responses. *Plant Molecular Biology*, *81*(3), 273–286.

Zhang, Y. H., Zhong, J. J., & Yu, J. T. (1996). Enhancement of ginseng saponin production in suspension cultures of Panax notoginseng: Manipulation of medium sucrose. *Journal of Biotechnology*, *51*(1), 49–56.

Zhao, D., Jiang, Y., Ning, C., Meng, J., Lin, S., Ding, W., & Tao, J. (2014). Transcriptome sequencing of a chimaera reveals coordinated expression of anthocyanin biosynthetic genes mediating yellow formation in herbaceous peony (*Paeonia lactiflora* Pall.). *BMC Genomics*, *15*(1), 689.

Zhao, J. L., Zhou, L. G., & Wu, J. Y. (2010). Effects of biotic and abiotic elicitors on cell growth and tanshinone accumulation in Salvia miltiorrhiza cell cultures. *Applied Microbiology and Biotechnology*, *87*(1), 137–144.

LIST OF PUBLICATIONS AND PAPERS PRESENTED

Abstract in proceedings:

Roslan, N.N., Teh, S.H., Mohamed, Z. (2015). *De novo* transcriptome assembly and analyses of *B. rotunda* cell suspension cultures with overexpresses *Chalcone synthase (CHS)* shows changes on the flavonoid biosynthesis and other related plant pathways. *Universiti Malaya-Academia Sinica (UM-AS) Joint Research Symposium, Date: 18-20th November 2015. Faculty of Science, UM (Oral presentation)*

Roslan, N.N., Teh, S.H., Wong, S.M., Khalid, N., Mohamed, Z. (2013). Overexpression of CHS in the cell suspension cultures of fingerroot, *Boesenbergia rotunda*. *The 10th Malaysia Genetic Congress, Date: 3-5th December 2013. Palm Garden Hotel IOI Resort, Putrajaya (Oral presentation)*

Roslan, N.N., Teh, S.H., Wong, S.M., Khalid, N., Mohamed, Z. (2012). Construction and overexpression of *chalcone synthase (CHS)* gene in cell suspension cultures of ginger (*Boesenbergia rotunda). The17th Biological Sciences Graduate Congress, Date: 8th -10th Dec 2012, Chulalongkorn University, Bangkok, Thailand (Oral presentation)*

Roslan, N.N., Teh, S.H., Wong, S.M., Khalid, N., Mohamed, Z. (2012). Construction and overexpression of *chalcone synthase (CHS)* gene in cell suspension cultures of ginger (*Boesenbergia rotunda).The 19th Scientific Meeting of MSMBB, Date: 31st Oct – 1st Nov 2012, University Malaya, Malaysia (Oral presentation)*