CHARACTERIZATION OF 4-COUMARATE: COENZYME A LIGASE FROM *BOESENBERGIA ROTUNDA*

NABEEL ATA ABDUL MUNEIM

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

2017

CHARACTERIZATION OF 4-COUMARATE: COENZYME A LIGASE FROM *BOESENBERGIA ROTUNDA*

NABEEL ATA ABDUL MUNEIM

DISSERTATION SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE

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

2017

UNIVERSITY OF MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: NABEEL ATA ABDUL MUNEIM

Matric No: SGR130068

Name of Degree: MASTER OF SCIENCE

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

CHARACTERIZATION OF 4-COUMARATE: COENZYME A LIGASE FROM *BOESENBERGIA ROTUNDA*

Field of Study:

I do solemnly and sincerely declare that:

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

Candidate's Signature Date:

Subscribed and solemnly declared before,

Witness's Signature Date:

Name: PROF. DR. NORZULAANI KHALID

Designation: PROFESSOR

ABSTRACT

Boesenbergia rotunda, formerly known as *Kaempferia pandurata Roxb.* or *Boesenbergia pandurata (Roxb. Schltr*), is a small perennial monocotyledonous plant belonging to the Zingiberaceae family. One of the important secondary metabolites produces by *B. rotunda* is flavonoids which derived from phenylpropanoid metabolisms. These compounds are essential for plant growth and development and have been used for biomedical applications. One of the key enzymes in the phenylpropanoid metabolism, 4-coumarate: coenzyme A ligase (*4CL*), is yet to be characterized. In this study, *B. rotunda* cell suspension cultures were first treated with phenylalanine precursor and the expression level of *4CL* was analyzed to determine the role of *4CL* in the production of flavonoids. In Real-time analysis, results show that the *4CL* transcript level for 14- and 10-day old phenylalanine-treated cultures were up-regulated to 22- and 17-fold, respectively, compared to 0-day, indicating the promotive effect of phenylalanine in enhancing the *4CL* expression. cDNA for *4CL* was then isolated and characterized from *B. rotunda* cell suspension cultures using combined methods of RACE and Degenerate Primers. The isolated *4CL* cDNA sequence consists of 1,578 bp open reading frame (ORF) which encoding 526 amino acids with a similarity of 82% to 4-coumarate-CoA ligase 1-like [*Musa acuminata subsp. malaccensis*]. Sequence analysis showed that the *4CL* contains AMP-binding domain (PFSSGTTGLPKGV), GEICIRG motif and conserved VPP and PVL domains. The *4CL* coding sequence was then transformed into *Escherichia coli* strain BL2 (DE3) and analyzed its protein expression by SDS-PAGE. Our preliminary protein expression result shows that the 4CL was successfully expressed in *E.coli*. These findings merit further studies such as overexpression of *4CL* gene in *B. rotunda* to enhance the flavonoid production.

ABSTRAK

Boesenbergia rotunda, sebelum ini dikenali sebagai *Kaempferia Pandurata Roxb*. atau *Boesenbergia Pandurata* (*Roxb.Schltr*), adalah tumbuhan monokot renek dari keluarga *Zingiberaceae*. Salah satu sebatian penting yang di hasilkan oleh *B. rotunda* adalah *flavonoid*s yang diperolehi daripada metabolisma fenilalanina. Sebatian ini adalah perlu dalam pertumbuhan dan perkembangan tumbuhan, dan berguna untuk aplikasi bioperubatan. Salah satu enzim utama dalam metabolisma *phenylpropanoid* adalah 4-coumarate: coenzyme ligase A (4CL) yang masih belum dicirikan. Dalam kajian ini, sel ampaian *Boesenbergia rotunda* telah dirawat oleh fenilalanina dan tahap ekspresi *4CL* dianalisa untuk menentukan peranan *4CL* dalam pengeluaran *flavonoids*. Keputusan analisa *Real-time* menunjukkan tahap transkrip *4CL* untuk hari ke 14 dan 10 bagi sel yang dirawat oleh fenilalanina meningkat ke tahap 17 kali ganda, dan 22 kali ganda, berbanding dengan hari ke-0, menunjukkan kesan penggalakan fenilalanina dalam meningkatkan tahap ekspresi *4CL* itu. cDNA dari *4CL* kemudiannya dipencilkan dan ditentukan ciri-cirinya daripada sel ampaian *B. rotunda* dengan menggunakan kaedah *RACE* dan *Degenerate Primers* . *4CL* yang telah dipencilkan mempunyai jujukan sepanjang 1,578 *bp* rangka (*ORF*) dan mengekodkan sebanyak 526 asid amino dengan persamaan sehingga 82% dengan 4-coumarate: CoA ligase 1-like [*Musa acuminata subsp. malaccensis*]. Analisis jujukan menunjukkan bahawa gen *4CL* mengandungi domain pengikat-AMP (PFSSGTTGLPKGV), GEICIRG motif, domain VPP dan PVL yang terpulihara. Gen *4CL* kemudiannya ditransformasikan ke dalam sel bakteria *Escherichia coli* jenis BL2 (DE3) dan tahap ekspresi protin yang terhasil di analisis oleh kaedah *SDS-PAGE*. Keputusan awal ekspresi protin menunjukkan bahawa 4CL telah berjaya diekspreskan ke dalam *E.coli*. Penemuan ini membolehkan kajian yang lebih lanjut dilakukan seperti penambahan ekspresi gen *4CL* dalam *B. rotunda* untuk meningkatkan penghasilan *flavonoid*.

ACKNOWLEDGEMENTS

In the name of ALLAH S.W.T the most Gracious and the most Merciful, all praises to Him who had given me blessings, strengths, and knowledge in completing this thesis entitled "Characterization of 4-coumarate: coenzyme A ligase in *Boesenbergia rotunda*".

First and foremost, I sincerely express my utmost appreciation to my supervisors, Professor Dr. Norzulaani Khalid and Dr. Yusmin Mohd Yusuf for their support, guidance, patience and advice throughout the whole project. I am really appreciated

Thank you to University of Malaya for providing me the fellowship (SBP) and PPP grants (RP003-2012D and PG218-2014B) for my study. I also want to express my gratitude to all my lab-mates for their technical and experimental assistances, information sharing and mental support. A special thanks to Dr. Tan Boon Chin, Dr. Jameel and Plant Biotechnology Research Laboratory (PBRL) lab members for providing good laboratory environment plus the consistent guidance and cool attitudes during my study.

Last but not least, I would like to extent my gratitude to my parents, siblings and friends for their continuous moral support throughout my degree of master study.

Alhamdulillah.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

%	:	percent
µg	:	microgram
µl	:	microliter
µM	:	micromolar
BAP	:	benzyl amino purine
bp	:	base pair
cDNA	:	complimentary deoxyribonucleic acid
Ct	:	threshold cycle
DNA	:	deoxyribonucleic acid
DNase 1	:	deoxyribonuclease acid 1
dNTP	:	deoxyribonucleoside triphosphates
DTT	:	dithiothreitol
E. coli	:	*Escherichia coli*
EDTA	:	ethylenediaminetetraacetic acid
EtBr	:	ethidium bromide
H_2O	:	water
IPTG	:	isopropyl β-D-1-thiogalactopyranoside
Kb	:	kilo base pair
L	:	liter
LB	:	Luria-Bertani
M	:	molar
mg	:	miligram
$MgCl_2$	:	magnesium chloride
ml	:	mililiter
mM	:	milimolar

mRNA	:	messenger ribonucleic acid
MS	:	Murashige and Skoog
MW	:	molecular weight
NAA	:	1-naphthylacetic acid
NaOH	:	sodium hydroxide
NCBI	:	National Center for Biotechnology Information
nm	:	nanometer
^{0}C	:	degree celcius
PCR	:	polymerase chain reaction
RNA	:	ribonucleic acid
RNase	:	ribonuclease
rpm	:	rotation per minute
RT	:	reverse transcriptase
RT-PCR:		reverse-transcriptase polymerase chain reaction
RT-qPCR:		real-time polymerase chain reaction or quantitative polymerase chain reaction
SDS	:	sodium dodecyl sulphate
TBE	:	tris-borate-EDTA
Tris	:	tris (hydroxyl methyl) amino methane
U	:	unit
UV	:	ultraviolet
v	:	volt
v/v	:	volume per volume
w/v	:	weight per volume

LIST OF APPENDICES

CHAPTER 1: INTRODUCTION

1.1 GENERAL INTRODUCTION

Boesenbergia rotunda, formerly known as *Kaempferia pandurata Roxb.* or *Boesenbergia pandurata (Roxb. Schltr*), is a small perennial monocotyledonous plant belongs to the Zingiberaceae family. Traditionally, it has been used as food ingredients and in folk medicine. *B. rotunda* is also known as Chinese keys, Temu Kunci or Finger root ginger. It is believed to have originated from India and South China.

B. rotunda produces flavonoids like other plants. Flavonoids are secondary metabolites derived from phenylalanine and acetate metabolisms and exhibit various functions in higher plants, such as provide flower colouring to attract pollinators, seed dispersers, influence the transport of the plant hormones, UV protection, and balance the levels of reactive oxygen species. Besides that, these flavonoids have been found to have a wide range of biochemical and pharmacologic activities, such as antioxidant, antibacterial, antifungal, anti-inflammatory, antitumor, and anti-tuberculosis activities (Tan et al. 2012; Tan et al. 2015), suggesting their possible role as health-promoting and disease-preventing dietary supplements. Previous studies have shown that these polyphenols or flavonoids can be abundantly found in the rhizomes of *B. rotunda* (Jaipetch et al., 1982; Tan et al., 2015).

However, the presences of some specific bioactive flavonoids are very little. In order to use flavonoid as main compounds for drug developments and commercial purposes, many enhancement strategies for compound accumulation have been published in various reports. These strategies include the use of precursors and manipulation of metabolic pathways, including phenylpropanoid pathway.

The phenylpropanoid pathway is an important pathway in plant secondary metabolisms, catalyzing the conversion of phenylalanine via the shikimate and aerogenate pathways to a myriad of phenolic secondary metabolites such as wall bounds phenolic acids, anthocyanin, benzoid, isoflavonoids, lignins, suberins, stilbenes and flavonoids. Phenylpropanoid pathway starts with the conversion of amino acid L-phenylalanine to trans-cinnamate by phenylalanine ammonia lyase (PAL). Trans-cinnamate is then hydroxylated by cinnamic-4-hydroxylases (C4H) and 4-coumarate: coenzyme A ligase (4CL) through the formation of a thioester bond to become a single thioester molecule of 4-coumaryl: CoA. These thioesters are central intermediates for the production of many other secondary metabolites including flavonoid compounds (Dixon and Paiva, 1995).

As one of the important enzyme in phenylpropanoid pathway, characterization of *4CL* gene in plant especially *B.rotunda* initiate the step toward understanding the gene regulation in this pathway. This information is also important in genetic manipulation studies. Furthermore, heterologous expression of this gene in *E.coli* can be exploited for the production of secondary metabolites compounds in either microbial host or transgenic *B.rotunda* plant. High copy numbers of this gene could lead to higher production of targeted bioactive compounds in the plant itself.

As previously reported, the production of secondary metabolites could be induced or increased by supplying precursors or intermediate compounds to the cell cultures (Hussain et al., 2012). In this study, phenylalanine was used as a precursor to observe its effects on the expression of *4CL* gene in *B.rotunda* suspension cells that may lead to the better understanding of the mechanism of phenylpropanoid pathway in order to increase the secondary metabolites production.

Later, *4CL* gene was characterized from *B. rotunda* cell suspension cultures and its expression profile was analyzed after being treated with phenylalanine at 0, 10, and 14 days. It was found that the *4CL* transcript levels for 14- and 10-day old phenylalanine-treated cultures were up-regulated for about 22- and 17-fold, respectively, compared to 0-day, indicating the promotive effect of phenylalanine in enhancing the *4CL* gene expression. A 1,578bp open reading frame (ORF) of cDNA fragment of *4CL* encoding 526 amino acids was isolated from *B. rotunda* cell suspension cultures. The isolated *4CL* (designated as *Br4CL*) showed 82% identity with 4-coumarate: CoA ligase 1-like [*Musa acuminata subsp. malaccensis*] and phylogenetically closed to *4CL* from *Musa acuminata subsp malaccensis*. Sequence analysis showed that most of the motifs/ domains were presence such as AMP binding domain, AMP binding domain c terminal and were grouped together with adenylate forming enzyme superfamily. The sequence encoded 4CL protein was synthesized and cloned into pET-28a vector and was successfully over-expressed in *E.coli* cells**.**

1.2 OBJECTIVES

1. To analyze the expression profile of 4-coumarate: coenzyme A ligase (*4CL*) gene in *B.rotunda* cell suspension cultures after being treated with phenylalanine over time

2. To isolate and characterize the full length cDNA of *4CL* gene in *B. rotunda* using Rapid Amplification of cDNA Ends (RACE) method

3. To transform genes that encodes 4-coumarate: coenzyme A Ligase enzymes responsible for catalyzing cinnamic acid to cinnamoyl-CoA into *E. coli*

CHAPTER 2: LITERATURE REVIEW

2.1 *Boesenbergia rotunda*

The Zingiberaceae is one of the largest families from the order Zingiberales, with approximately 50 genera and over 1,000 species. It is estimated about 150 species of ginger belonging to 23 genera are found in Peninsular Malaysia. *Boesenbergia rotunda* formerly known as *Kaempferia pandurata Roxb* is a small perennial monocotyledonous plant which belongs to the Zingiberaceae family (**Table 2.1**). It is traditionally known as Chinese keys, Temu Kunci or Fingerroot ginger. *B. rotunda* is well-known as folk medicine and food ingredient either as a spice or flavouring in South East Asia (**Figure 2.1**).

One of the important secondary metabolites produced by *B.rotunda* is flavonoid. Flavonoids are derived from phenylalanine and acetate metabolisms with a variety of essential functions in higher plants. They are ubiquitous plant natural products and have low molecular weight composed of a three-ring structure like tocopherols. In plants, these metabolites play a big role in the colour of the plants and are involved in plant defense mechanism against UV light and pathogen infections. Some of the major subgroups of flavonoids are chalcones, flavanones, flavones, flavonols, aurones, isoflavonoids, and flavanediols, which comprises of the largest group of natural products (Winkel-Shirley, 2001). Flavonoids are also remarkably known as safe nutrients with a wide range of biochemical and pharmacologic activities, suggesting their possible role as health-promoting and disease-preventing dietary supplements. These metabolites are biosynthesized through the phenylpropanoid pathway.

.

Table 2.1: Scientific classification of *Boesenbergia rotunda*

Kingdom	**: Plantae**
Division	**: Magnoliophyta**
Order	**: Zingiberales**
Genus	**: *Boesenbergia***
Species	**: *rotunda***

A

B

C

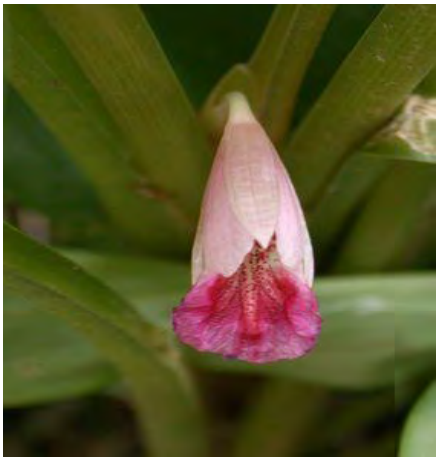


Figure 2.1.: *B. rotunda* whole plant (A), their matured rhizomes (B) and their flowers(C). Bar: 2.5 cm

2.2 Medicinal properties of *B. rotunda* compounds

Rhizomes of *B. rotunda* have been reported to contain many useful secondary metabolites, such as pinostrobin, pinocembrin, cardamonin, alpinetin and boesenbergin A (Jaipetch et al., 1982). Its biological activities include antimutagenic, antitumour, antibacterial, antifungal, analgesic, antipyretic, antispasmodic, anti-inflammatory and insecticidal activities (Cheenpracha et al., 2005).

Isa and co-workers(2012) reported that boesenbergin A has an anti-inflammatory, cytoxicity and anti-oxidant activities in *in-vitro* studies with multiple pharmacological properties whereas pinocembrin has an anti-oxidant and apoptosis role where it protects rat's brain from oxidation and induced apoptosis by ischemia-reperfusion (Liu et al., 2008). Pinocembrin is a potential candidate for cancer treatment as most of the anticancer drugs presently used in clinical settings caused cell death by apoptosis (Cen et al., 2002). Prenylated flavonoids such as panduratin A and 4-hydroxypanduratin A exhibited antioxidant, antibacterial, antifungal, anti-inflammatory, antitumor, and anti-tuberculosis activities (Trakoontivakom et al., 2001; Tan et al., 2012; Tan et al., 2015;).

Interestingly, there are more than 60% of prenylated flavonoids that specifically exist only in *B.rotunda*. Prenylated flavonoids, exhibited a wide range of pharmacological effects with very promising biological activity (Chahyadi et al., 2014). The presence of lipophilic prenyl group in a molecule could help enhance the transmembrane transport and interaction of the molecule with the target protein. Besides that, structure-activity relationship (SAR) analysis showed that compounds having prenyl group in their structure will have higher biological activities such as anticancer, anti-bacterial, anti-inflammatory compared to compounds without prenyl group. Some prenylated aromatic compounds were shown to interrupt the RAS signal cascade (intracellular signal transduction system) that is responsible in regulating many cell

functions activity such as proliferation, differentiation and inflammation by interfering with the binding of RAS to its plasma membrane anchor protein (Tanaka et al., 2010). However, further studies need to be carried out in order to understand the detailed mechanisms by which prenyl group influences pharmacological activity.

Generally, flavonoids are readily ingested by humans and they seem to display important anti-inflammatory, anti-allergic and anti-cancer activities. Flavonoids are also found to be powerful anti-oxidants and researchers are looking into their ability to prevent cancer and cardiovascular diseases (Kris-Etherton et al., 2002). Flavonoids help prevent cancer by inducing certain mechanisms that may help to kill cancer cells, and researchers believe that when the body processes extra flavonoid compounds, it triggers specific enzymes that fight carcinogens (Lall et al., 2016).

2.3 Phenylpropanoid biosynthetic pathway

Flavanoids are derived from phenylpropanoid pathway. This phenylpropanoid pathway is an important pathway in plant secondary metabolism that is able to converts phenylalanine to numerous secondary metabolites, such as phenolic acids, lignins, coumarins, tannins, anthocyanins, chlorogenics acids, caffeic acids, stilbenes and flavonoids (**Figure 2.2**). These diverse structures of compounds play essential roles in plant development, including mechanical support, signalling compounds, flower pigments, UV protectants and are also involved in disease resistance mechanism. Besides that, structure compounds such as flavonoids have been exploited for human diet purpose as they have pharmacological characteristics. In general, there are three enzymatic transformations redirecting the carbon flow from primary metabolism, transforming phenylalanine into the Coenzyme A (CoA)-activated hydroxycinnamoyl

or activated thioester. They are phenylalanine ammonia-lyase (PAL), cinnamic acid 4-hydroxylase (C4H) and 4-coumarate: coenzyme A Ligase (4CL).

4CL is an enzyme responsible for the abundance of activated thioesters reservoir capable of entering into two major downstream pathways which are monolignol and flavonoid biosynthesis. Other specialized compounds such as stillbenes, isoflavonoids and aurones biosynthesis are also produced through general phenylpropanoid pathway (**Figure 2.3**). The pathway **(Figure 2.3)** starts with the deamination of phenylalanine by PAL to produce cinnamic acid followed by introduction of a hydroxyl group at the *para* position of the phenyl ring of cinnamic acid by C4H to form *p*-coumaric acid. The carboxyl group of *p*-coumaric acid is then catalyzed by 4CL to form a thioester bond with CoA which results in the formation of *p*-coumaroyl-CoA (Ferrer et al., 2008). These thioesters are the key intermediates that can later be directed to different subsequent pathways, such as lignin, flavanoids or hormone metabolisms (Ferrer et al., 2008). These general phenylpropanoid enzymes usually would be activated by stresses such as wounding, UV radiation, or pathogen attack and are largely controlled at the transcriptional level (Bauer et al., 2009; Sullivan 2009; Gutiérrez-Carbajal et al., 2010).

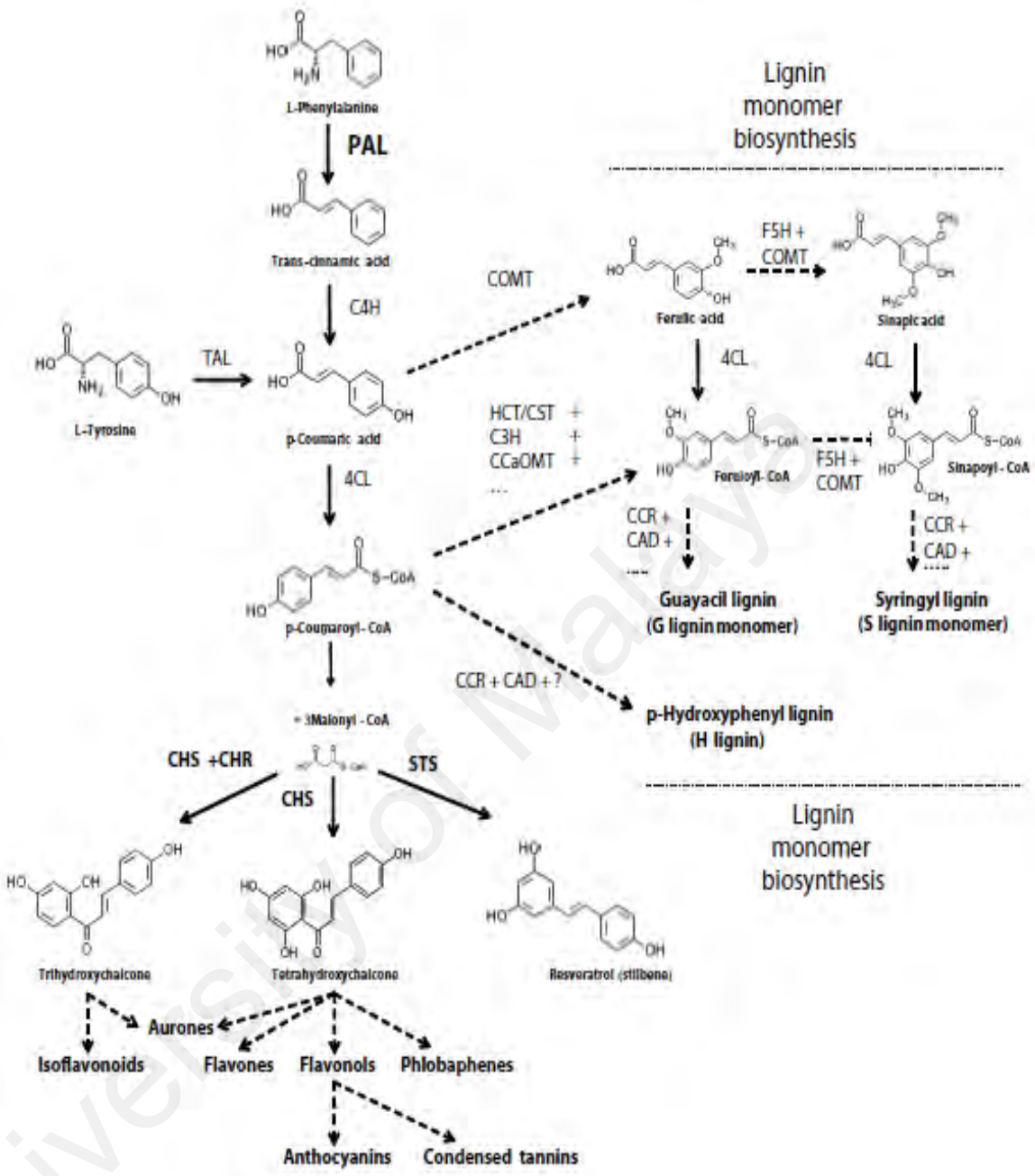


Figure 2.2: The general phenylpropanoid pathway (top left, reactions from L-phenylalanine to *p*-Coumaroyl-CoA). Two separated branches lead to the production of lignin monomers (right) and of flavonoids (bottom). Solid arrows indicate a single step enzymatic reaction while dashed arrows show multiple sequential enzymatic reactions. Enzymes are reported with a three letter code: PAL, phenylalanine ammonia lyase; C4H, cinnamate 4-hydroxylase; 4CL, 4-coumarate: CoA ligase (Emiliani et al., 2009).

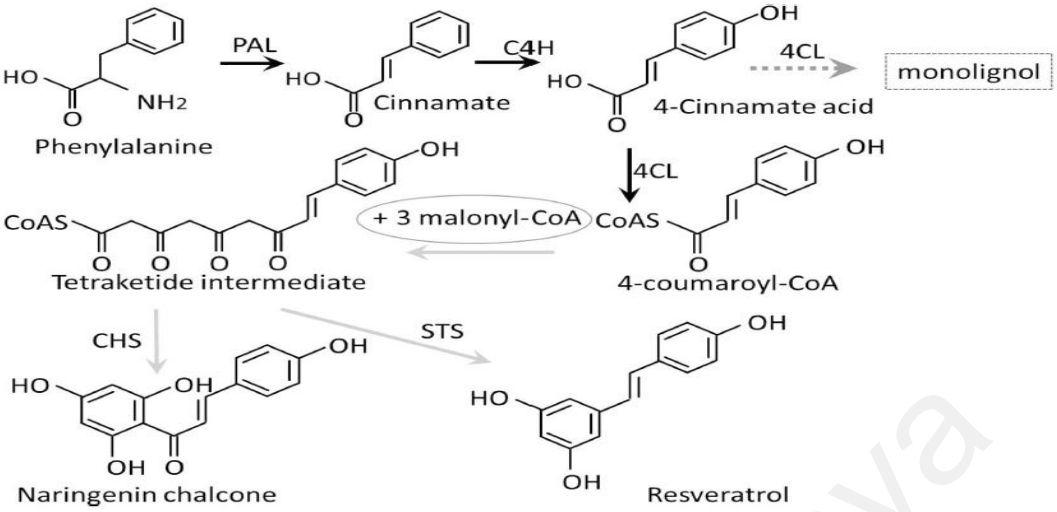


Figure 2.3: Phenylpropanoid biosynthetic pathway. 4CL enzyme is the last enzyme in the phenylpropanoid pathway where the products of 4CL (thioesters) serve as substrates for other subsequent enzymes such as chalcone synthase (CHS) or stillbene synthase (STS). Dashed arrows indicate a branch pathway for lignin biosynthesis emanating from the general phenylpropanoid metabolic pathway, which is also catalyzed by 4CL. The gray arrows indicate the pathways for chalcone and resveratrol biosynthesis. Source from Wang et al., (2016)

2.4 4-coumarate: coenzyme A Ligase (EC 6.2.1.12), a flavonoid- related gene

4CL is the final enzyme in the phenylpropanoid pathway that regulates carbon flux into downstream biosynthetic pathways. It is also one of the members of acyl adenylate forming enzymes (Cukovica et al., 2001) and adenosine monophosphate AMP-binding proteins (Lindermayr et al., 2003). 4CL is encoded by multiple paralogs and exist as a multi-gene family or isoforms in most plant species.

Expression of *4CL* is restricted in certain organs and their substrate preferences are also varied across the species (K_m). For instance, expression of *Pt4CL1* gene was restricted to leaf and stem epidermal cells in aspen (Hu et al., 1998). In an aspen trees study, two *4CL* genes were cloned and their expression were clearly distinct, with one in epidermal and leaf tissue, and the others specifically in developing xylem tissue (Harding et al., 2002; Hu et al., 1998). Earlier studies in Arabidopsis indicated that

different isozymes of 4CL showed different substrate preferences and gene expression patterns where *At4CL1*, and *At4CL2*, were involved in the monolignol biosynthetic pathway, while *At4CL3* participated in flavonoid and other nonlignin biosynthetic pathways (Lee et al., 1995; Ehlting et al., 1999; Cukovica et al., 2001). The *At4CL1* and *At4CL2* genes were constitutively expressed in lignified bolting mature stems and at the onset of lignin deposition in roots and cotyledons (Lee et al., 1995), whereas *At4CL3* was expressed in all light-exposed organs, such as siliques, leaves, and flowers. Besides that, *Ptr4CL1* in poplar plant which was involved in lignin biosynthesis was found in developing xylem tissues, whereas *Ptr4CL2* was important for flavonoid biosynthesis in epidermal cells (Hu et al., 1998). Grand and co-workers (1983) reported that three isoforms of 4CL showed different substrate specificities and gene expressions in poplar stems too. In contrast, two genes encoding virtually identical 4CL enzymes were found to have the same substrate specificity in parsley (Lozoya et al., 1988). Interestingly, Voo et al. (1995) found only one form of 4CL in loblolly pine xylem which was quite rare. In switchgrass plant, silencing of *Pv4CL1* has led to a reduction in lignin content without affecting the plant growth (Xu et al., 2011). The results above demonstrated that the regulation of *4CL* genes is complex, and thus, different 4CL family members show distinct expression patterns in response to biotic or abiotic stress.

It has been shown that the 4CL enzyme is not only found in plant, but also in bacteria called *Streptomyces coelicolor* A3(2) (Kaneko et al., 2003). Interestingly, the protein has higher sequence similarity to plant 4CLs compared to bacterial acyl-CoA ligase. However, this *4CL* gene could activate the conversion of cinnamic acid to cinnamoyl-CoA, which is a very poor substrate for plant 4CL (Kaneko et al., 2003). This indicated that although the enzymes from these 2 origins were similar in protein homology sequence, they differed in terms of substrate affinities. This showed the possibility of cross species conservation.

2.4.1 Reaction mechanism of *4CL* gene

The reaction of 4CL-catalyzed CoA ester formation takes place via two-step reactions (**Figure 2.4**). During the first step, 4-coumarate and ATP form coumaroyl-adenylate intermediate with the simultaneous release of pyrophosphate. In the second step, the coumaroyl group is transferred to the sulfhydryl group of CoA, and AMP is released (Knobloch et al., 1975; Becker-Andre et al., 1991). 4CL enzyme ligates CoA with their corresponding carboxylic acids to form thioesters.

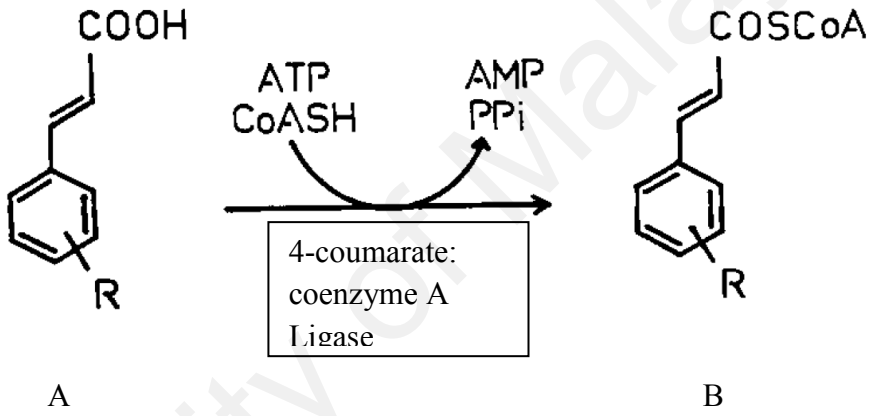


Figure 2.4: The conversion of cinnamic acids (A) to cinnamoyl-CoA (B) by 4-coumarate: coenzyme A Ligase.

2.5 Production of natural compounds in recombinant organism

Natural products have been the single most abundant compounds that lead to the development of drugs. These include compounds from plants, microbes and animals, as well as synthetic or semi-synthetic compounds based on natural products (Harvey, 2008).The importance of natural products in medicine and health is undeniable and documented by civilizations throughout human history (Ji et al., 2009). Today, natural products from plants remain an abundant source for pharmaceuticals products. Bioactive candidates of small molecules such as secondary metabolites or phenolic compounds which are difficult to isolate or found in trace amount in nature can be exploited for large scale production by using the *E.coli* expression system or using recombinant microorganism as a host or factories for targeted bioactive compounds production. For example, *E.coli* expression system.

E.coli is one of the most well characterized microorganisms that have been utilized as a recombinant host for protein expression and also for the production of small molecules/compounds through heterologous expression of clustered gene (Marienhagen et al., 2013). The *E. coli* cells are modified by introducing a partial or entire biosynthetic pathway gene cluster for the production of the target compound. The product can be either final bioactive compounds or intermediates that need downstream chemical synthesis. The *E.coli* cells growth conditions can be optimized to maximise product formation (Pandey and Sohng, 2013). Some of the compounds that have been developed by pharmaceutical company using recombinant microorganism were listed in **Table 2.2.**

Microorganism such as *Pseudomonas putida*, *Bacillus subtilis* and *Streptomyces* spp have also have been used as a heterologous host systems for natural product biosynthesis (Zhang et al., 2016). Besides that, *Saccharomyces cerevisiae* is

increasingly being used for the overproduction of plant and fungal natural products, such as isoprenoids, terpenoids, avonoids and alkaloids (Siddiqui et al., 2012).

Although there are many advantages that can be exploited through this method such as clean chemical background for downstream natural product detection and separation, and fast growth rates to harvest the compound but in term of industrial production for large scale commercialization purpose, the yield is still a problem and is not sustainable. Nevertheless, there are lots of strategies being optimized for increasing the target compound to an industrial level or demand. The strategies included strains improvement, manipulating regulatory pathways and engineering cellular translational and transcriptional machineries of the host and also searching new enzymes from other organism that are capable to perform the desired catalytic activity.

Table 2.2: Compounds developed by pharmaceutical company

Target compound	**Company**	**References**
Paclitaxel	BMS	Banerjee *et al*. (1994)
Lamivudine	Glaxo	Mahmoudian *et al*. (1993)
Omapatrilat	BMS	Patel *et al*. (2001)
Xemilofiban	Monsanto	Landis *et al*. (2002)
β-Lactams	Glaxo	Stead *et al*. (1996)

*Companies: BMS, Bristol-Myers Squibb; Glaxo, Glaxo Wellcome/Glaxo Smith Kline;

2.5.1 Production of compounds using *4CL* gene in microbes

In order to produce the targeted 4CL proteins, supplementation of natural on unnatural substrates, such as L-tyrosine, L-phenylalanine, or carboxylic acids (cinnamic acids and ferulic acids) into *E. coli* culture is needed. These substrates are then converted into targeted compounds through enzymatic reactions. For instance, in 2007 Katsuyama and co-workers developed recombinant *E. coli* containing *4CL-1, ACC, CHS,* and chalcone isomerase (*CHI)* genes to convert exogenous fed natural carboxylic acids (coumaric acid, sinapic acid, ferulic acid) and unnatural carboxylic acids (fluorocinnamic acid, acrylic acid) into their corresponding flavanones. In addition, successful production of plant-specific flavanones in bacteria through combinatorial biosynthetic approaches have resulted in construction of libraries of ‘‘unnatural’’ natural compounds (Kaneko et al., 2003). Watts and co-workers (2004) successfully produced flavanone (naringenin) using cinnamic acids as a substrate instead of phenylalanine by expressing *4CL* and *CHS* genes obtained from *A. thaliana* in *E. coli* host (Watts et al., 2004). In 2005, Watts and co-workers have demonstrated engineering of a recombinant, flavanone-producing *Saccharomyces cerevisiae* strain by simultaneous expression of four flavanone biosynthetic enzymes which are C4H, 4CL, CHS, and CHI. Furthermore, it has been shown that recombinant *E. coli* cultures expressing 4CL from *A. thaliana* and stilbene synthase (STS) cloned from *Arachis hypogaea*, converted the externally added precursor 4-coumaric acid to resveratrol (100 mg/L) and externally added caffeic acid to piceatannol (10 mg/L) (Watts et al. 2006). Besides that, an artificial curcuminoid biosynthetic pathway was constructed in *E. coli* by co-expressing PAL from the yeast *Rhodotorula rubra*, 4CL from *Lithospermum erythrorhizon* and curcuminoid synthase (CUS) from rice (*Oryza sativa L.;Poaceae*), which resulted in the production of curcuminoids by the recombinant *E. coli* (Hwang et al. 2003).

2.5.2 Production of compounds using *4CL* gene in plants

The yield of a desired plant product or metabolites can be enhanced through, a channelling of intermediates into specific products. Such channelling might be obtained by blocking unwanted side branches through gene-silencing of the first enzymatic step of the side-branch pathway or any other identified candidate gene (Van Der Krol et al., 1988). Besides that, product formation may be enhanced by overexpressing those genes whose products rate are limiting in the synthesis of the desired metabolites. In plants, *4CL* genes or their isoform were one of the candidate genes that can be manipulated as these enzymes may play a role in the biosynthesis of secondary metabolites such as flavonoids. The product of 4CL enzymes which is *p*-coumaroyl-CoA is also a precursor for the production of secondary plant metabolites such as stilbenes, lignins and flavonoids. Thus, silencing or over expressing *4CL* gene may affect the production of these flavonoids too. In addition, it is also affect the production of the building blocks of lignin which is their major pathway as shown by many studies in *Nicotiana tabacum*, *Arabidopsis thalian*, and *Populus tremuloides* (Lee et al., 1996; Hu et al., 1998; Li et al., 2015).

Mutation study on *4CL* has been reported to affect the secondary metabolite production as some of their isoforms may be involved in flavonoids pathway. Mutation on *At4CL3* significantly reduced the flavonol glycoside content in Arabidopsis by approximately 80% compared to wild-type plants, while mutations in the other *At4CL* genes did not significantly affect the flavonol glycoside content (Li et al., 2015). Besides that, flavonol glycosides were found accumulated in the aerial part of the Arabidopsis plant, and this 4CL3 has been shown to be important in floral flavonol biosynthesis (Dobritsa et al., 2011). This study proved the complexity and the role of 4CL enzymes in secondary metabolites production in plants particularly flavonoids.

CHAPTER 3: MATERIALS & METHODS

3.1 Materials

Cell suspension culture derived from the shoot explants (sprouts) of *B.rotunda* plants was used for both isolation of *4CL* gene and the phenylalanine treatment experiment.

The primer sets were prepared from reference sequence of *4CL* that was extracted from *B.rotunda* transcriptomic database, Unigene36813_All (Diyana et al., 2015) (**Appendix C14**) for the amplification of the coding region of *4CL* gene and for the relative expression study of *4CL* in response to phenylalanine treatment in *B.rotunda* suspension cultures. From prior experiments, it was found that suspension cultures treated with phenylalanine produced higher accumulation of selected flavonoids (Tan et al., 2012). cDNA template was used for all PCR reactions during the amplification of 4CL coding region and real-time PCR analysis.

For the phenylalanine treatment experiment, five groups were formed namely C0, C10, T10, C14, and T14. Each group had 3 replicates which were named as C0/1, C0/2 and C0/3. The replicates represented 3 different 250 ml conical flasks of suspension cultures of *B. rotunda*. Subsequently, for each replicate, 6 samples were taken and placed in microcentrifuge tubes (1.5 ml) for RNA extraction. These tubes were labelled as C0/1-1 until C0/1-6. Similar labels were done for other replicates (**Appendix D**). Therefore, for each group, a total of 18 tubes were prepared for RNA samples extraction.

The nomenclature for C is for control and T is for treated (phenylalanine) suspension cultures and the numbers represent the days the samples were harvested for RNA extraction. C0 and C10 groups represent control cultures (not treated with

phenylalanine) that were harvested for RNA extraction during 0 day and after 10 days respectively. While, C14 is control cultures harvested after 14 days. Samples T10 and T14 represent samples that underwent phenylalanine treatment and were incubated for 10 days and 14 days respectively before samples were harvested for RNA extraction. The overview of the project is as shown in **Figure 3.1**.

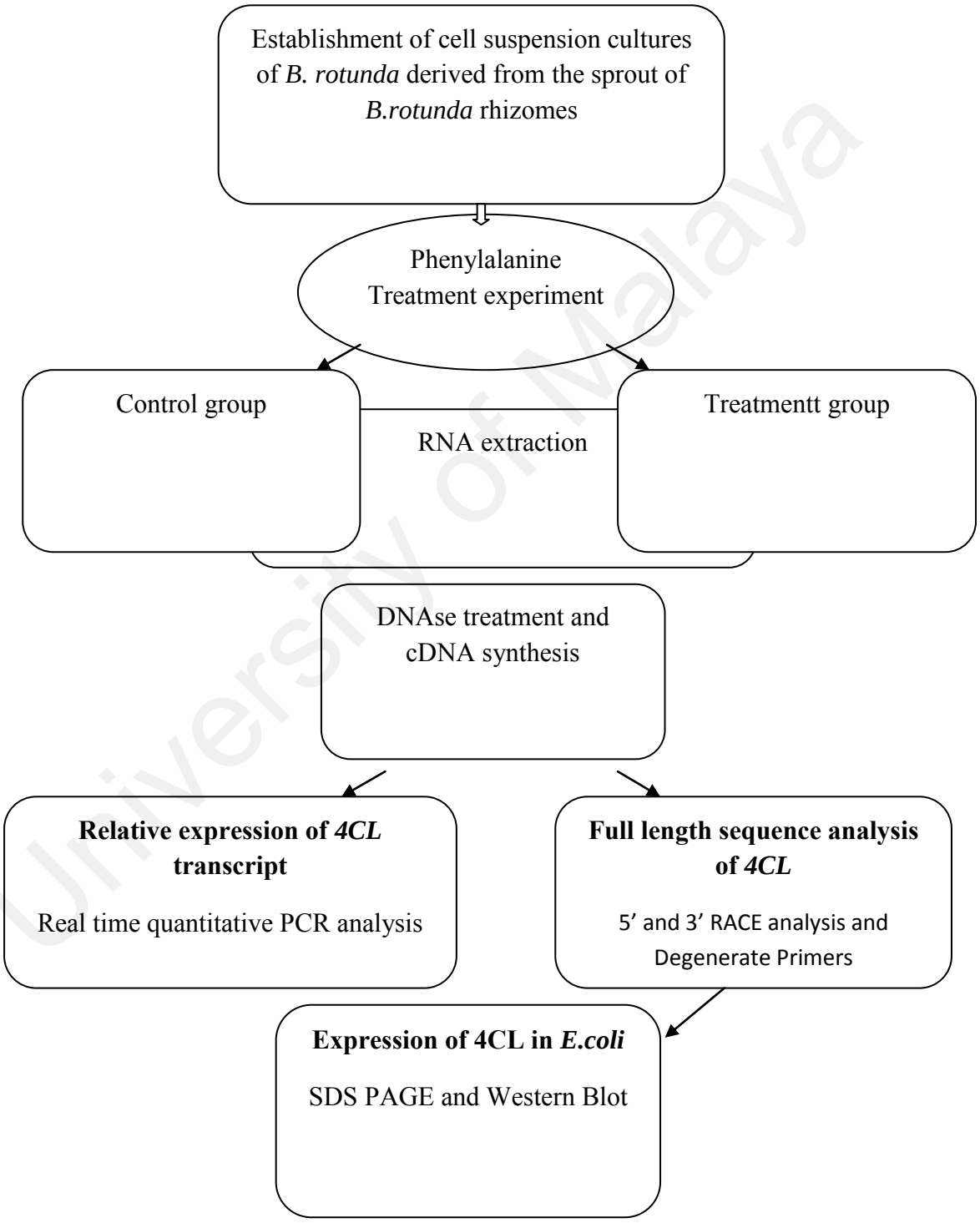


Figure 3.1: Flow chart of the project

3.2 Methods

3.2.1 Sterilisation

All glasswares, micropipette tips, microcentrifuge tubes, PCR tubes, Falcon tubes, stock solutions and buffers for DNA and RNA experiments were autoclaved at 121^0C with 15 psi for 20 minutes and 45 minutes respectively. The culture media for bacterial and plant tissues/cells were autoclaved at 121^0C with 15psi for 20 minutes.

For explant sterilization, sprouts were excised from the rhizome and were placed into 20 % (v/v) Clorox solution (active ingredient 5.25 % (v/v) sodium hypochlorite) with 2 drops of tween-20 detergent. The sprouts were then washed with 70% (v/v) ethanol solution and subsequently washed three times with distilled water prior to splicing the basal part of the sprout. In order to ensure sterility, explant sterilisation, handling of plant tissue and bacterial cultures were all carried out in a laminar flow (GelmanSciences). Prior to use, UV light was switched on in the laminar flow for 15 minutes, followed by swapping of the surface of the laminar air flow and containers with 70% ethanol. All culture tools were placed in a hot bead steriliser (250 ^{0} C) and cooled at room temperature before being used.

3.2.2 Media preparation

There are three types of media that were prepared, namely the callus initiation media (CIM), propagation media (PM), and cell suspension media (SM). The chemical components for each media were described in details in **Appendix A1**. The pH for all media was then adjusted to 5.8 using 1M NaOH prior to autoclaving. SM was prepared as liquid media whereas PM and CIM were solidified with 2 % gel rite (**Appendix A1**). All media consisted of Murashige and Skoog (MS) basal medium and supplemented with different types and concentration of plant growth regulators.

3.2.3. Callus initiation

Basal part of the sprouts was sliced horizontally, of about 5 mm thickness before placing onto the agar plate (0.5 cm) (**Figure 3.2**) containing MS medium for callus formation (CIM). After 4 weeks, the callus was formed and a clump of callus was transferred into propagation media (PM) for establishment of cell suspension media. Callus cultures were placed in a culture room maintained at 25 ± 2°C under a dark condition.

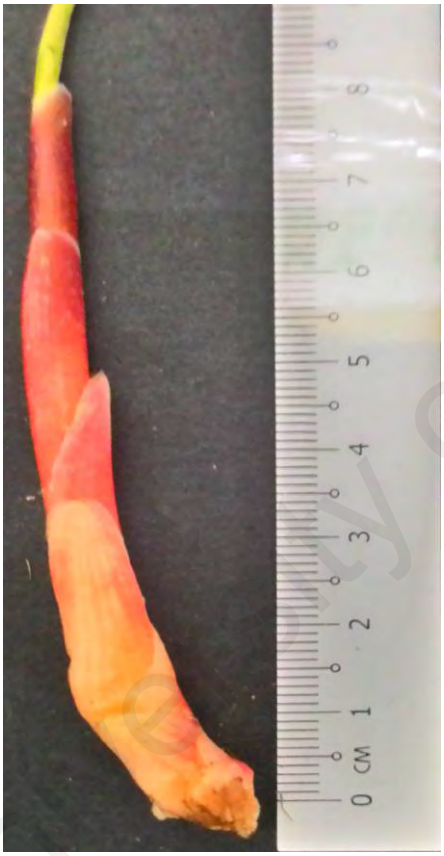


Figure 3.2: Sprout excised from rhizome of *B.rotunda* plant

3.2.4 Establishment of cell suspension cultures of *Boesenbergia rotunda*

Two grams of the callus which were cultured on PM media was added to 50 ml of the suspension medium (SM) in 250 ml Erlenmeyer flasks. The cells were propagated for one month. A sterile stainless steel sieve, with pore size 450 µm, was used to separate the fine cells from debris and clumpy cells. When homogenous fine cells were formed, subculturing was done on a routine basis at 14-day intervals and maintained in suspension medium (SM). During subculture, 10 ml of the conditioned media was re-used with an addition of 40 ml new SM media. The cell cultures were placed on an orbital shaker (New Brunswick Scientific, USA) at 80 rpm in a culture room maintained at 25 ± 2°C under a continuous 16/8 light/dark condition.

3.2.5 Phenylalanine treatment on cell suspension of *Boesenbergia rotunda*

The concentration of phenylalanine used was 0.04mg/l for each treatment. In preparing this concentration, 2 ml of 1mg/ml of phenylalanine was added into 50 ml suspension cultures containing 3 ml settled cell volume (SCV) of *B.rotunda* cells. Subsequently, cell suspension cultures were placed onto shakers (80rpm) and were incubated for 10 (T10) and 14 (T14) days. Cultures were harvested on day 10 and 14 by filtering the culture medium through a 9cm Whatman filter paper (USA) and washed three times with distilled water. As for control, phenylalanine was replaced with sterile distilled water and cells were harvested at 3 phases, in 0 day (C0), after 10 days (C10) and after 14 days (C14). Three biological replicates were used per treatment. The cells were weighed, frozen in liquid nitrogen and kept at -80°C for storage and subsequent use for RNA extraction.

3.2.6 Relative quantification analysis via Reverse-transcriptase PCR

3.2.6.1 RNA extraction

Total RNA was isolated from cell suspension cultures using CTAB method as described by Liu *et al* (1998). About 0.2 g of ground samples were harvested into 2.0 ml microcentrifuge tube. The samples were mixed with 1ml of CTAB extraction buffer [2% (w/v) CTAB, PVP-40, 100 mM Tris-Cl, pH 8.0, 25 mM EDTA, 2 M NaCl, 0.6% (v/v) β-mercaptoethanol]. Equal volume of extraction solution (chloroform: isoamyl alcohol, 24:1) were mixed in the tubes and the tubes were inverted for 2 minutes. After that, the tubes were centrifuged at 13,000 x g for 15 minutes at room temperature. Clear supernatants were transferred into each new tube. Equal volume of extraction solution was mixed in each tube and the steps were repeated. Approximately 0.5 ml of lithium chloride was mixed in each tube to a final concentration of 2 M and the mixtures were incubated overnight in 4^0C refrigerator. The overnight mixtures were centrifuged at 13,000 × g for 30 minutes at 4^0C. The pellets were re-suspended with 1 ml of NTES buffer (1 M NaCl, 10 mM Tris-HCL, pH 8.0, 1 mM EDTA, 0.5% (w/v) SDS). Equal volume of extraction solution was mixed and the mixtures were centrifuged at 13,000 × g for 25 minutes at 4^0C. The supernatants were transferred into each new 1.5 ml microcentrifuge tube. Equal volume of ice-cold isopropanol was mixed and the mixtures were precipitated overnight at -80^0C laboratory freezer (News Brunswick scientific). The overnight mixtures were centrifuged at 13,000 × g for 30 minutes at 4^0C. The supernatants were decanted and the pellets were air-dried. The RNA pellets were dissolved with 100 µl double distilled water and stored in -80^0C laboratory freezer.

Using a UV Biophotometer (Eppendorf, Germany), the optical density (OD) at 260 and 280nm were determined. The purity and concentrations of RNA of the samples were analysed based on the ratio of $OD_{260/280}$. For measuring RNA concentration, 4 µl of samples were mixed with 96 µl double distilled water. The diluted samples were

measured at wavelength of 260 nm and 280 nm. The readings were taken against a blank of 100 µl double distilled water.

3.2.6.2 DNAse treatment and Reverse-transcriptase PCR

After measuring the RNA concentration, DNAse I (Invitrogen) treatment was carried out on the RNA samples to remove the remaining DNA. The reaction mixture was prepared as showed in **Table 3.1**

Table 3.1: DNase I treatment reaction mixture

Component	Working concentration	Volume (µl)
RNA sample	1 µl	Concentration dependent
10x DNaseI reaction buffer	1x	1
DNaseI (1U/ µl)	1U	1
DEPC-treated H_2 O	-	10

The RNA extract was incubated with RNase-free DNase I (1 U/µl) (Invitrogen, U.S.A) at 37°C for 10 minutes. The reaction was inactivated with 1 µl of 25 mM EDTA solution and heated at 65^0 C for 10 minutes. Then, agarose gel electrophoresis (**Appendix A2**) was carried out on the treated RNA samples. DNAse treatment was performed to remove DNA molecules prior to Random Amplification cDNA Ends (RACE) and for Real-time quantification Polymerase Chain Reaction (RT-qPCR). RNA quality and concentration were checked by ethidium bromide (EtBr)-stained agarose gel electrophoresis and a UV Biophotometer (Eppendorf, Germany), respectively.

For reverse transcription experiment, the first strand of cDNA was synthesized using a cDNA synthesis Kit (EURx Ltd, Poland). The reaction mixture was prepared as showed in **Table 3.2.** The thermal cycling condition of Reverse-transcriptase PCR is as follows;

25^0C	10 minutes
50^0C	30 minutes
85^0C	5 minutes
4^0C	∞

The concentration of cDNA was quantified using a NanoDrop ND 1000 spectrophotometer (Thermoscientific, USA). The synthesized cDNAs (1 µl) from 5 selected RNA samples were assessed for their purity and concentration using a Nanodrop.

Table 3.2: Reverse-transcriptase PCR reaction mixture

Component	Volume in 1x reaction (µl)
5x cDNA buffer	4
5mM dNTP mix	4
Random hexamers primers	1
dART Reverse Transcriptase.	1
RNase Inhibitor 12.5 U/µ	1
RNAse-free water	to 13
RNA (10ng- 5µg)	-
Total	20

3.2.6.3 Real-Time PCR

Real-Time quantification Polymerase Chain Reaction (RT-qPCR) was performed in a 7500 Fast Real-Time PCR System (Applied Biosystems, U.S.A) using the SYBR Green mixture (Applied Biosystem, U.S.A). Each 12.5 µl reaction mixture contained 6.25 µl of SYBR Green mixture (Applied Biosystem, U.S.A), 10 µM (each) specific primers and 100 ng of template of cDNA. The amplification of each sample was performed in triplicate. The reaction mixtures were prepared as mentioned in **Table 3.3**. The amplification conditions were as follows: 95°C for 10 minutes, followed by 40 cycles of 95°C for 15 seconds and 60°C for 1 minute. The primers *4RTF* (5'- GCC CCC TCT GGA AAA ATC TT-3') and *4RTR* (5'-TGC CAC CAT GAG AGC ACT ATG-3) were used to amplify a 125 bp *Br4CL* fragment (**Table 3.4**). A 100 bp fragment of *B. rotunda* actin gene was amplified as an internal control for normalization using primers: *ARTF* (5'-GCC TCA CGC TCT TCT TTC GAT- 3') and *ARTR* (5'-AGC AGT GGT GGT GAA TGA ATC TC- 3') (**Table 3.4**). The target and the endogenous primers specificity were performed with dissociation program provided in a 7500 Fast Real-Time PCR system (Applied Biosystem, U.S.A). The primer set with single peak indicates the specificity of primer. Four replicates of each primer pair and a negative control (without cDNA template) were prepared and assayed with the following dissociation program;

30 cycles	95^0 C for 15 seconds
1 cycle	60^0 C for 1 minute
1 cycle	95^0 C for 15 seconds

Data of Relative Quantification plates were analyzed using Relative Quantification Study (SDS software 1.3.1) (Applied Biosystems, U.S.A) for comparing the gene expression level between the samples.

Table 3.3: Reaction mixture for RT-qPCR

Component	Volume for standard 96 well plate (µl)
Power SYBR green PCR Master mix 2X	6.25
10 µM Forward primer of 4CL(200 nM)	0.625
10 µM Reverse primer of 4CL(200 nM)	0.625
Nuclease free water	4.22
cDNA template (100 ng/µl)	0.78
Total volume	12.5

3.2.7 Amplification of full length *4CL* gene from *B.rotunda*

Two methods were used to amplify the coding region of 4CL from *B.rotunda*, which were Rapid Amplification of cDNA Ends (RACE) and Polymerase Chain Reaction (PCR) using degenerate primers. The sequence of primer sets for both methods were shown in **Table 3.4.**

3.2.7.1 Rapid Amplification of cDNA Ends (RACE) method

The amplification of 5' and 3' cDNA ends were performed using a RACE kit (Invitrogen, U.S.A) according to the manufacturer's protocol. The 5' and 3' gene specific primers were designed from *4CL* Unigene36813_All from transcriptome data (**Appendix C14**). The transcriptome data was obtained from previous transcriptome

profiling data of phenylalanine-treated *B.rotunda* of cell suspension (Diyana et al., 2015). This technique is based on RNA ligase-mediated (RLM-RACE) and oligo-capping rapid amplification of cDNA ends (RACE) methods, and results in the selective ligation of an RNA oligonucleotide to the 5′ ends of decapped mRNA using T4 RNA ligase. The raw material for this experiment was RNA molecules that have undergone DNAse treatment. Before RNA samples were converted into RACE-ready cDNA, the RNA samples underwent 3 types of treatment for 5' RACE. Initially, a dephosphorylated reaction was done to remove the 5′ phosphates. This eliminates truncated mRNA and non-mRNA from subsequent ligation with the GeneRacer™ RNA Oligo. Subsequently, the mRNA cap structure was removed followed by ligation reaction using T4 RNA ligase in 10 µl reaction mix containing ligase buffer, ATP, and 40 unit of RNAseOUT^{TM}. The reaction was carried out at 37^{0}C for an hour. RNA was then retrieved by phenol: chloroform precipitation according to manufacture manual. Finally, RNA pellet was resuspended in 10 µl of distilled water and 1 µl sample was taken to be analyzed by agarose gel electrophoresis. For cDNA conversion, 10 µl RNA ligated with GeneRacer^{TM} RNA oligo was added to 1 µl of random primers, 1 µl of dNTP mix and 1 µl of nucleasefree-H_20. The mixture was incubated at 37^{0}C for 5 minutes to remove any RNA secondary structures and chilled on ice for 1 minute. High capacity RT master mix containing buffer, reverse transcriptase, DTT and RNAse OUT^{TM} were then added to the mixture and was briefly centrifuged to bring all the ingredients to the bottom of the tube. Reverse transcription was carried out in a thermocycler with the following incubation conditions: 25^{0}C for 5 minutes, 50^{0}C for 60 minutes, 70^{0}C for 15 minutes and hold at 4^{0}C. RNase H (2U) of 1 µl was added into the reaction mixture and incubated at 37^{0}C for 20 minutes.

For 3'RACE, RNA was reverse transcribed using GeneRacer^{TM} oligo dT to replace GeneRacer^{TM} RNA oligo with the same conditions as 5' RACE. RACE-ready

cDNA with known priming sites on each end were used to amplify the 5' end and 3' end using HotStarTaq *Plus* DNA polymerase (Qiagen, U.S.A). The components and the volume of PCR reaction were set as mention in **Table 3.5.** The PCR cycle conditions were as follows: initial denaturation at 95°C for 5 minutes, followed by 34 cycles of 94°C for 1 minutes, 50°C for 50 seconds, 72°C for 1 minute followed by a final extension at 72°C for 10 minutes. The amplified PCR products were analyzed by electrophoresis on EtBr-stained agarose gel. Purification of the excised gel fragments were carried out using a Gel extraction kit (Qiagen, U.S.A) according to the manufacturer's protocol before sending it for sequencing. Sequencing was done by the NextGene Sdn Bhd Company by preparing 10 µM of primers (forward and reverse) and 10 µl of sample.

3.2.7.2 Multiple alignments and designing degenerate primer

A total of 6 proteins encoding 4CL enzymes, namely *Zea mays* (NM_001111788.1), *Panicum virgatum* (ADZ96250.1) *Sorghum bicolour* (XP_002451647.1), *Lolium perenne* (AAF37732.1), *Hibiscus cannabinus* (ADK24217.1), and *Galtonia saundersiae* (KF241990.1) were used to design degenerate primers. After extracting the sequences (amino acids and nucleotides) from NCBI database, the amino acid sequences were aligned using CLUSTALW Omega programme (**Appendix C7**). Three sets of primers were used namely Degenerate outer forward (*DOF*), Degenerate outer reverse (*DOR*), Degenerate inner forward (*DIF*), Degenerate inner reverse (*DIR*) and Specific 4CL reverse (*SR*) to obtain the coding region of *4CL* gene in *B.rotunda.*

Table 3.4: Primers used for the amplification of 4CL coding region from *B.rotunda*

Primer Set	Sequence
3' RACE	*3RF* : AGCGAAGCTAGCTGAGCAAT
	3RR:GCTGTCAACGATACGCTACGTAACGGCATGACAGTG$(T)_{24}$
5' RACE	*5RF*: GGACACTGACATGGACTGAAGGAGTA
	5RR: CTCTTGTAGAAAACCACCTGTC
Degenerate outer	*DOF*: TST TYCGKTCSARKCTBCCVGACAT
	DOR: TCYTTCCKB ARR RAT CTTGCC NGANGG
Degenerate inner	*DIF*: GTVGTSGCSCTKCCNTWCTCVTCS
	DIR: T CRTYRTCRTCRTCVAYRTANCCRA
Degenerate outer forward	*DOF*: TSTTYCGKTCSARKCTBCCVGACAT
Specific reverse	*SR*: TGGGGTTTTCGCCGTCCACTTGC
4CL real time primers	*4RTF*: GCCCCCTCTGGAAAAATCTT
	4RTR: TGCCACCATGAGAGCACTATG
Actin real time primers	*ARTF*: GCCTCACGCTCTTCTTTCGAT
	ARTR: AGCAGTGGTGGTGAATGAATCTC

Table 3.5: Mastermix 7x PCR reaction mixtures

Components	Volume/reaction (µl)	Final concentration
10x CoralLoad PCR buffer (with 15mM $MgCl_2$)	10	1x
dNTP mix (5mM of each)	4	200µM of each dNTP
Downstream primer (10µM)	5	0.5 µM
Upstream primer (10µM)	5	0.5 µM
Template	3	$\leq$ 1µg/reaction
HotStarTaq *Plus* DNA polymerase (2.5 U/µl)	0.5	2.5 units/reaction
Sterile distilled H_2O	72.5	-
Total reaction volume	100	-

Note: 10 µl was taken out from the mixture for negative control before adding 3 µl of cDNA as a template and 15 µl were loaded into 6 PCR tubes.

3.2.7.3 Overlapping sequences analysis

Sequences of PCR product obtained from RACE method and degenerate primers were analyzed using BioEdit Sequence Alignment Editor Version 7.2 software to detect the overlapping region between the sequences obtained. The chromatogram peak of nucleotide sequences were viewed through FinchTV 1.4 software.

3.2.7.4 BLAST analysis

Sequences were analyzed using Blastx where the translated nucleotide query was used in searching for protein database (blast.ncbi.nlm.nih.gov/Blast).

3.2.8 MEGA 5.2 analysis

Phylogenetic tree was constructed using MEGA 5.2 software program. Thirty sequences of 4CL enzymes including 4CL from *B. rotunda* were extracted to construct the trees, including 4CL isoforms from Arabidopsis, aspen and *Musa acuminata subsp malaccensis*. Neighbour joining method from the software was selected to construct the trees. This method will reveal the evolutionary relationship of *B.rotunda*'s 4CL enzymes with other related plants. In addition, this method was fast, due in part to its being a polynomial-time algorithm. Neighbor joining is basically a clustering algorithm that clusters haplotypes based on genetic distance. All positions containing gaps and missing data were eliminated from the data set.

3.2.9 Expression of 4CL in *E.coli*

3.2.9.1 Vector construction and Gene synthesizing

A vector plasmid namely pET-28a was used to express open reading frame of *4CL* gene in *E. coli* strain BL2DE3. Codon optimization was carried by the NextGene Sdn Bhd Company. The finalised sequence coding open reading frame of *4CL* gene was confirmed by blastx search and ExPASy analysis and was sent to NextGene Sdn.Bhd Company to be synthesized. The synthesized coding region of *4CL* gene was cloned into pET-28a expression vector and expressed in *E. coli* strain BL2DE3 (**Figure. 3.3**). *4CL* gene was tagged with histidine sequence. The sequence was inserted into the multiple cloning site of pET-28a vector by using Nde1/Xho1 restriction enzymes.

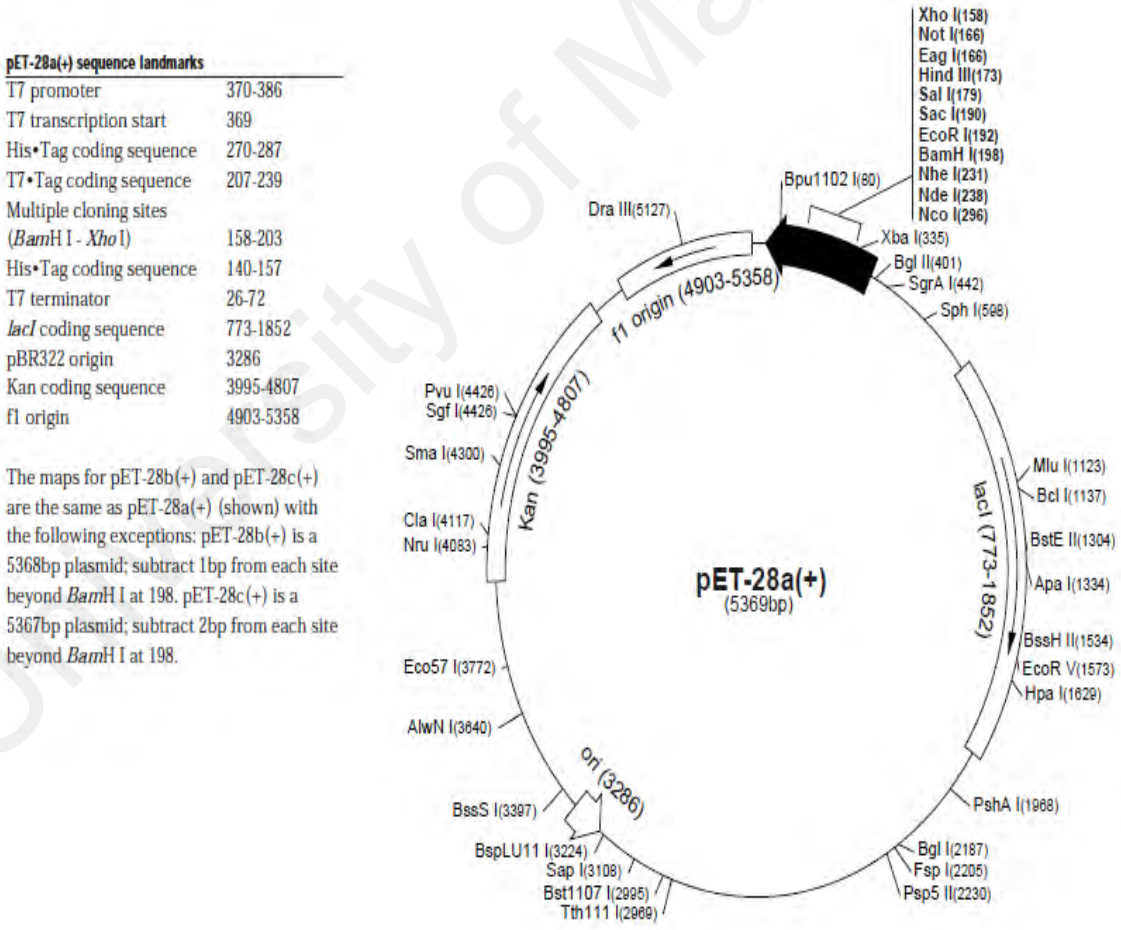


Figure 3.3: Expression vector of pET-28a where the synthesized *4CL* gene of *B. rotunda* was inserted into the multiple cloning sites with His-tagged at the N terminal of the gene.

3.2.9.2 Protein expression

Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE) and Western blot analysis were used to verify the expression of the recombinant proteins. Bacterial cultures containing plasmids with target insert (pET28-4CL-sequence) supplied by the company were streaked on the surface of a luria bertani (LB) agar plate containing an antibiotic, 25 µg/ml kanamycin. The plates were incubated overnight at 37°C. A single isolated bacterial colony from freshly streaked plates was used to inoculate 25 ml liquid LB medium containing the same concentration of the antibiotic. Culture was grown overnight on a shaker at 250 rpm at 37 ^{0}C. One ml of overnight culture was transferred into 25 ml liquid LB medium having the same concentration of kanamycin. Culture was incubated while shaking at 150 rpm at 37 ^{0}C. Once the cultures reached OD 0.4 - 0.5 at 600nm, recombinant protein expression was induced by the addition of isopropyl -ß-D-thiogalactopyranoside (IPTG), and the culture was grown for another 4 hours at 37 ^{0}C at a shaking speed of 150 rpm. Cells were harvested by centrifugation at 13000 rpm for 10 minutes at 4°C. Pellets were resuspended in lysis buffer before loading into SDS PAGE gel. A total of 20 µl of sample was loaded into the gel with some loading dye. The gel was electrophoresed at 120 V for 70 minutes. Following that, the gel was put in a plastic container with Coomasie Blue Staining solution and placed on a Rocker Gyrotory 30 (Stuart Scientific, U.K) at 20 rpm .The gel was placed at room temperature for about 2 hours. The gel was destained with destaining solution for three times to wash excessive blue stain until bands were visible.

The PAGE gel was prepared by preparing 4 % of stacking gel and 12% of separation gel as shown in **Table 3.6** and **3.7**.

For western blot, the analysis was also done by outsourcing to Next Gene Sdn Bhd Company. Western blot is often used in research to separate and identify proteins. In this technique a mixture of proteins was separated based on molecular weight

through gel electrophoresis. These results were then transferred to a membrane producing a band for each protein. The membrane was then incubated with labels antibodies specific to the protein of interest. The unbound antibody was washed off leaving only the bound antibody to the protein of interest. The bound antibodies were then detected by developing the film. As the antibodies only bind to the protein of interest, only one band should be visible.

Table 3.6: SDS-PAGE Separating gel (12 %)

Components	Volumes required (µl)
H_2 O	1650
1.5 M Tris	1250
10 % SDS	25
Acrylamide-Bis (30%)	2000
APS (10%)	75
TEMED	10

Table 3.7: SDS-PAGE Stacking gel (4 %)

Components	Volumes required (µl)
H_2 O	1700
1 M Tris	315
SDS (10%)	25
Acrylamide (30%)	415
APS (10 %)	12.5
TEMED	10

CHAPTER 4: RESULTS

.4.1 RNA extraction and DNAse treatment

RNAs isolated from phenylalanine-treated cells at days 10 and 14 (T10 and T14) and untreated cells at days 0, 10 and 14 (C0, C10, and C14) were assessed for purity and concentration. The isolated RNA exhibited good purity within a range of 2.0 (at wavelength 260nm/280nm), indicating the minimal contamination from proteins and other contaminants (**Tables 4.1 and 4.2**) while secondary measurement (wavelength 260nm/230nm) was higher than 260nm/280nm respective value indicating minimal presence of contaminants. Lower than 2.0 (260nm/230nm) may indicate the presence of contaminants which absorb at 230 nm. Besides, distinct 28 S and 18 S rRNA bands were also observed on 1% agarose gel (**Figure 4.1)** indicating the presence of RNA. Although DNA residual might still be present in the RNA extract, the DNAse-treated RNA showed a reduction of DNA band intensity. The samples were further treated with DNAse to ensure no contamination from genomic DNA.

Table 4.1: Concentration and purity of RNA samples

Sample	Concentration (µg/mL)	Purity (260nm/280nm)	Purity (260nm/230nm)
C10 (1)	26.8	1.92	2.67
C14 (2)	32.0	1.93	2.23
T10 (3)	35.7	1.93	2.99
T14/ (4)	27.8	1.90	2.97

Table 4.2: Concentration and purity of the RNA samples after DNAse treatment

Sample	Concentration (ng/µL)	Purity (260nm/280nm)	Purity (260nm/230nm)
C10 (A)	0.2	2.18	1.90
C14 (B)	0.2	2.15	2.18
T10 (C)	0.2	2.17	2.09
T14 (D)	0.1	2.20	1.57

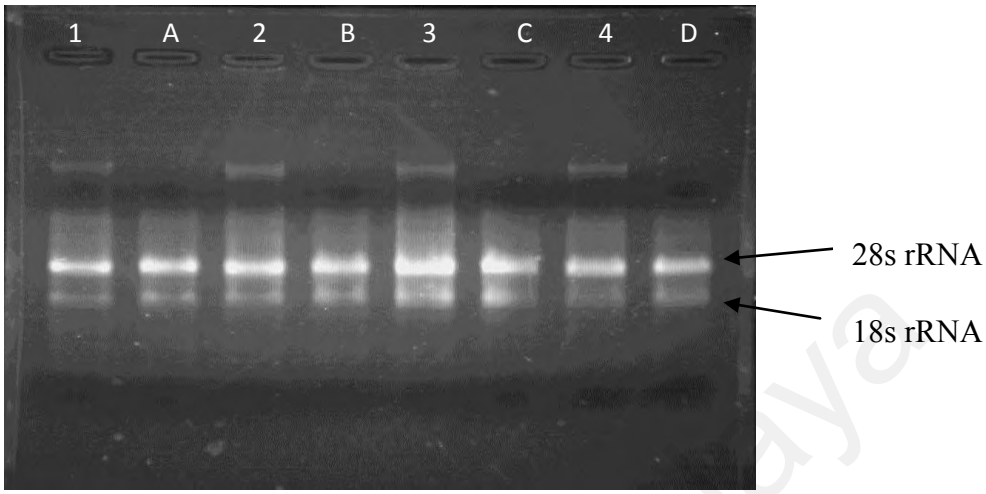


Figure 4.1: Typical gel electrophoresis of RNA isolated from *B.rotunda* cell suspension cultures before and after DNAse treatment. Lanes 1, 2, 3 and 4 referring to samples before DNAse treatment while lanes A, B, C, and D referring to samples after DNAse treatment

Lane 1-A: sample C10

Lane 2-B: sample C14

Lane 3-C: sample T10

Lane 4-D: sample T14

For sample C0, the gel electrophoresis image of of RNA and DNAse treated RNA samples from control group at 0 day was shown in **Appendix E.** Five RNA samples from each group (C0, C10. T10, C14, T14) were selected for real-time PCR analysis based on the integrity and their purity after DNAse treatment. The selected samples are indicated in (**Appendix E**).

4.2 cDNA synthesis

The synthesized cDNAs (1 µl) from 5 selected RNA samples were assessed for their purity and concentration using a Nanodrop (**Table 4.3**). The purity of all selected cDNAs was within acceptable range by the 260 nm/280 nm ratio (range 1.8-2.0).

Table 4.3: Concentration and purity of cDNA samples

Sample	Concentration (ng/µL)	Purity (260 nm/280 nm)	Purity (260 nm/230nm)
C0	1337.7	1.77	1.90
C10	1296.0	1.79	1.92
T10	1279.0	1.79	1.91
C14	1273.5	1.78	1.95
T14	1265.3	1.79	1.95

4.3 Expression analysis by Real-time PCR

Prior to RT-qPCR analysis, a dissociation curve for the house keeping gene, *actin* and targeted gene, *4CL* was performed. The dissociation curves for both *4CL* and actin genes showed a single peak, indicating the specificity of the designed primers and confirming only one PCR product being produced (**Figures 4.2 and 4.3**).

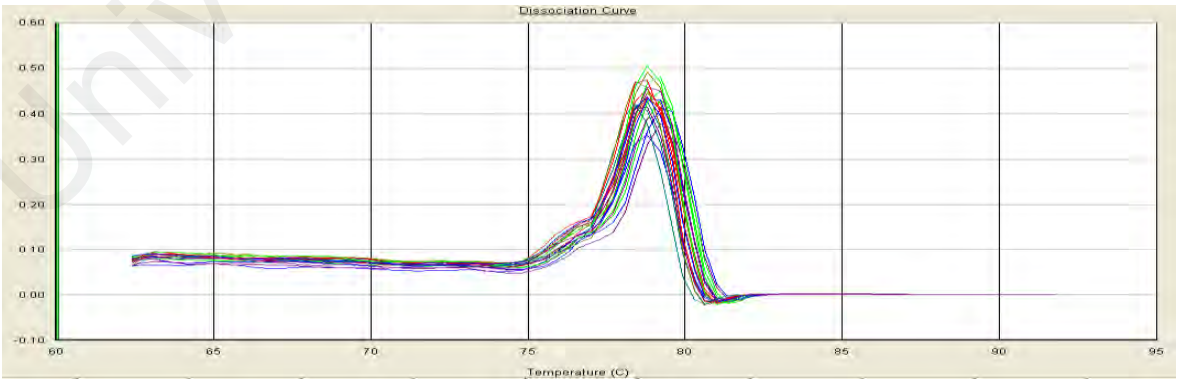


Figure 4.2: Dissociation curve for 4CL PCR product. 4CL dissociation melting point: 78°C

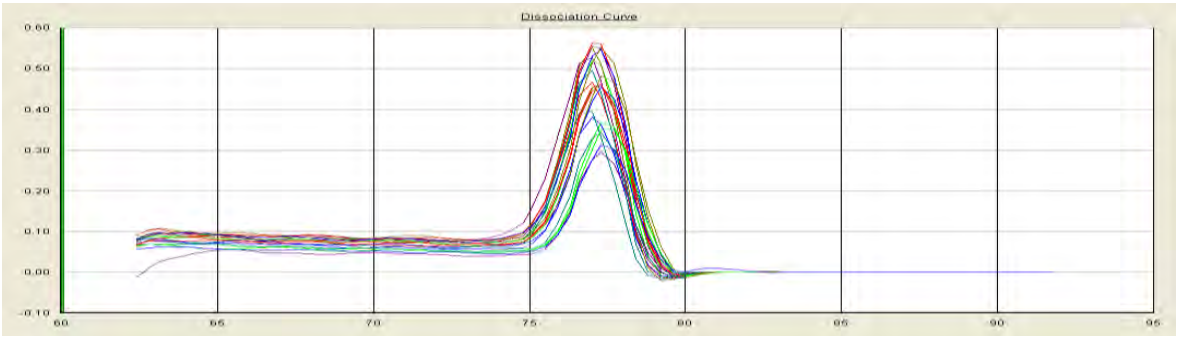


Figure 4.3: Dissociation curve for actin PCR product. Actin dissociation melting point: 76°C

The *4CL* gene expression of *B. rotunda* cells treated with (treatment group) or without phenylalanine (control group) was investigated over 14 days. A clear separation of amplification plot between different groups of sample showed that there were differences of expression level between the samples (**Figure 4.4**)

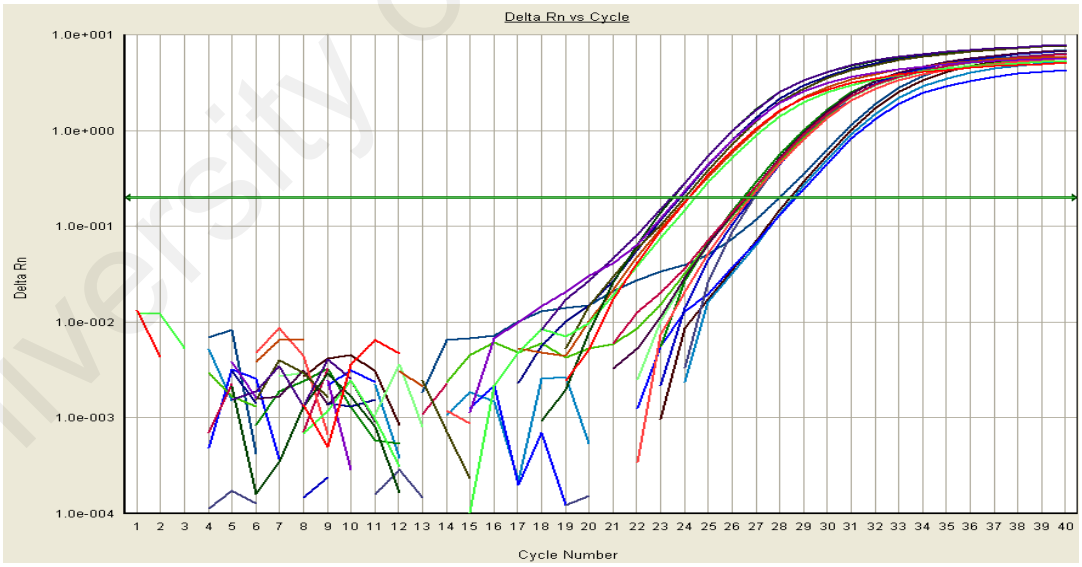


Figure 4.4: Amplification plot of cDNA sample of C10/3-2, C14/1-5, T10/1-3 and T14/3-1

The results showed that 10-days old (C10) and 14-days old (C14) control cultures were 7- and 8- fold, respectively, higher than cell cultures at 0 day (C0) (**Figure 4.5**). When cells are treated with phenylalanine, the transcription levels of *4CL* in 14-days old (T14) and 10-days old (T10) cultures were increased 22- and 17-fold,

respectively, compared to C0 (**Figure 4.5**). T10 and T14 cultures exhibited 2- and 3-fold, respectively, greater than their relative controls (C10 and C14). These findings further confirm the promotive effect of phenylalanine in enhancing the expression level of *4CL* gene.

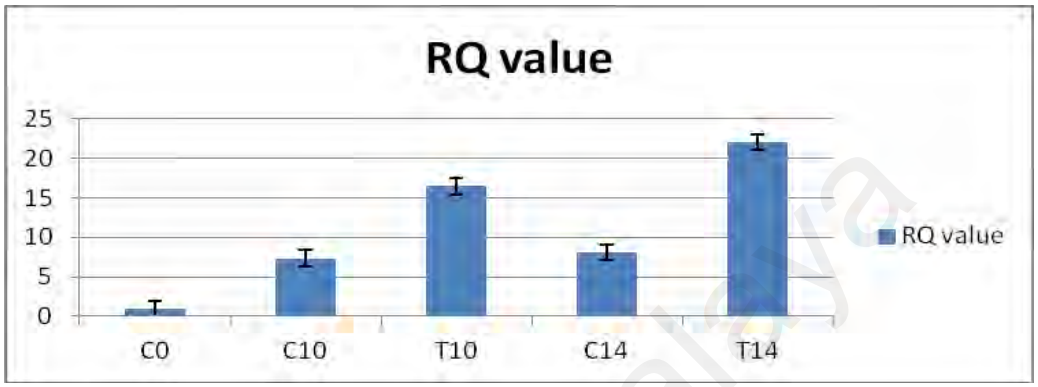


Figure 4.5 : The expression levels of *4CL* in *B. rotunda* cell suspension cultures after being treated with phenylalanine at days 10 and 14. Bars represent the standard deviation of three biological replicates. RQ value: Relative quantification expression value on *Y* axis represents *Br4CL* transcripts level relative to actin. The value in C0 (Control at 0 day) was set to 1

4.4 Full length analysis of *Br4CL* gene

The verification of the Unigene36813_All existence was done first before amplifying the full length of *4CL* from *B.rotunda* using PCR reaction (**Appendix C15).** Two methods have been adopted in this study to amplify the full length of *Br4CL* in **section 3.2.4** i.e RACE-PCR and Degenerate primers-PCR methods.

4.4.1 RACE-PCR method

Amplification of 3' end region of the Unigene36813_All with primer (*3RF*) and reverse primer (*3RR*) result in 450bp fragment of PCR product (**Figure 4.6**). Sequencing results showed that the PCR product from 3' RACE amplification was at the 3' end region of initial reference sequence of *4CL* in *B.rotunda* that were extracted from transcriptomic database (Unigene36813_All) **(Appendix C14)** as shown by the presence of Poly-A sequences (**Appendix C1**). In addition, the Blastx results showed that the target sequence was 96% identical to *4CL* 1-like *Musa acuminata subsp.malaccensis* (accession number : XP009414743 .1) with the coverage of 23% (**Figure 4.7**).

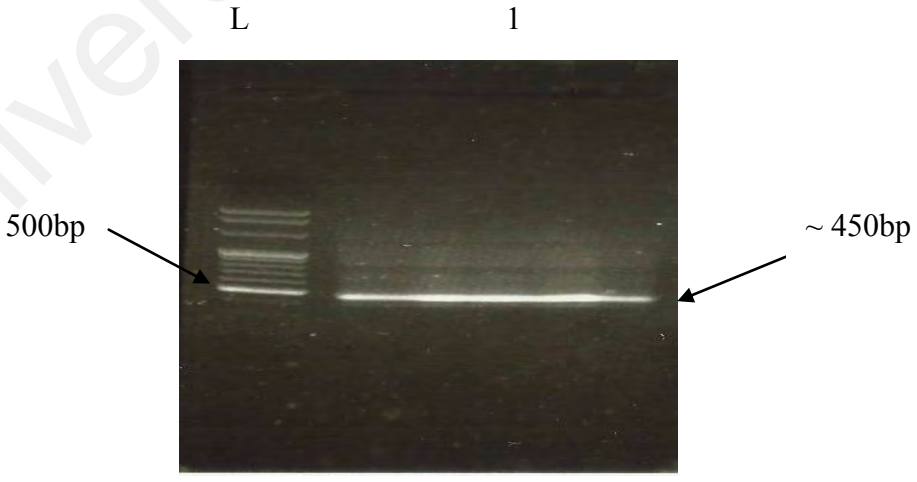
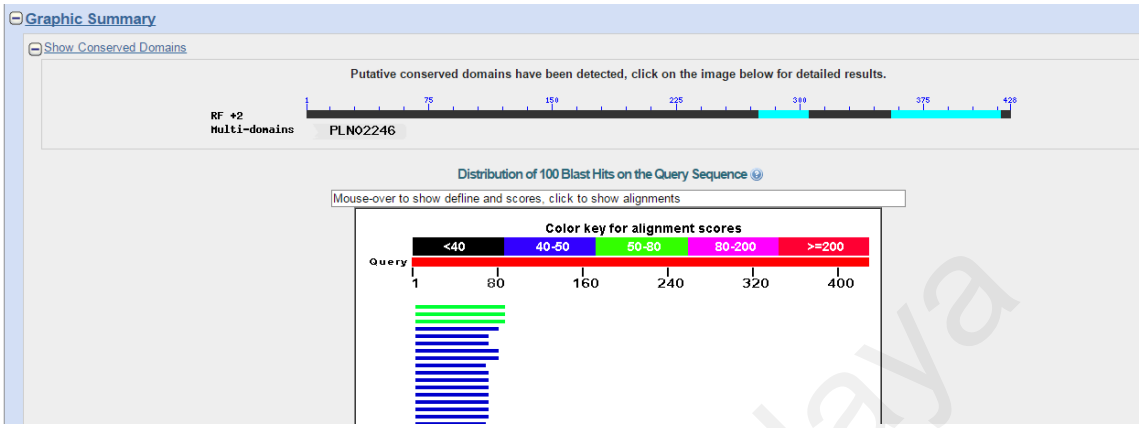
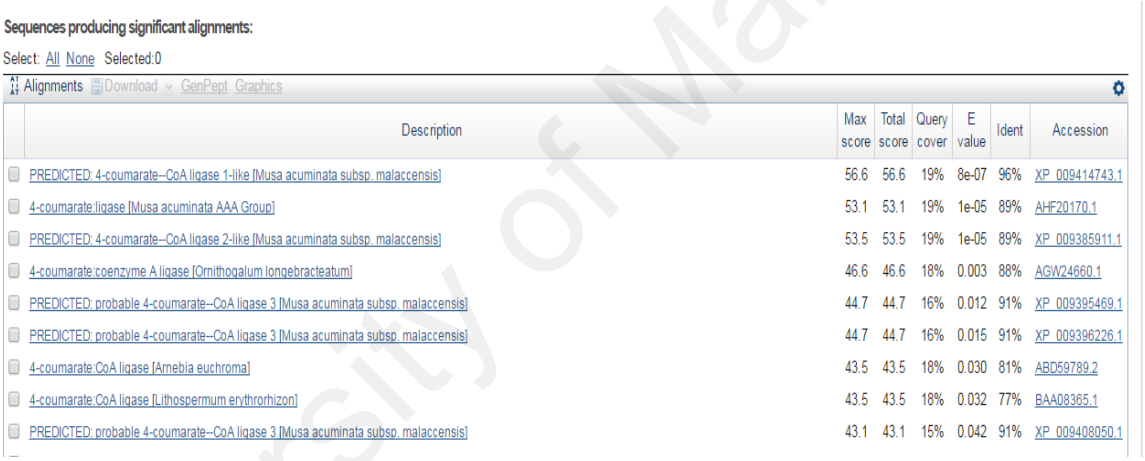


Figure 4.6: Typical gel electrophoresis of PCR product for 3' RACE experiment

L: Marker: 1kb DNA ladder (EUR_X)

Lane 1: PCR product





Sequences producing significant alignments:

Select: All None Selected:0

Alignments Download GenPept Graphics

Description	Max score	Total score	Query cover	E value	Ident	Accession
PREDICTED: 4-coumarate--CoA ligase 1-like [Musa acuminata subsp. malaccensis]	56.6	56.6	19%	8e-07	96%	XP_009414743.1
4-coumarate:ligase [Musa acuminata AAA Group]	53.1	53.1	19%	1e-05	89%	AHF20170.1
PREDICTED: 4-coumarate--CoA ligase 2-like [Musa acuminata subsp. malaccensis]	53.5	53.5	19%	1e-05	89%	XP_009385911.1
4-coumarate:coenzyme A ligase [Ornithogalum longebracteatum]	46.6	46.6	18%	0.003	88%	AGW24660.1
PREDICTED: probable 4-coumarate--CoA ligase 3 [Musa acuminata subsp. malaccensis]	44.7	44.7	16%	0.012	91%	XP_009395469.1
PREDICTED: probable 4-coumarate--CoA ligase 3 [Musa acuminata subsp. malaccensis]	44.7	44.7	16%	0.015	91%	XP_009396226.1
4-coumarate:CoA ligase [Arnebia euchroma]	43.5	43.5	18%	0.030	81%	ABD59789.2
4-coumarate:CoA ligase [Lithospermum erythrorhizon]	43.5	43.5	18%	0.032	77%	BAA08365.1
PREDICTED: probable 4-coumarate--CoA ligase 3 [Musa acuminata subsp. malaccensis]	43.1	43.1	15%	0.042	91%	XP_009408050.1

Figure 4.7: Blastx result showed sequences was 96% identical to *4CL* 1-like *Musa acuminata subsp.malaccensis* (accession number: XP009414743 .1) with the coverage of 23%

Amplification of 5' end region of Unigene36813_All with primer (*5RF*) and reverse primer (*5RR*) result in 250bp fragment of PCR product (**Figure 4.8**). Sequencing results of 5'RACE were able to overlap with Unigene36813_All (**Appendix C2).** In addition, the Blastx results from NCBI showed that the *4CL* target sequence was 90% identical to *Musa acuminata* AAA group (accession number: AHF 20170.1) with the coverage of 100% (**Figure 4.9).**

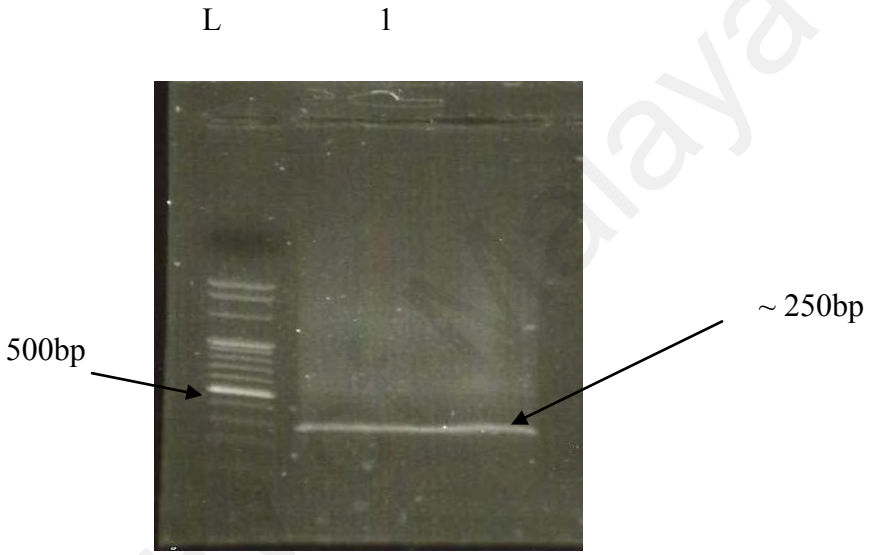
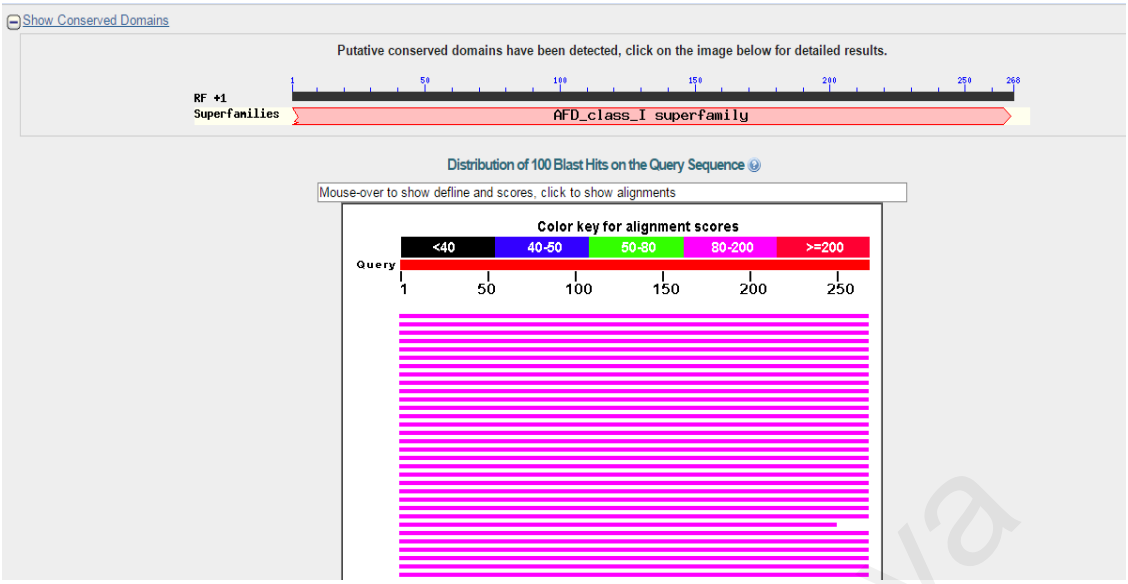


Figure 4.8: Typical gel electrophoresis of PCR product from 5' RACE experiment

L: Marker: 1kb DNA ladder (EUR_X)

Lane 1: PCR product



Sequences producing significant alignments:

Select: All None Selected:0

Alignments Download GenPept Graphics

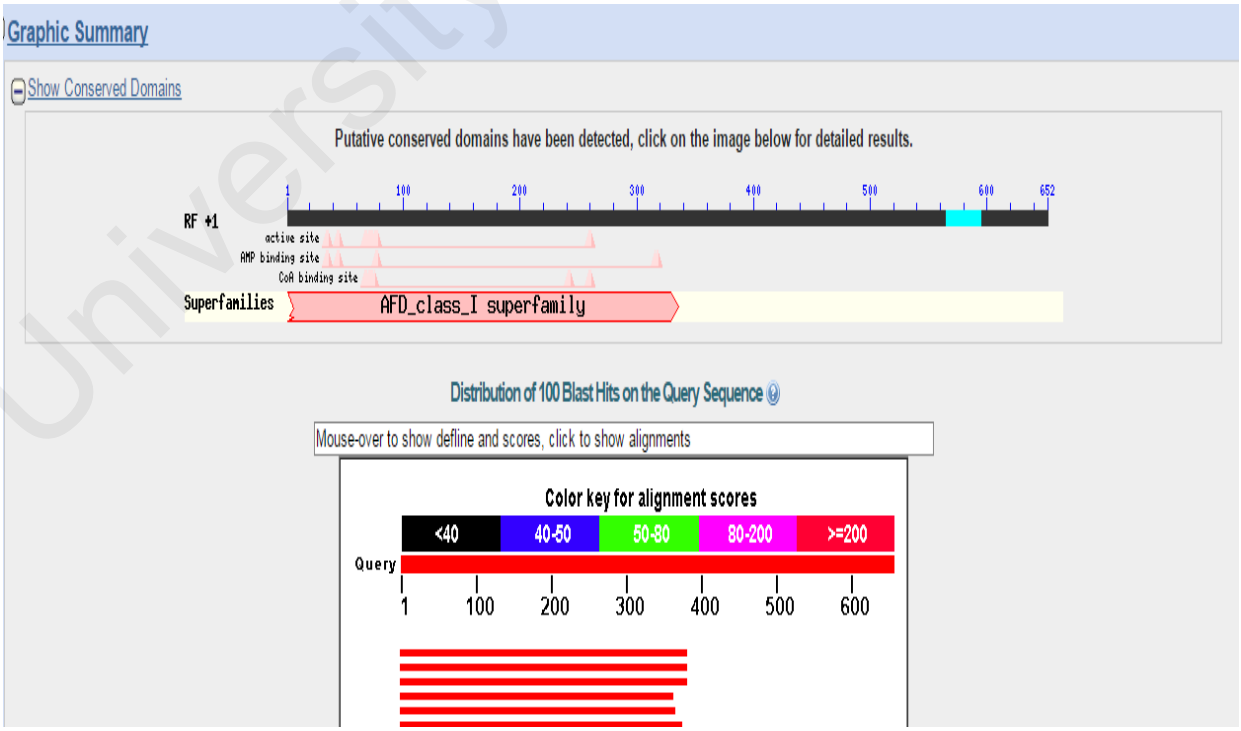
Description	Max score	Total score	Query cover	E value	Ident	Accession
4-coumarate:ligase [Musa acuminata AAA Group]	166	166	100%	2e-49	90%	AHF20170.1
PREDICTED: 4-coumarate--CoA ligase 1-like [Musa acuminata subsp. malaccensis]	169	169	100%	2e-48	91%	XP_009414743.1
PREDICTED: 4-coumarate--CoA ligase 2-like [Ziziphus jujuba]	169	169	100%	3e-48	88%	XP_015877593.1
PREDICTED: 4-coumarate--CoA ligase 2-like [Jatropha curcas]	169	169	100%	4e-48	88%	XP_012074070.1
PREDICTED: 4-coumarate--CoA ligase 2-like [Citrus sinensis]	168	168	100%	5e-48	89%	XP_006474168.1
hypothetical protein CICLE_v10007733mg [Citrus clementina]	169	169	100%	8e-48	89%	XP_006453393.1
PREDICTED: 4-coumarate--CoA ligase 1-like [Populus euphratica]	157	157	100%	1e-47	84%	XP_011007584.1
PREDICTED: 4-coumarate--CoA ligase 2-like [Musa acuminata subsp. malaccensis]	167	167	100%	4e-47	90%	XP_009385911.1
4-coumarate:coenzyme A ligase [Lycium ruthenicum]	158	158	100%	7e-47	83%	AHH86096.1
PREDICTED: 4-coumarate--CoA ligase 2-like [Pyrus x bretschneideri]	165	165	100%	7e-47	85%	XP_009347855.1
PREDICTED: 4-coumarate--CoA ligase 2-like [Pyrus x bretschneideri]	164	164	100%	2e-46	84%	XP_009334566.1
4-coumaroyl CoA ligase [Pyrus x bretschneideri]	164	164	100%	2e-46	84%	AKC34063.1

Figure 4.9: Blastx result showed sequences was 90 % identical to *4CL Musa acuminata* AAA group (accession number: AHF 20170.1) with the coverage of 100%

Initially, the raw sequence which is the size of Unigene36813_All was only 428bp in length. The 5' RACE experiment has successfully recovered around 228bp which is equal to 76 amino acids, while the 3' RACE experiment has recovered around 450bp which made the total recovered sequence to 650bp (**Appendix C3)**. Within 650bp, only 350bp was matched with 4CL putative superfamily, such as AMP binding

site, active site and CoA binding site, whereas the remaining 300bp might due to the unconserved region and specificity of *B. rotunda 4CL* sequence at 3' region **(Appendices C1 and C4)**. Blastx search showed that the sequences recovered from both 3' and 5' RACE experiment was 90% identical with *4CL* from *Musa acuminata* AAA Group (accession no.AHF20170.1) with query coverage of 100% (**Figure 4.10**). The deduced amino acid sequence recovered from the RACE experiment was then subjected to protein blast, BLASTP (Basic Local Alignment Search Tool). The results showed that several conserved domains belonged to 4CL superfamily such as AMP binding domain, AMP binding enzyme C-terminal and adenylate forming enzymes superfamily (**Figure 4.11**).

Therefore, it was postulated that the 3' region of the gene was completed but not for 5' region. Next, degenerate primers were designed by aligning multiple sequences of 4CL from other organisms to generate the full coding region of *4CL* gene.



Sequences producing significant alignments:

Select: All None Selected:0

Alignments Download GenPept Graphics

Description	Max score	Total score	Query cover	E value	Ident	Accession
4-coumarate:ligase [Musa acuminata AAA Group]	233	233	58%	3e-72	90%	AHF20170.1
PREDICTED: 4-coumarate--CoA ligase 1-like [Musa acuminata subsp. malaccensis]	240	240	58%	4e-72	92%	XP_009414743.1
PREDICTED: 4-coumarate--CoA ligase 2-like [Musa acuminata subsp. malaccensis]	233	233	58%	8e-69	90%	XP_009385911.1
4-coumarate:ligase [Musa acuminata AAA Group]	216	216	55%	1e-64	87%	AHF20171.1
hypothetical protein JCGZ_08679 [Jatropha curcas]	218	218	55%	1e-63	85%	KDP36410.1
4-coumarate:coenzyme A ligase [Galtonia saundersiae]	217	217	57%	2e-63	83%	AGW24660.1
PREDICTED: 4-coumarate--CoA ligase 2-like [Citrus sinensis]	217	217	54%	2e-63	88%	XP_006474168.1
hypothetical protein CICLE_v10007733mg [Citrus clementina]	218	218	54%	6e-63	88%	XP_006453393.1
4-coumarate:CoA ligase 1 [Populus trichocarpa x Populus deltoides]	214	214	62%	3e-62	76%	AAC39366.1
PREDICTED: probable 4-coumarate--CoA ligase 3 [Musa acuminata subsp. malaccensis]	213	213	55%	5e-62	86%	XP_009408050.1

Figure 4.10: Blastx search for 5' and 3' RACE PCR product sequence showed the highest similarity (90% max identity) with *4CL* from *Musa acuminata* AAA Group (accession no.AHF20170.1) with query coverage of 58% and E value of 3e-72

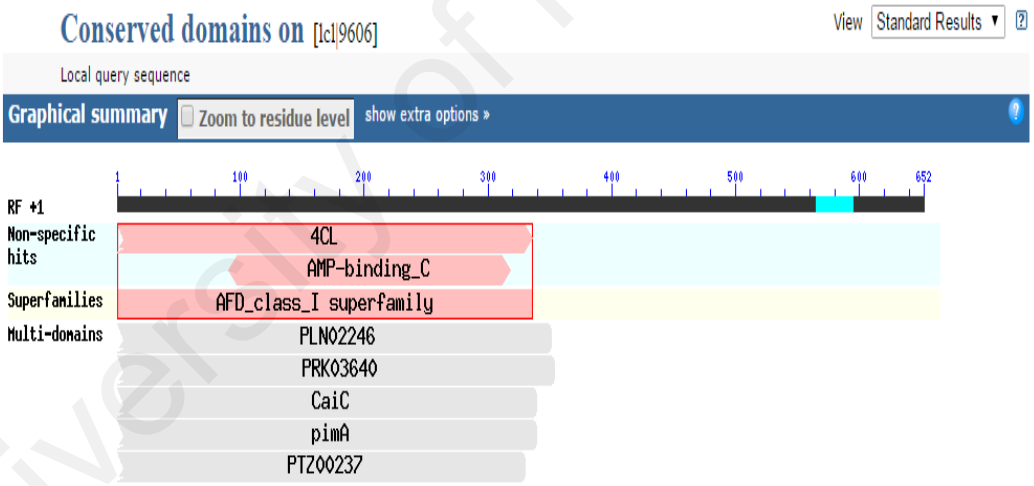


Figure 4.11: Graphical summary of BLAST results of deduced amino acids of 4CL on conserved domain

4.4.2 Degenerate primers-PCR method

Amplification by RACE method only recovered approximately 650bp of *B.rotunda 4CL* with missing of 5' region. Further amplification by Degenerate-PCR method was carried out. The primers were designed based on amino acids sequence from the multiple alignment of 4CL organism (**Appendix C7**). Outer and inner degenerate pair of primers (*DOF*, *DOR*) and (*DIF*, *DIR*) were used during primary and secondary PCR reaction respectively (**Table 3.4**). This strategy was used in order to make the amplification of the target region became more specific and to increase the concentration of the target region.

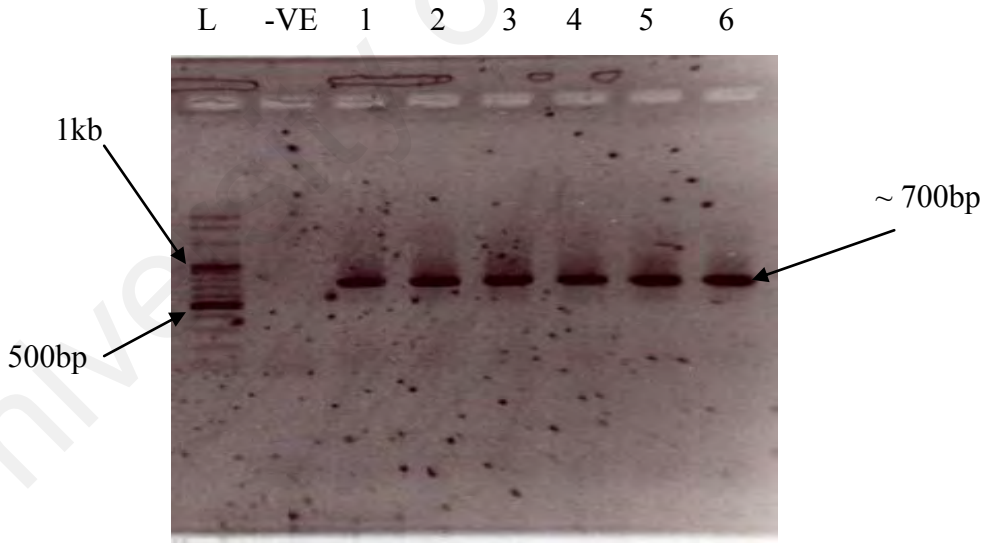


Figure 4.12: Amplification of *Br4CL* using Degenerate inner primer pairs (*DIF*, *DIR*). Amplified product is around 700bp

L: Marker: 1kb DNA ladder (EUR_X).

Lane 1-6: PCR product

-VE: Negative control; without cDNA template

A single band with size of around 700bp was successfully PCR amplified using Degenerate inner primers and the template was from PCR product using outer degenerate primers (**Figure 4.12**).

The sequences recovered from this method matched with predicted *4CL*- 1- like from *Musa acuminata subsp malaccencis* (accession number: XP 009414743.1) with 84 % identity and 91% coverage through Blastx analysis (**Figure 4.13**).

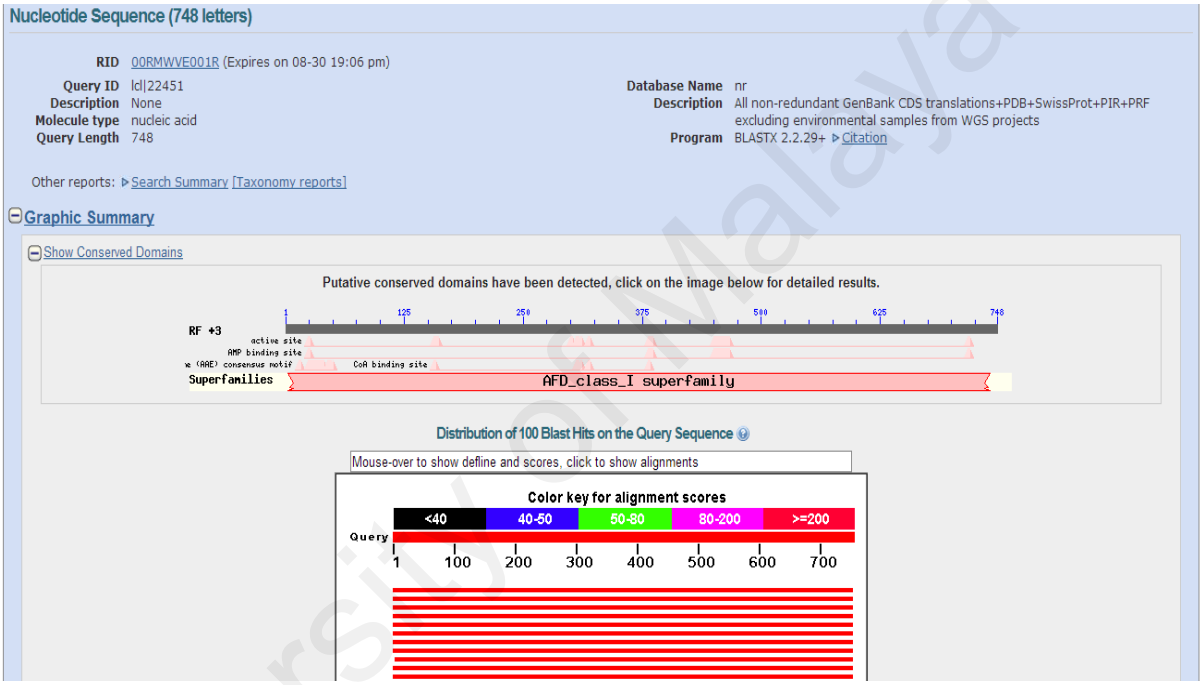


Figure 4.13: Blastx search result showed that the sequence matched with 4CL from other organisms

The sequences were aligned with the sequences recovered from RACE-PCR method using BioEdit Sequence Alignment Editor. The alignment results showed that 18 nucleotides overlapped at the end of the 5' region as shown in **Figure 4.14**.

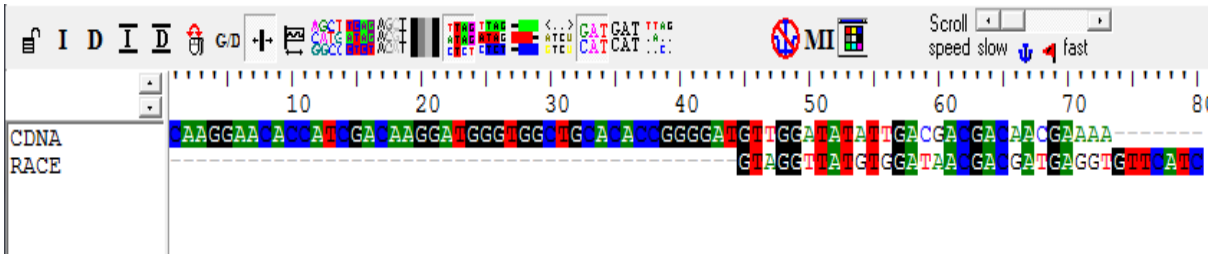


Figure 4.14: Eighteen overlapping nucleotide sequences detected between PCR product sequences from RACE-PCR method and Degenerate primers PCR method.

Subsequently, both sequences were separately aligned with *4CL* sequences from *Musa acuminata* (AAA) to reconfirm the alignment result (**Figure 4.15)**. Both results showed same position of overlapping nucleotides (1,186 nucleotides). As expected, the alignment result also revealed the position of stop codon at 3' region (**Figure 4.16**).

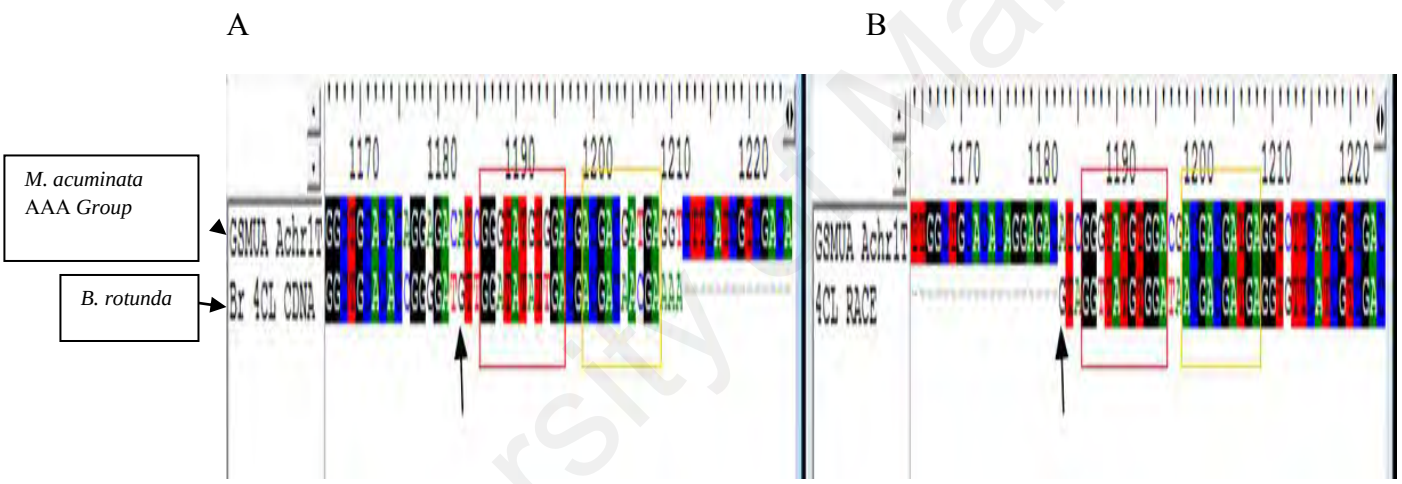


Figure 4.15: Reconfirmation by aligning nucleotides sequences with 4CL isoform from *Musa acuminata* AAA Group[Same overlapping nucleotides at 1186 nucleotides position for both PCR product sequencing from Degenerate Primers-PCR method (A) and RACE-PCR method (B)]

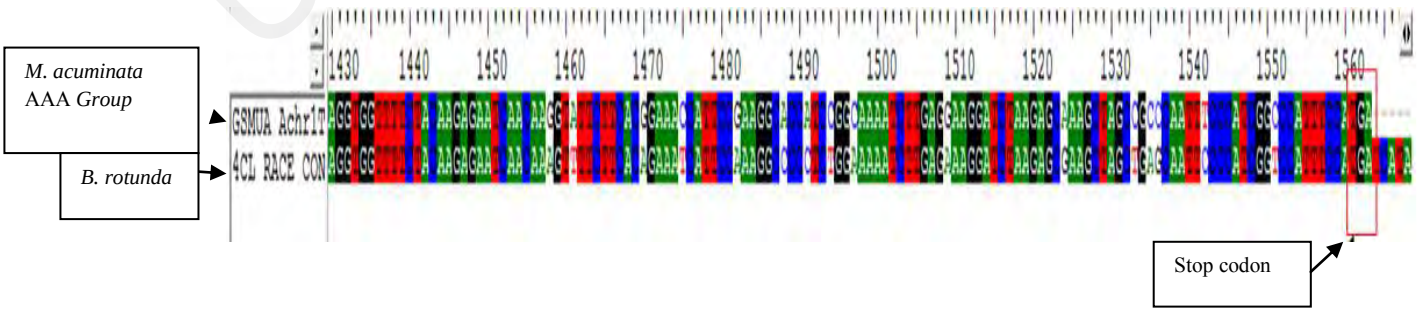


Figure 4.16: Stop codon was detected at the end of 3' region of align sequence from RACE-PCR method with 4CL from *Musa acuminata* AAA Group

4.4.3 Amplification of 5' region of *Br4CL*

In order to find the remaining 5' region, amplification was carried out with the *DOF* primer (**section 4.4.2)** and new primer *SR* designed from partial sequence recovered from **4.4.2** (**Appendix C6 and C8**).

A single band with size around 600bp was successfully amplified (**Figure 4.17**) and their sequencing result matched the 5' region of *4CL* in other organisms in NCBI database as shown in **Figure 4.18.**

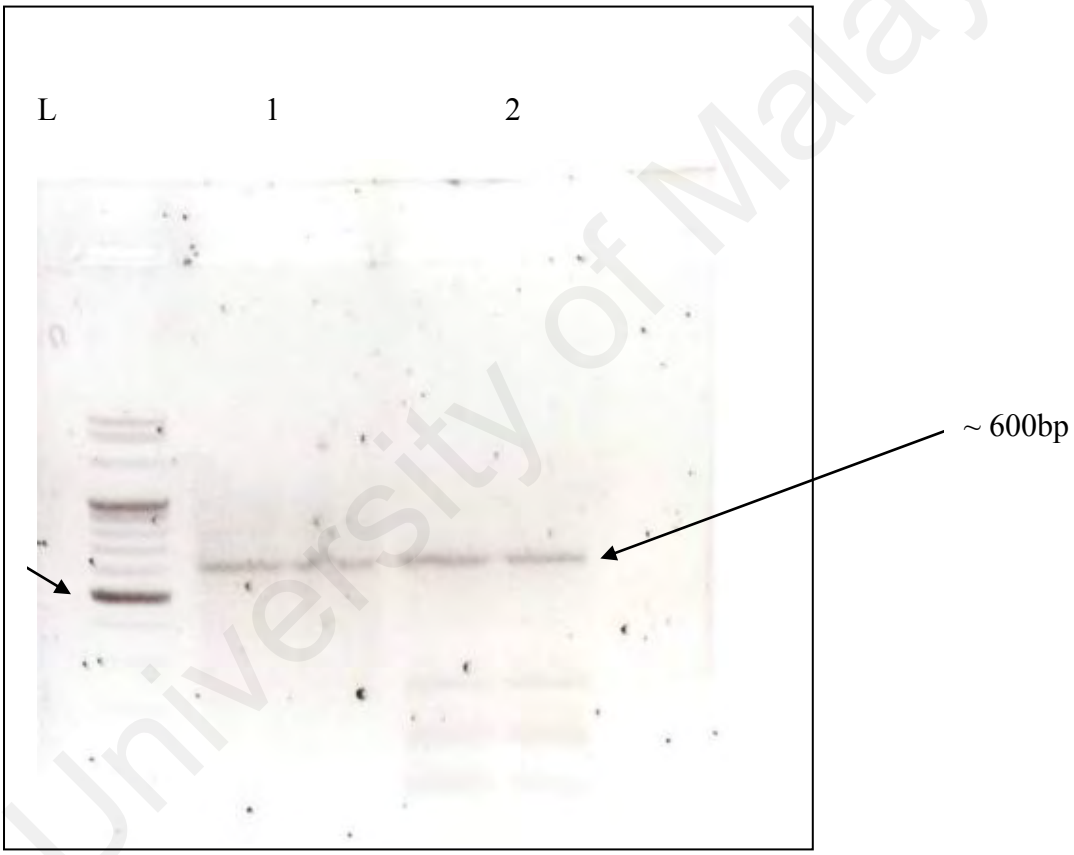
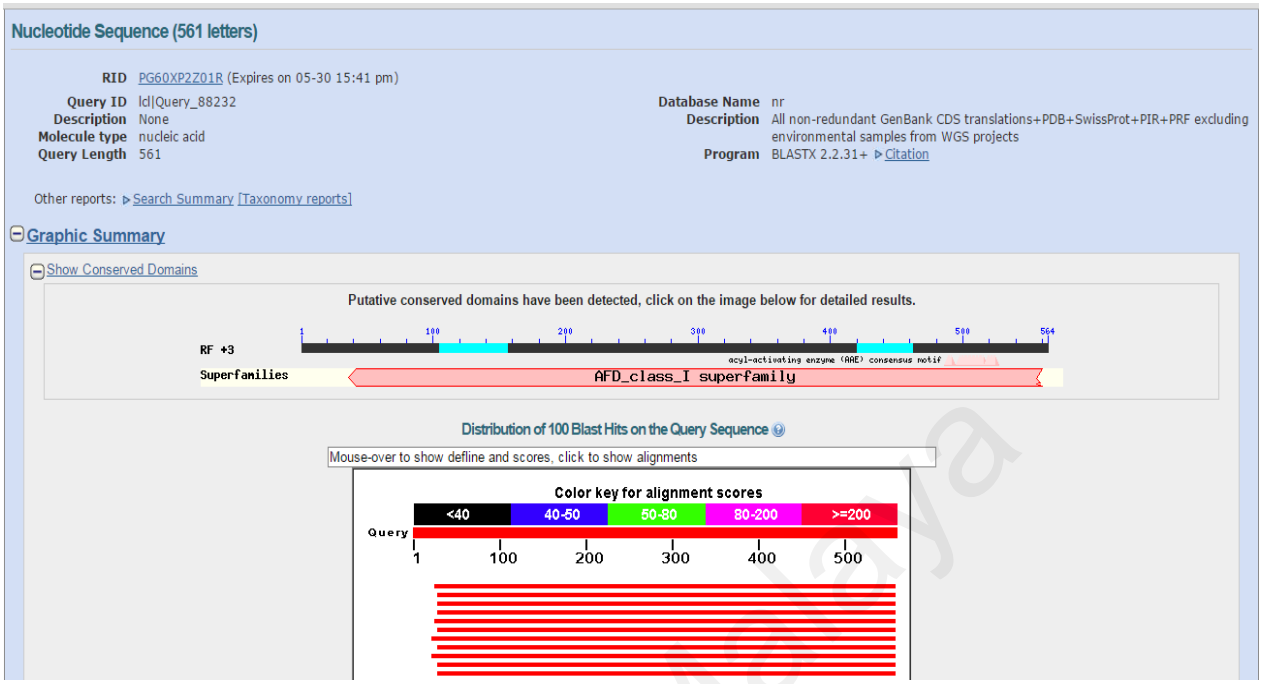
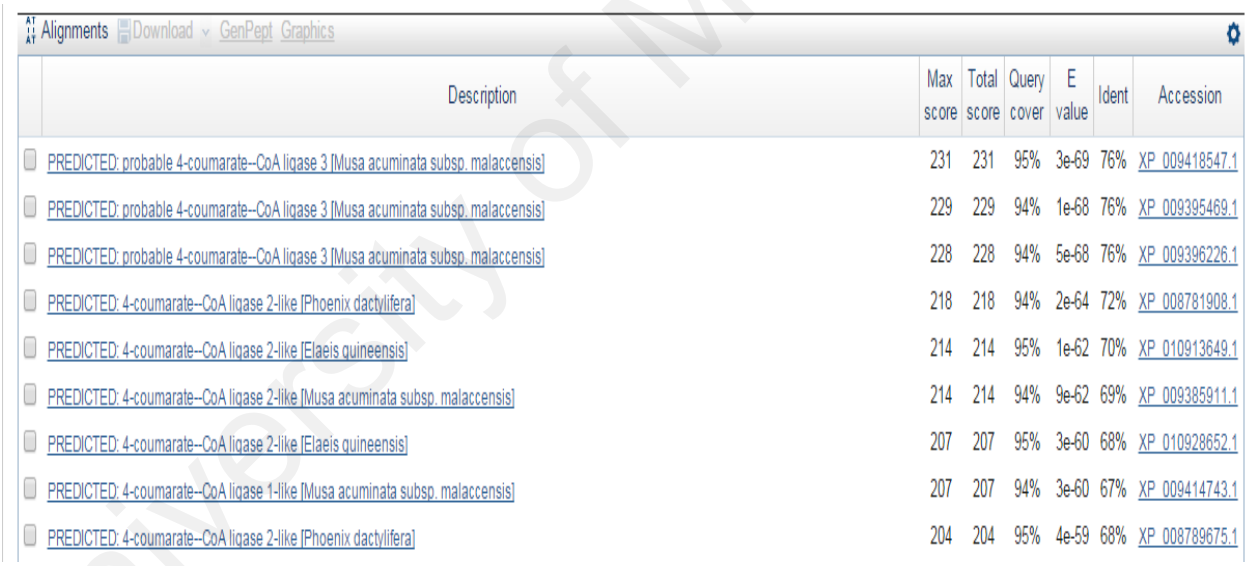


Figure 4.17: Amplification of 5' region of *Br4CL* using Forward (*DOF*) and Reverse primers (*SR*)

L: Marker: 1kb DNA ladder (EUR_X).

Lane 1-2: PCR product





Alignments Download GenPept Graphics

Description	Max score	Total score	Query cover	E value	Ident	Accession
PREDICTED: probable 4-coumarate--CoA ligase 3 [Musa acuminata subsp. malaccensis]	231	231	95%	3e-69	76%	XP_009418547.1
PREDICTED: probable 4-coumarate--CoA ligase 3 [Musa acuminata subsp. malaccensis]	229	229	94%	1e-68	76%	XP_009395469.1
PREDICTED: probable 4-coumarate--CoA ligase 3 [Musa acuminata subsp. malaccensis]	228	228	94%	5e-68	76%	XP_009396226.1
PREDICTED: 4-coumarate--CoA ligase 2-like [Phoenix dactylifera]	218	218	94%	2e-64	72%	XP_008781908.1
PREDICTED: 4-coumarate--CoA ligase 2-like [Elaeis guineensis]	214	214	95%	1e-62	70%	XP_010913649.1
PREDICTED: 4-coumarate--CoA ligase 2-like [Musa acuminata subsp. malaccensis]	214	214	94%	9e-62	69%	XP_009385911.1
PREDICTED: 4-coumarate--CoA ligase 2-like [Elaeis guineensis]	207	207	95%	3e-60	68%	XP_010928652.1
PREDICTED: 4-coumarate--CoA ligase 1-like [Musa acuminata subsp. malaccensis]	207	207	94%	3e-60	67%	XP_009414743.1
PREDICTED: 4-coumarate--CoA ligase 2-like [Phoenix dactylifera]	204	204	95%	4e-59	68%	XP_008789675.1

Figure 4.18: Blastx search result of PCR product matched with the predicted probable *4CL3* from *Musa acuminata subsp malaccacensis* with 95% coverage and 76% identity

An ATG start codon was successfully amplified and detected on the sequence as shown in **Figure 4.19** and **Appendix C10.**

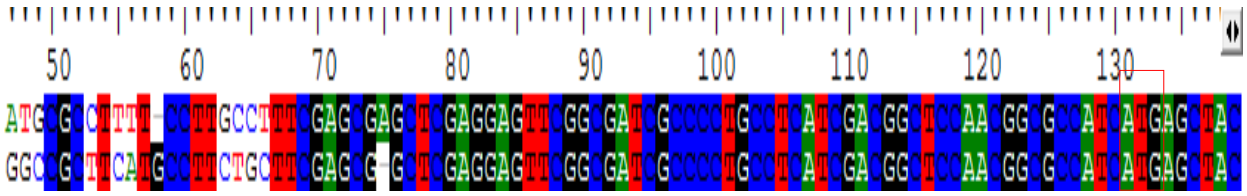


Figure 4.19: Overlapping nucleotides region between PCR product of forward and reverse sequences. ATG for the start codon was identified (Red box)

In addition, overlapping region was detected where 67 nucleotides were overlapped with partial sequences (**Section 4.3.2**) as shown in **Figure 4.20** and **Appendix C11.**

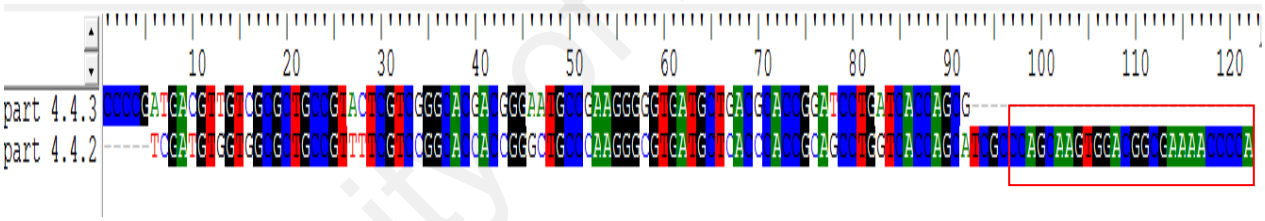


Figure 4.20: Sixty seven overlapping nucleotide sequences were detected between PCR product and partial sequence. Red box showed the location of specific reverse primer (*SR*)

Analysis of the upstream sequences at 5' region showed that the sequence (before the ATG) only matched 63% identity with 4CL3 from *Musa acuminata subsp malaccensis*. While downstream sequences (after the ATG) were 82% identical and covered 99% of 4CL3 from *Musa acuminata subsp malaccensis* (**Figures 4.21 and 4.22**). This further indicates the location of the ATG found in the sequence could be the start codon for the proteins.

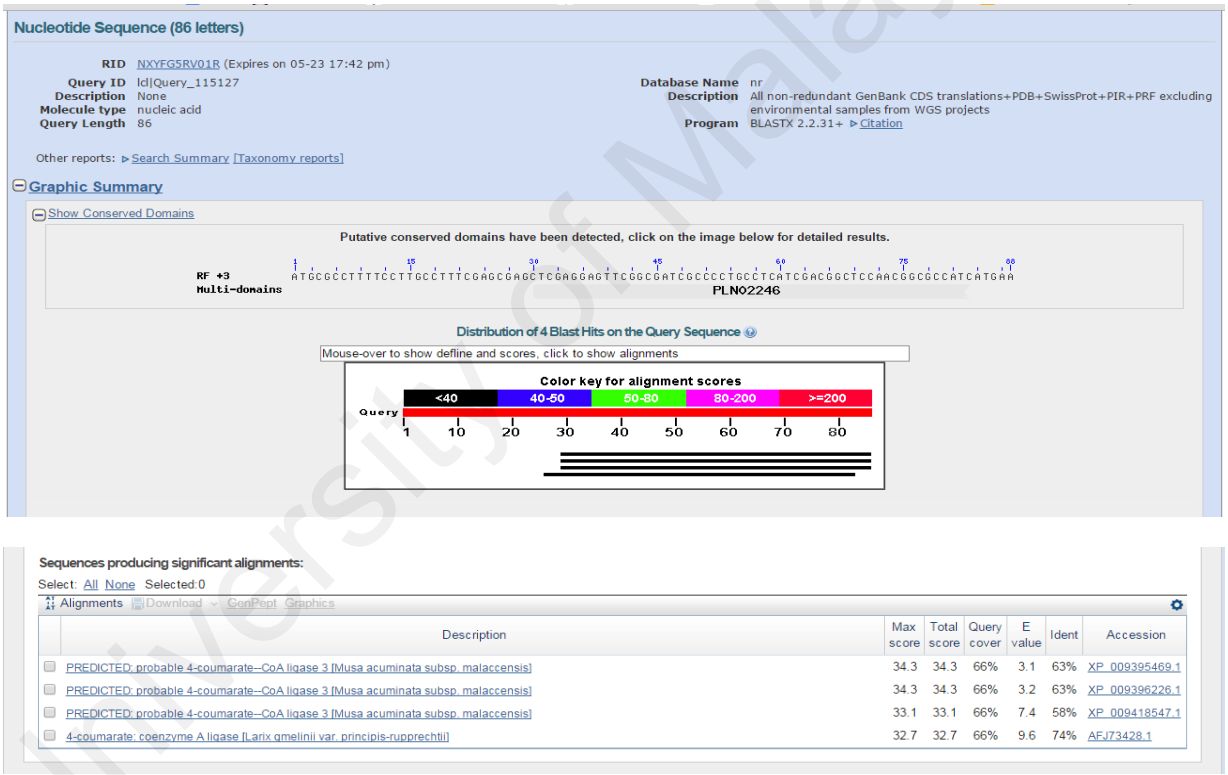
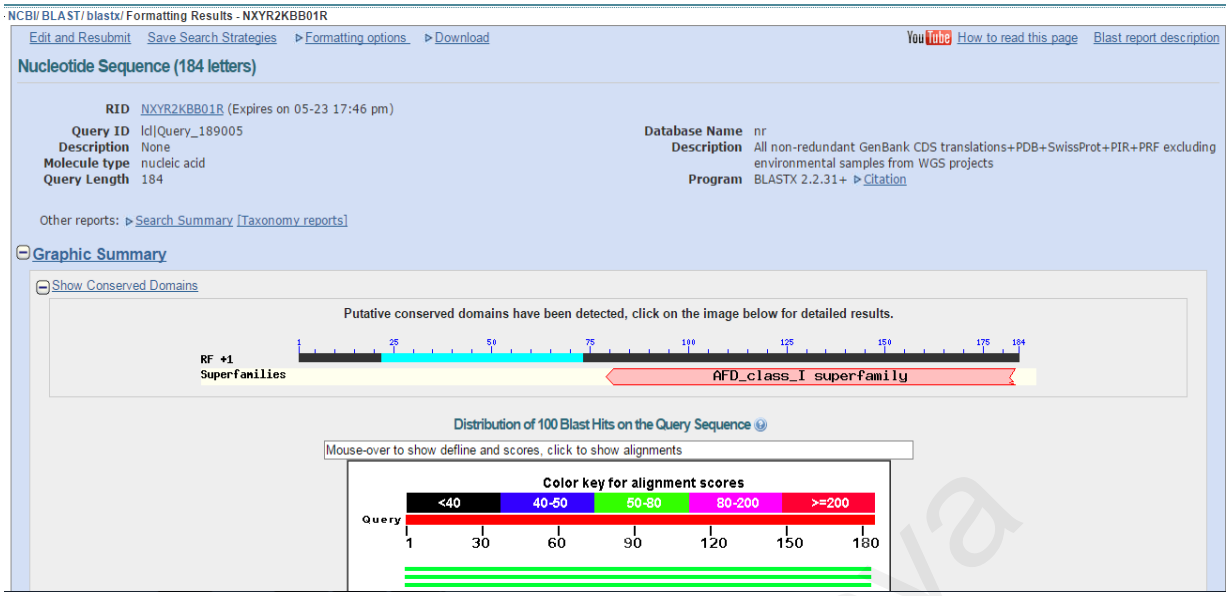
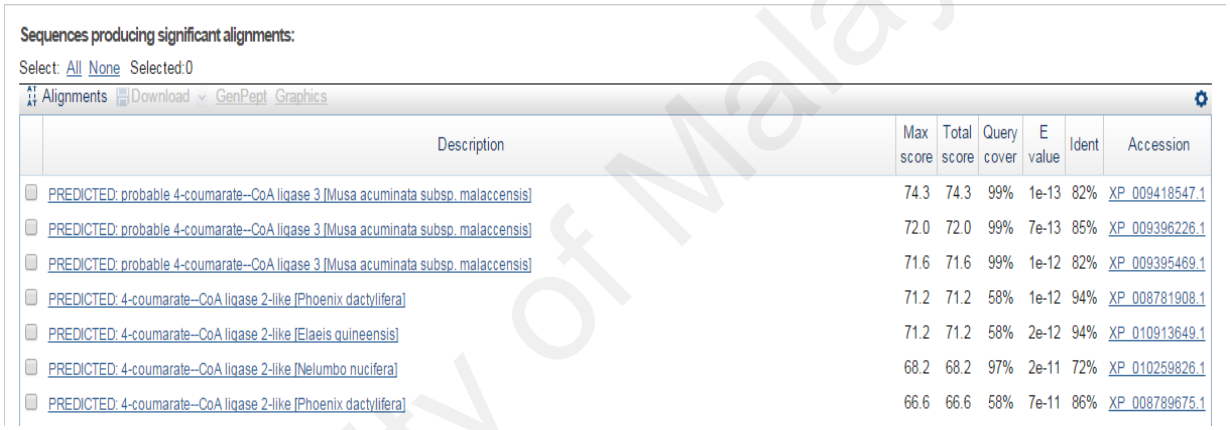


Figure 4.21: Blastx search for sequence before ATG match with the predicted probable *4CL3* from *Musa acuminata subsp malaccensis* but covered only 66% with 63% identity



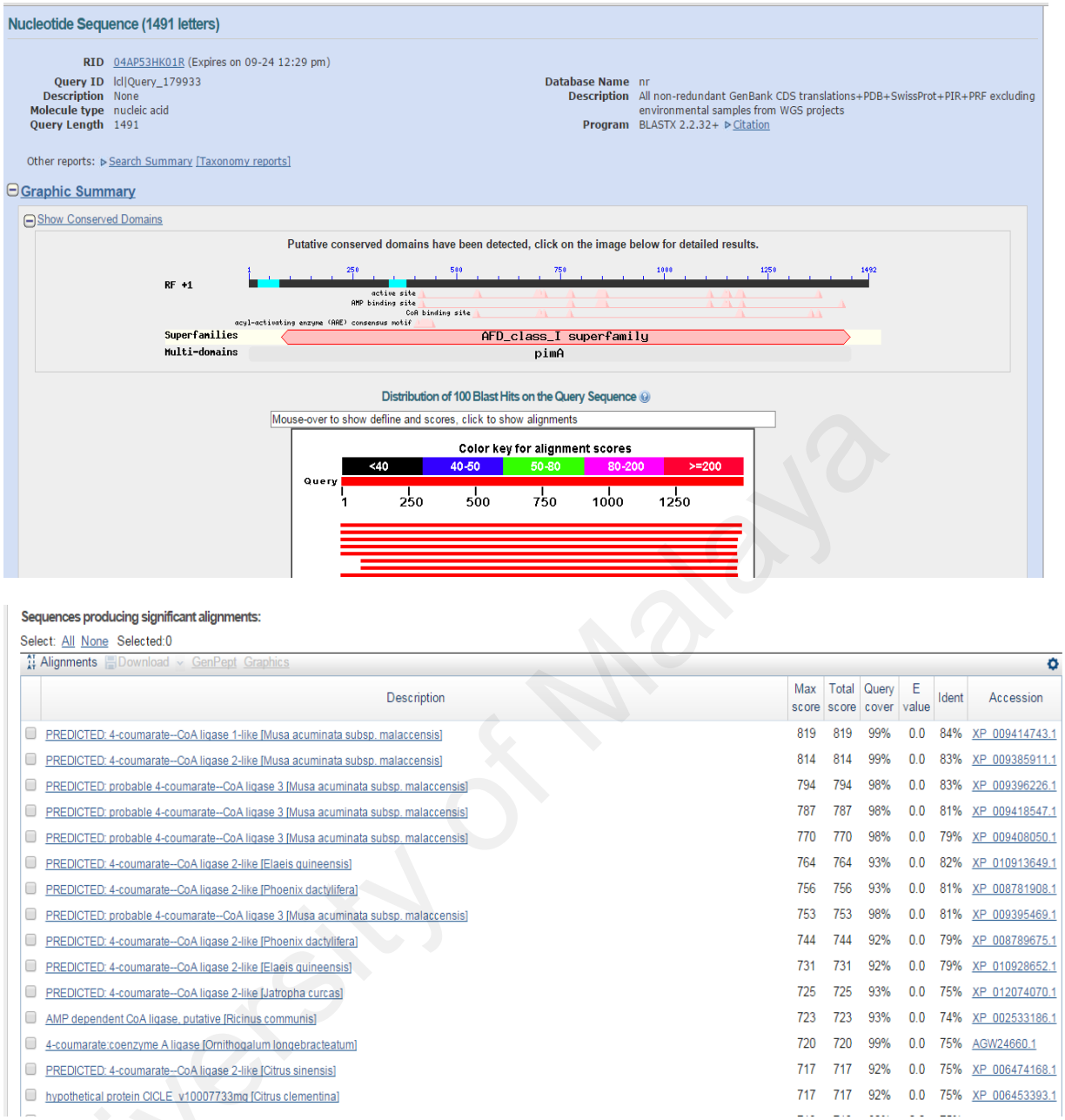


Description	Max score	Total score	Query cover	E value	Ident	Accession
PREDICTED: probable 4-coumarate--CoA ligase 3 [Musa acuminata subsp. malaccensis]	74.3	74.3	99%	1e-13	82%	XP_009418547.1
PREDICTED: probable 4-coumarate--CoA ligase 3 [Musa acuminata subsp. malaccensis]	72.0	72.0	99%	7e-13	85%	XP_009396226.1
PREDICTED: probable 4-coumarate--CoA ligase 3 [Musa acuminata subsp. malaccensis]	71.6	71.6	99%	1e-12	82%	XP_009395469.1
PREDICTED: 4-coumarate--CoA ligase 2-like [Phoenix dactylifera]	71.2	71.2	58%	1e-12	94%	XP_008781908.1
PREDICTED: 4-coumarate--CoA ligase 2-like [Elaeis guineensis]	71.2	71.2	58%	2e-12	94%	XP_010913649.1
PREDICTED: 4-coumarate--CoA ligase 2-like [Nelumbo nucifera]	68.2	68.2	97%	2e-11	72%	XP_010259826.1
PREDICTED: 4-coumarate--CoA ligase 2-like [Phoenix dactylifera]	66.6	66.6	58%	7e-11	86%	XP_008789675.1

Figure 4.22: Blastx search for sequence after ATG showed significance match with 99% of query coverage with 82 % identity with predicted probable *4CL3* from *Musa acuminata subsp malaccensis*

4.4.4 Analysis of open reading frame (ORF)

Finally, the full-length of *Br4CL* was successfully amplified with 1,976bp in length.The coding region was 1,578 base pair and code for 526 amino acids as determined by ExPASy (translate tool) program (**Appendix C12**). The final sequence showed highest similarity (84% maximum identity) with *4CL* 1-like from *Musa acuminata subsp malaccensis* (accession no.XP009414743.1) with query coverage of 99% and E value of 0 (**Figure 4.23**). In summary, RACE-PCR method covered the end region of 3' end of target gene; while Degenerate Primers-PCR method covered the centre region of target gene where else the third method (**4.3.3**) covered the 5' region of the target gene (**Appendix C11**). Three strategies were used to obtain the coding region of *4CL* gene in *B.rotunda.*



Description	Max score	Total score	Query cover	E value	Ident	Accession
PREDICTED: 4-coumarate--CoA ligase 1-like [Musa acuminata subsp. malaccensis]	819	819	99%	0.0	84%	XP_009414743.1
PREDICTED: 4-coumarate--CoA ligase 2-like [Musa acuminata subsp. malaccensis]	814	814	99%	0.0	83%	XP_009385911.1
PREDICTED: probable 4-coumarate--CoA ligase 3 [Musa acuminata subsp. malaccensis]	794	794	98%	0.0	83%	XP_009396226.1
PREDICTED: probable 4-coumarate--CoA ligase 3 [Musa acuminata subsp. malaccensis]	787	787	98%	0.0	81%	XP_009418547.1
PREDICTED: probable 4-coumarate--CoA ligase 3 [Musa acuminata subsp. malaccensis]	770	770	98%	0.0	79%	XP_009408050.1
PREDICTED: 4-coumarate--CoA ligase 2-like [Elaeis guineensis]	764	764	93%	0.0	82%	XP_010913649.1
PREDICTED: 4-coumarate--CoA ligase 2-like [Phoenix dactylifera]	756	756	93%	0.0	81%	XP_008781908.1
PREDICTED: probable 4-coumarate--CoA ligase 3 [Musa acuminata subsp. malaccensis]	753	753	98%	0.0	81%	XP_009395469.1
PREDICTED: 4-coumarate--CoA ligase 2-like [Phoenix dactylifera]	744	744	92%	0.0	79%	XP_008789675.1
PREDICTED: 4-coumarate--CoA ligase 2-like [Elaeis guineensis]	731	731	92%	0.0	79%	XP_010928652.1
PREDICTED: 4-coumarate--CoA ligase 2-like [Jatropha curcas]	725	725	93%	0.0	75%	XP_012074070.1
AMP dependent CoA ligase, putative [Ricinus communis]	723	723	93%	0.0	74%	XP_002533186.1
4-coumarate:coenzyme A ligase [Ornithogalum longebracteatum]	720	720	99%	0.0	75%	AGW24660.1
PREDICTED: 4-coumarate--CoA ligase 2-like [Citrus sinensis]	717	717	92%	0.0	75%	XP_006474168.1
hypothetical protein CICLE_v10007733mg [Citrus clementina]	717	717	92%	0.0	75%	XP_006453393.1

Figure 4.23 : Blastx search result of full length showed highest similarity (84% maximum identity) with *4CL* 1-like from *Musa acuminata subsp malaccensis* (accession no. XP009414743.1) with query coverage of 99% and E value of 0

The final sequence of 1,976bp was then subjected to protein blast, BLASTP (Basic Local Alignment Search Tool). The results showed that several conserved domains were belonging to 4CL superfamily such as AMP binding domain, AMP binding enzyme C-terminal and adenylate forming enzymes superfamily (**Figure 4.24**).

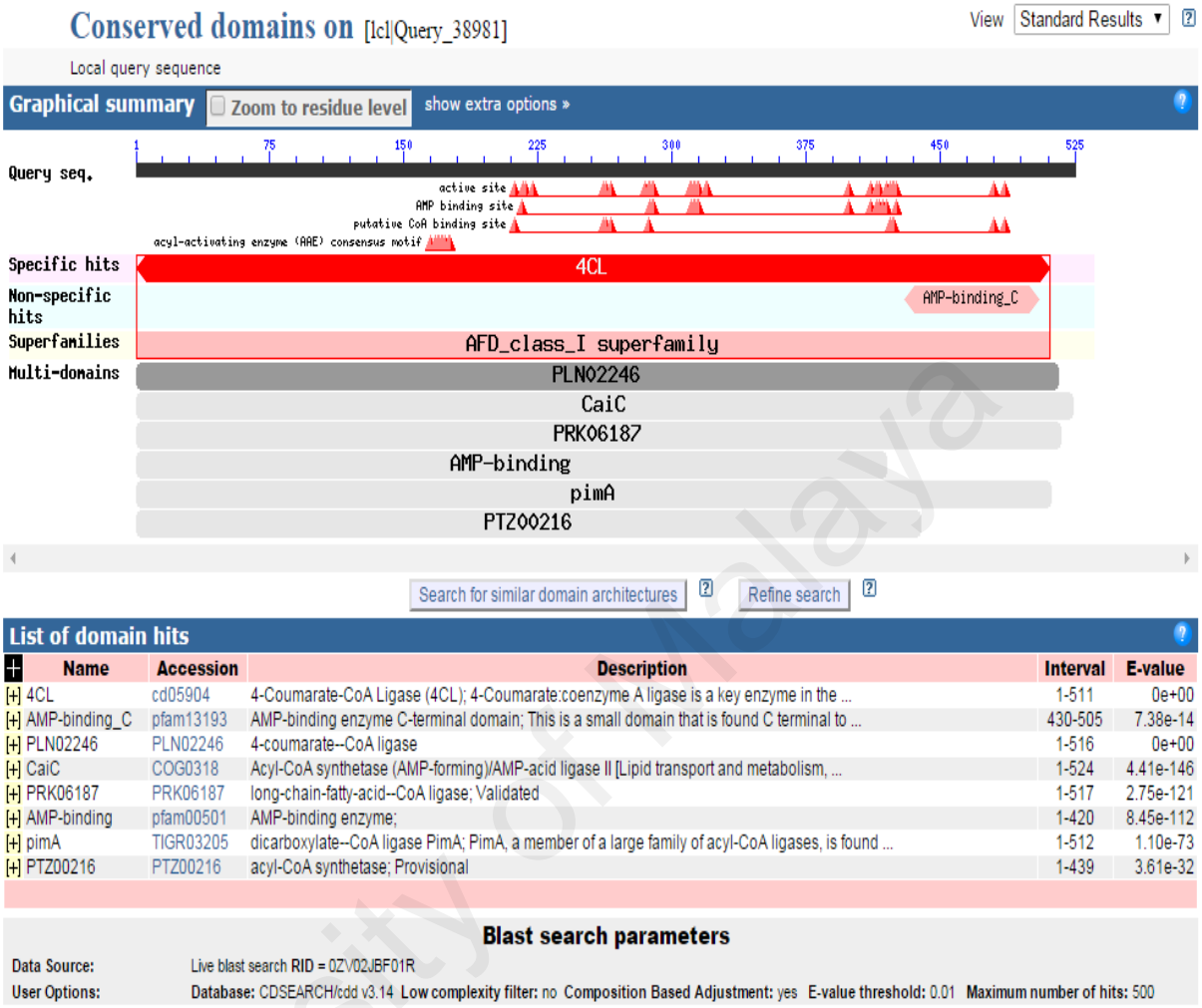


Figure 4.24: Graphical summary of BLASTP results of deduced amino acids of 4CL on conserved domain

4.5 Phylogenetic analysis of *Br4CL*

Construction of phylogenetic trees was done using MEGA 6.0 programme. Sequences of 4CL enzymes from other resources were extracted from NCBI. A total of 30 sequences of 4CL enzymes were used, these include 4CL from *B. rotunda*, Arabidopsis, aspen and *Musa acuminata subs malaccensis*, for phylogenetic trees comparison. The accession numbers of these sequences are provided in **Table 4.4.** Neighbour joining method was used for the construction of the trees after the alignment of amino acid sequence of the proteins using CLUSTALW alignment in BioEdit programme. It was found that 4CL protein from *B. rotunda* was phylogenetically related to 4CLs from *M. acuminata subsp malaccensis* (Red box) as shown in **Figure 4.25.** Green boxes showed characterized 4CL enzymes that devoted to the monolignol biosynthesis pathway while purple boxes showed characterized 4CL enzymes mainly participate in the flavonoid biosynthesis pathway based on description by Xu and co-workers (2011).

Table 4.4: *4CL* gene database from NCBI

Species	Gene names	GenBank Accession umbers
Oryza sativa	*Os4CL1*	XP_015650724.1
	Os4CL2	XP_015624111.1
	Os4CL3	XP_015625716.1
Populus tremuloides	*Popt4CL1*	AF041049
	Popt4CL2	AF041050
Phyllostachys edulis	*Pe4CL*	FP101648.1
Eucalyptus camaldulensis	*Ec4CL*	ACX68559.1
Zea mays	*Zm4CL*	NM_001111788
Populus tomentosa	*Pto4CL1*	AAL02145.1
Arabidopsis thaliana	*At4CL1*	U18675
	At4CL2	AF106086
	At4CL3	AF106088

Continued table 4.4

Species	Gene names	GenBank Accession numbers
Giycine max	*Gm4CL1*	AF279267
	Gm4CL2	AF002259
	Gm4CL3	AF002258
Lolium perenne	*Lp4CL1*	AAF37732.1
	Lp4CL2	AAF37733.1
Selaginella moellendorffii	*Sm4CL*	XP_002985214.1
Sorghum bicolour	*Sb4CL2*	XP_002451647.1
Nicotiana tabacum	*Nt4CL1*	O24145.1
	Nt4CL2	O24146.1
Musa acuminata subsp.	*Ma4CL1*	XP_009414743.1
Malaccensis	*Ma4CL2*	XP_009385911.1
	Ma4CL3	XP_009396226.1
	Ma4CL3	XP_009418547.1
	Ma4CL3	XP_009408050.1
	Ma4CL3	XP_009395469.1
Pinus taeda	*Pt4CL1*	PTU12012
	Pt4CL2	PTU12013

PHYLOGENETIC TREES IMAGE

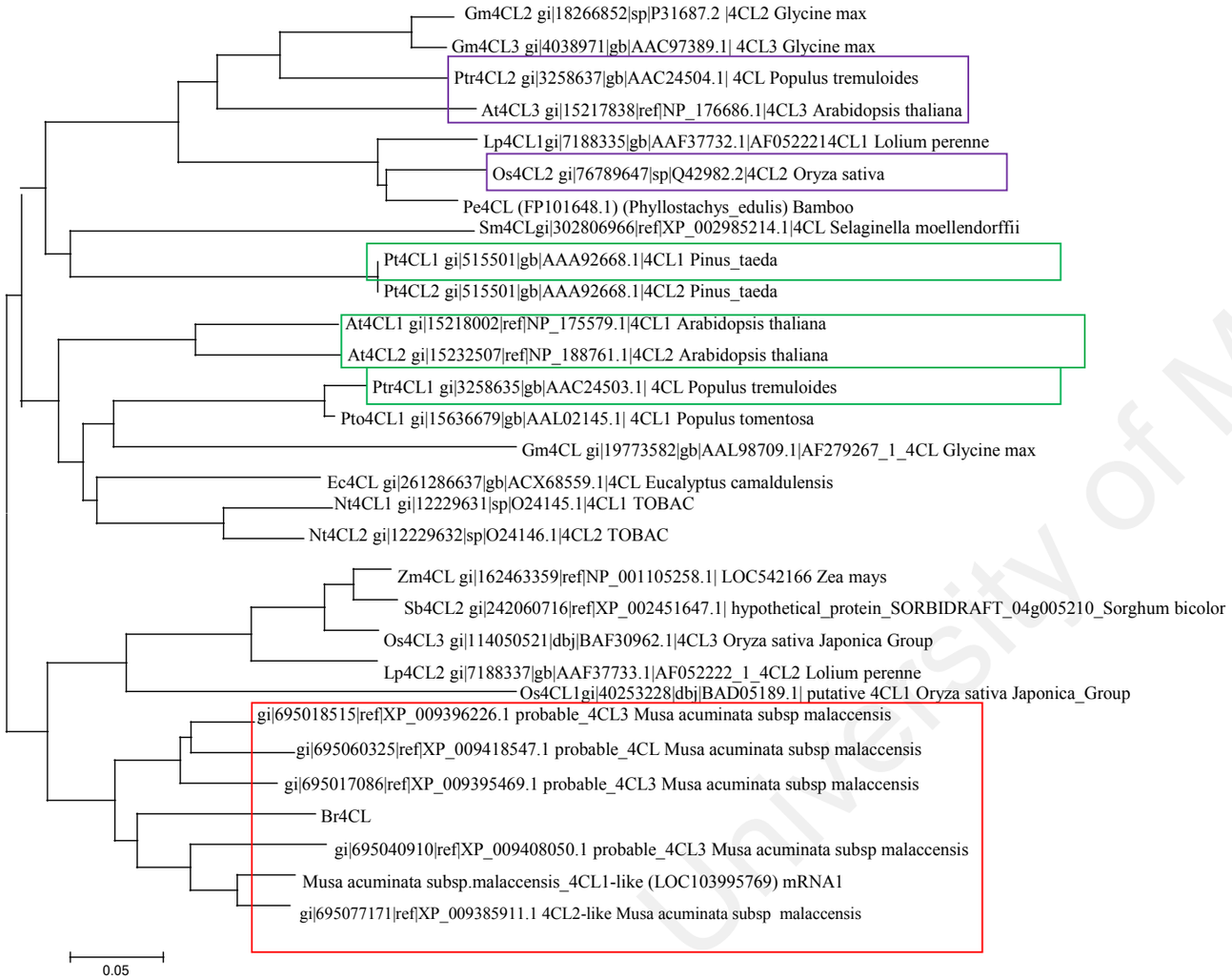


Figure 4.25: Phylogenetic tree of *Br4CL* and its homologs in some model plant. Scale indicates amino acid substitutions per position. 4CL enzyme from *B.rotunda* has close evolutionary relationship with 4CL enzymes from *Musa acuminata subsp malaccencis* (Red box).Neighbor-joining (NJ) tree was built using Mega 5.2 software (Tamura et al., 2007)

4.6 Protein expression of *Br4CL* in *E.coli*

In protein expression study, the coding sequence of *Br4CL* was successfully cloned into pET28a vector plasmid and transformed into *E.coli* BL21 (DE3). The SDS-PAGE and Western blot results showed that the sequence was successfully expressed in *E. coli* (**Figure. 4.26**). The size of the band was around 70 kDa as expected (1578bp codes for 526 amino acids)

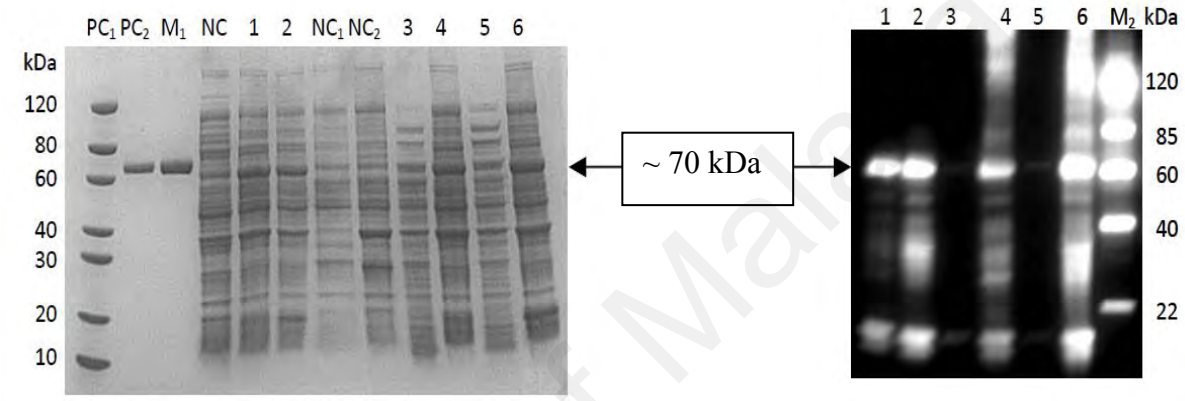


Figure 4.26: SDS-PAGE (left) and Western blot (right, using anti-His antibody, GenScript Cat.No. A00186) analysis of sequence expression in construct pET28a vector

Lane M_1 : Protein marker

Lane PC_1: BSA (1 μg)

Lane PC_2: BSA (2 μg)

Lane NC: Cell lysate without induction

Lane NC_1: Supernatant of cell lysate non-induced

Lane NC_2: Pellet of cell lysate non-induced

Lane 1 : Cell lysate with induction for 16 h at 15 °C

Lane 2 : Cell lysate with induction for 4 h at 37 °C

Lane 3 : Supernatant of cell lysate with induction for 16 h at 15 °C

Lane 4 : Pellet of cell lysate with induction for 16 h at 15 °C

Lane 5 : Supernatant of cell lysate with induction for 4 h at 37 °C

Lane 6 : Pellet of cell lysate with induction for 4 h at 37 °C

CHAPTER 5: DISCUSSION

5.1 Introduction

In this study, the expression level of *4CL* gene in cell suspension cultures of *B. rotunda* in response to phenylalanine treatment was observed. The transcription level of *4CL* has been found to be increased over 14 days after being treated with phenylalanine. In line with our previous studies (Tan et al., 2012; Md-Mustafa et al., 2014), these findings indicated the promotive effect of phenylalanine in enhancing the selected flavonoid production.

The coding region of *4CL* gene from *B.rotunda* was successfully amplified by performing several PCR reactions and Blast search analysis. Our first attempt with RACE analysis only results in partial coding sequence of 4CL. We then performed PCR with degenerate primers to complete the coding sequence. Degenerate primers were designed based on alignment of sequences of 4CL proteins from other monocots and dicots. The amplified sequence of *4CL* gene was then searched against the public database of NCBI using Blastx search analysis and analysed using ExPASy translate program. This *4CL* gene sequence obtained in this study (i.e *Br4CL* gene) was 82 % similar to *4CL*-1 like *Musa acuminata subsp malaccensis* with 99% coverage where it codes for 526 amino acids. In addition, a phylogenetic tree suggested that *Br4CL* belonged and is grouped together with 4CL isoform from *Musa acuminata subsp malaccensis*. In various plant species, the main function of class II 4CL has been suggested to be associated with flavonoid synthesis. While the biosynthesis of monolignols has been considered to be the main function of class I 4CL.

Expression of *4CL* gene from *B.rotunda* in *E.coli* was performed to obtain the protein expression in prokaryotes cells. Successful 4CL protein expression in microbial cells act as preliminary proof for further functionality study or for co-expression study with other biosynthetic enzymes to produce compounds in microbes, as an alternative either to

extracting compounds from plants or obtaining it via chemical synthesis. The details of each analysis that were carried out in the present study were discussed as follows.

5.2 Relative expression of *4CL* gene in response to phenylalanine treatment

As mentioned earlier in 1.1, phenylalanine is one of the key players in the production of the large and diverse family of phenylpropanoid specialized metabolites derived from shikimic acid pathway (Herrmann et al., 1999). Besides that, its structural backbone (6-carbon ring and 3-carbon side) chain serve as a basic structure for secondary metabolites such as flavonoids. Phenylalanine is a natural amino acids that can be found in plants and having this aromatic amino acid as a precursor simply means increasing the substrate for the PAL enzyme, which is the initial enzyme in general phenylpropanoid metabolism and may affect continuous downstream gene expression such as *C4H* and *4CL*. In order to test this hypothesis, RT- qPCR experiment was performed to analyse the effect of phenylalanine, used as a precursor, towards *4CL* transcript of *B.rotunda* cell suspension cultures.

Studies have shown that bioactive compounds can be induced in cell suspension cultures or organ cultures by supplying precursors such as methyl jasmonate, tyrosine, cinnamic acids, salicylic acids and phenylalanine (Sivakumar et al., 2004; Kováčiket et al., 2007; Bauer et al., 2009; Gutiérrez-Carbajal et al., 2010; Tan et al., 2012). For instance, the addition of phenylalanine was reported to stimulate taxol production in *Taxus cupidata* cell suspension culture (Fett-Neto et al., 1994). In addition, repetitive feeding of strawberry suspension culture with L-Phe led to 31 and 80% higher anthocyanin accumulation as compared to single Phe-feeding and non-feeding culture, respectively (Edahiro et al., 2005). In *Gloriosa superba,* after application of 30 mM phenylalanine and tyrosine, not only colchicine production was increased, the activities of PAL and tyrosine ammonia lyase (TAL) detected was high (Sivakumar et al., 2004). On the other hand, a recent study has

shown increased endogenous concentrations of phenylalanine and tyrosine in transgenic lines of grape cell cultures transformed with 3-deoxy-D-arabino-heptulosonate7-phosphate synthase enzyme (AroG∗) which mimicked a stress response in the cells leading to a significant increase in several specialized metabolites representing different phenylpropanoid pathways in *Vitis vinifera* cv. Gamay Red (Manela et al., 2015). The study also suggest that increased phenylpropanoid accumulation in the transformed lines of *Vitis vinifera* cv. Gamay Red cell suspension was probably because of increased production of the common substrates for these pathways, namely phenyalanine and tyrosine. This study further support the assumption that indeed phenylalanine is a limiting factor for metabolites production.

In order to analyse the expression of *4CL* transcript, relative quantification analysis using Real-Time PCR was used. The relative quantification is a method for studying and analyzing the transcript expression of certain target genes by analyzing the C_t value of each PCR reaction. The expression of target gene that has been up regulated or down regulated in a particular situation or treatment can be analysed at real time (Live). Real-time PCR amplifies a specific target sequence in a sample and subsequently monitors the amplification progress using fluorescent technology.

Our result showed that *4CL* transcript from *B.rotunda* suspension cultures was generally upregulated by external phenylalanine treatment. The level of *4CL* transcript was increasing between 0 to 14 days of culture and highest at day fourteen. In another study, *Pueraria lobata* suspension cultures treated with methyl jasmonate (MeJA) also showed an up regulation at the highest transcript level of *4CL* in comparison to PAL, C4H ,CHS, CHI and IFS (Li et al., 2014). This shows that the regulation of *4CL* transcripts could be influenced by various mechanisms caused by precursors and stress inducers.

The relationships between the expression of genes and compounds production have been elucidated in many studies (Van Der Krol et al., 1988; Li et al., 2014; Manela et al.,

2015). In this study, although the transcript level of *4CL* seemed to be significantly affected by phenylalanine, it may not directly affect production of all compounds in the phenylpropanoid pathway but probably certain metabolites only. Thus, in the previous studies (Tan et al., 2012; Md-Mustafa et al., 2014) the correlation between the levels of 4CL enzymes and the production of flavonoids showed that higher expression of *4CL* may play a role in increasing the production of certain flavonoids such as pinostrobin, pinocembrin, alpinetin, 4-hydroxypanduratin A, and panduratin A. The thiol esters that have been produced by the 4CL enzymes may explain the connection between *4CL* expression level and compounds production as thioesters serve as an intermediate compound for many other subsequent specialized downstream compounds such as stillbenes, lignin, coumarins and also flavonoids.

The *4CL* gene family in *B.rotunda* is of particular interest since rhizomes of matured *B.rotunda* accumulate a novel phenylpropanoid derived compounds, such as pinostrobin, pinocembrin and panduratin A which is a primary/trace amount flavonoids found in *B.rotunda*'s plant (Yusuf et al., 2013; Tan et al., 2015). Biosynthesis of these flavonoids such as panduratin A has been shown to proceed via multiple-step pathway branching from the general phenylpropanoid pathway (Md-Mustafa et al., 2014). The first enzyme in this branch, chalcone synthase condensing 4-coumaroyl-CoA and 3 acetate moieties from malonyl-CoA as substrates to produce naringenin chalcone and dependent on activated 4-coumaroyl-CoA as a precursor. The existence of different 4-coumarate-derived end-products such as pinocembrin and panduratin A in rhizome of *B.rotunda* plant and also in their cell suspension cultures raises the possibility that distinct *4CL* genes might be associated with the biosynthesis of individual metabolic end-products in rhizome of *B.rotunda*. To address this question, it was first necessary to characterize the *4CL* gene family in this species.

5.3 Sequence analysis of 4CL enzyme in *B. rotunda*

The major differences in the *Br4CL* sequences when compared with others 4CLs was the discovery of the SELLEKKKTEEEEE amino acid sequence that may cause 82% identity with 4CL from *Musa acuminata subsp malaccensis* (**Figure 5.1**). Glutamic acid (E) amino acid was the most repeated amino acid followed by lysine (K) and then leucine (L). In the amino acid sequence, S denotes serine, E denotes glutamic acid, L denotes leucine, K denotes lysine and T denotes threonine. These amino acid sequences were found at the beginning of the *Br4CL* coding region near the start codon. Since the sequence alignment between *B.rotunda* and *Musa acuminata subsp malaccensis* was slightly different, the nucleotide sequence of the repeated amino acids was checked and confirmed free from contamination as shown by the single peak in the chromatogram (**Figure 5.2**). Therefore, this repeated sequence is verified to be the coding region of *Br4CL* gene. The presence of this repeated amino acid sequences had caused difficulty in the isolation of the coding region using a single set of primer. To overcome this problem, different sets of primers were used to amplify the coding regions by part. Following that, the sequences of the coding regions were realigned and similar sequences were overlapped to obtain the final coding region of *Br4CL*.

PREDICTED: 4-coumarate--CoA ligase 1-like [Musa acuminata subsp. malaccensis]
Sequence ID: ref|XP_009414743.1| Length: 548 Number of Matches: 1

Range 1: 53 to 548 GenPept Graphics Next Match Previous Match

Score	Expect	Method	Identities	Positives	Gaps	Frame
819 bits(2115)	0.0	Compositional matrix adjust.	415/497(84%)	447/497(89%)	2/497(0%)	+1

```
Query  1     MSYTEVDrasrraaaglrrfgvgrgQVIMILLRNSPEFVLAFLGASRRGAVATTANPFYT  180
            ++Y EVD A+R  AAGL   G+GRG V MILLRNSPEFV+AFL AS RGAVATTANPFYT
Sbjct  53    LTYAEVDAAARGFAAGLFSVGIGRGDVFMILLRNSPEFVIAFLAASYRGAVATTANPFYT  112

Query  181   PAEIHKQAVASGARLIVTESCYVDKVRDFARERGVVVVCVDD-PLPEGCRRFSelleeee  357
              EIHKQA  SGARLI+TESCYVDK+R+FA ER + +V V D P P+GCR F++L+  +
Sbjct  113   TGEIHKQAAGSGARLIITESCYVDKIREFAGERDITIVTVGDGPAPDGCRLFADLMGTDA  172

Query  358   kteeeeeeIDPDDVVALPYSSGTTGMPKGVMLTHRILITSVAQQVDGENPNLNFRPDDVL  537
                   E  DPDDVVALPYSSGTTG+PKGVMLTHR LITSVAQQVDG+NPNL    DDVL
Sbjct  173   GALSAAE-FDPDDVVALPYSSGTTGLPKGVMLTHRSLITSVAQQVDGDNPNLYLHQDDVL  231

Query  538   LCVLPLFHIYSLNSVLLCSLRVGAAILILRRFEVGPVLELVQRFRITIAPLVPPIVLEFV  717
              LCVLPLFHIYSLNSVLLC LRVGAAILI+RRFEVGP+LELVQR R+TIAP VPPIVLEFV
Sbjct  232   LCVLPLFHIYSLNSVLLCGLRVGAAILIMRRFEVGPLLELVQRHRVTIAPFVPPIVLEFV  291

Query  718   KSPLVDSFDLSSIRMVMSGAAPMGKEVEEKFMAKLPNARLGQGYGMTEAGPVLSMCLAFA  897
              KSPLVD +DLSSIRMVMSGAAPMGKE+E+KFM KLPNA+LGQGYGMTEAGPVLSMCLAFA
Sbjct  292   KSPLVDGYDLSSIRMVMSGAAPMGKELEDKFMTKLPNAQLGQGYGMTEAGPVLSMCLAFA  351

Query  898   KHPFEVKSGACGTVVRNAEMKIVDPETGAALGRNQRGEICIRGAQIMKGYINDPEATRNT  1077
              K PFEVKSGACGTVVRNAEMKIVDP TGA+LG NQRGEICIRGAQIMKGYINDPEATRNT
Sbjct  352   KEPFEVKSGACGTVVRNAEMKIVDPATGASLGSNQRGEICIRGAQIMKGYINDPEATRNT  411

Query  1078  IDKDGWLHTGDVGYVDNDDEVFIVDRLKEIIKFKGFQVAPAELEALLVMHPNISDAAVVP  1257
              IDKDGWLHTGD+GYVD+DDEVFIVDRLKEIIK+KGFQVAPAELEALL+ HPNI+DAAVVP
Sbjct  412   IDKDGWLHTGDIGYVDDDDEVFIVDRLKEIIKYKGFQVAPAELEALLITHPNIADAAVVP  471

Query  1258  MKDEAAGEVPVAFVVRSNGSKITEDEIKQYISRQVVFYKRINKVFFTEIIPKAPSGKILR  1437
              MKDEAAGEVPVAFVVRSNGSKI+EDEIKQYIS+QVVFYKRINKVFFTE IPKAPSGKILR
Sbjct  472   MKDEAAGEVPVAFVVRSNGSKISEDEIKQYISKQVVFYKRINKVFFTETIPKAPSGKILR  531

Query  1438  KDLRAKLAEQFPIGPFP  1488
              KDLRAKLA QFPIGPFP
Sbjct  532   KDLRAKLAAQFPIGPFP  548
```

Figure 5.1: Amino acids alignment between coding region of *4CL* gene from *B.rotunda* and *4CL*-1-like from *Musa acuminata subsp malaccensis*. Red boxes showed repeated amino acid sequences of *Br4CL.*

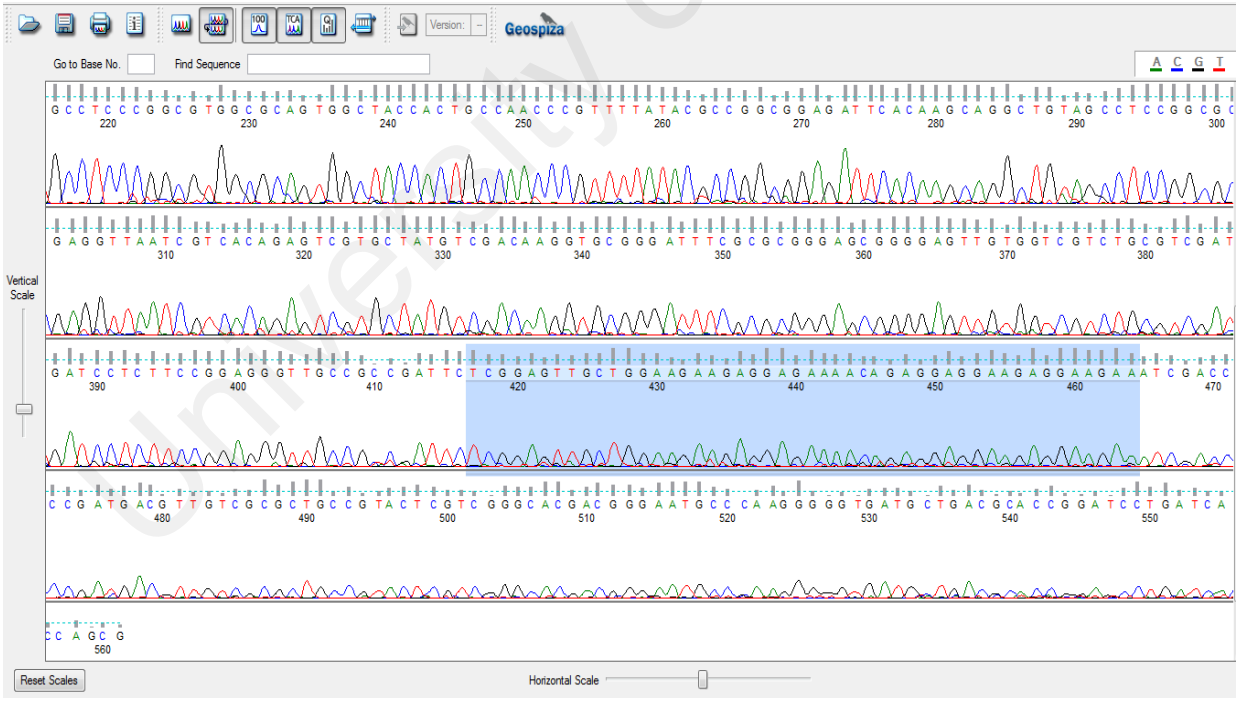


Figure 5.2: Chromatogram peak of repeated nucleotide sequences (TCGGAGTTGCTGGAAGAAGAGGAGAAAACAGAGGAGGAAGAGGAAGAA) found in open reading frame of 4CL from *B.rotunda* that code for SELLEEEEKTEEEEEE amino acid protein.

These repeated amino acid sequence may also change the 3D structure of the enzyme and may affect its functionality. Nevertheless, the common domains for 4CL enzymes such as 4CL catalytic active site, AMP binding site, putative CoA binding site and acyl-activating enzyme consensus motif were detected on the amino acid sequence. The conserved putative AMP binding domain has been used as the most important criterion to group certain proteins in one superfamily of adenylate-forming enzymes (Fulda et al., 1994). The full length of *Br4CL* that has been recovered was 1976 bp comprising of the open reading frame at 1578bp in size. In addition, the conserved sequences for most of 4CL enzyme in other plants were also present in *Br4CL* sequences. This was in line with studies conducted by Stuible & Kombrink, (2001) and Schneider et al., (2003), where the AMP binding domains (PFSSGTTGLPKGV) and the motif (GEICIRG) called Box 1 and Box II were detected respectively on the sequences and also VPP and PVL domains (**Appendix C13**). Box I, described as a putative nucleotide-binding motif, has been used as a signature for the superfamily of adenylate-forming enzymes (Babbit et al., 1992), whereas the conservation of the box II motif seemed to be restricted to 4CLs (Becker – andrea et al., 1991).

As for the phylogenetic analysis, previous studies have shown that, in dicots plant, 4CL can be divided into two types of classes; type I class is responsible for providing precursor intermediates for lignin formation while type II class responsible for non-lignin phenylpropanoids intermediates formation such as flavonoid pathways that produce flower pigments and production of defence mechanisms metabolites (Ehlting et al., 1999). The 4CLs from monocots can also be classified into two groups, which are designated as monocots class I and monocots class II as suggested by Xu and co-workers (2011) in switchgrass study. The *4CL* genes in monocots class I are associated with lignin accumulation as shown in *Panicum vaticum* plant where RNA interference of *Pv4CL1* reduces the activity of extractable 4CL by 80% leading to a reduction in lignin content and also the guaiacyl unit

composition (Xu et al. 2011). However, *4CL* genes in monocots class II are likely to be involved in the flavonoid biosynthetic pathway based on phylogenetic analysis by Xu and co-workers (2011) where they hypothesized that *Pv4CL2* may participates in the flavonoid biosynthesis pathway in switchgrass. In addition, in rice, a model species for monocots plant, 5 isoforms of *4CL* genes were identified where 4 of the isoforms (Os*4CL*1/3/4/5) were grouped in separate cluster namely type III, which was distinct from the lignin-associated type I 4CLs found in dicots while one isoform of 4CL which was *Os4CL2* belonged to the same clade as type II 4CLs in dicots (Gui et al., 2011). Besides that, Kumar and Ellis (2003), suggested that their novel outcome from the phylogenetic analysis study on *4CL* gene family of *Rubus idaeus* where putative "class III" of *4CL* genes which consisted of 4CL sequences from monocot plants, represented by ryegrass 4CL2, ryegrass 4CL3 and rice 4CL1 were found to be clustered together. In summary, it was shown that some of the 4CL isoforms from monocots were separated into different type of classes from the dicot's 4CL and exhibited their own separate cluster of class.

Br4CL sequence and some of the 4CL isoforms from *M.acuminata subsp malaccensis* have been added for the phylogenetic analysis to cluster the type of classes of *Br4CL*. Interestingly, *Br4CL* and isoforms of 4CL from banana were grouped together and were separated from the rest of the other sequences (Class I and II) suggesting close evolutionary relationship between them (**Figure 4.25**). This might indicate different biological functions of 4CL in monocots other than being involved in monolignol synthesis and flavonoid synthesis. This finding was expected since *B.rotunda* and *M.acuminata* shared the same order of Zingiberaceae and were categorised as monocots plant. In addition monocotyledons plant contains different monolignol compositions and different wall-bound phenolic compounds compared to dicots (Hatfield et al., 2009). Thus, this may result in the differences in the phylogenetic class of *4CL* genes between dicot and monocot plants.

The functionality and the substrate preferences need to be done for further characterisation of *Br4CL* sequences involved in either lignin or flavonoids biosynthesis. Recent studies of 4CL isoform in Arabidopsis has showed that despite having 4 isoforms of 4CL, mutagenesis analysis showed the functionality of the isoform could be overlapping yet has different roles in phenylpropanoid metabolism (Li et al., 2015). Furthermore, the enzymatic activities of 4CL members from Arabidopsis, aspen, loblolly pine, soybean tobacco, and many other species were found to have different substrate specificities towards cinnamic acid derivatives (Hu et al., 1998; Lindermayr et al., 2003; Voo et al., 1995). As the 4CL catalytic kinetics vary among species, it is also likely that the mainstream pathway mediated by 4CL may not be exactly the same in all plant species or tissues. Nevertheless, monolignol biosynthesis is tightly controlled by 4CL compared to flavonoid (Khan, 2012). In addition, previous studies showed that the K_m of 4CL for flavonoids was usually lower than that for monolignols (Ehlting et al., 1999; Lindemayr et al., 2003). Whether the substrate specificity of the 4CL members relates to different metabolic pathways that synthesized particular products is still unknown (Khan, 2012).

The substrate specificity of 4CL enzyme may largely depend on the sequences between the two conserved boxes (Box I and Box II) as suggested by many researchers. For instance, the structure study of peptide synthetase conducted by Conti and co-workers (1997), where the location of substrate binding pocket constituents of PheA (phenylalanine activating domain) of gramicidin *S*-synthetase 1 was identified within a 100-amino acid residue-comprising region, which is flanked by the conserved core motifs A3 and A6 of peptide synthetases exhibiting high sequence identity with the Box I (LPFSSGTTGLPKG) and Box II (GEICIRG) sequences of 4CLs, respectively. Therefore, the region between these two boxes was postulated to be the substrate specificity region that determines the substrate specificities of 4CL in activating various hydroxylated and methoxylated cinnamic acid

derivatives to the corresponding thiol esters. Besides that, according to Schneider and co-workers (2003), 12 amino acid residues located between these two boxes (Box I and Box II) are responsible for the formation of the substrate binding pocket and substrate specificity of 4CL. The 12 amino acid residues of *Br4CL* are shown in **Appendix C13**. Furthermore, study on mutagenesis on the two boxes (Box I and Box II) of conserve region in *Arabidopsis* showed no significant changes on the enzyme activity which showed that the substrate binding specificity may not involve these two boxes. On the other hand, a unique sequence that only can be found in class II 4CL and are absent from Class I sequences, which is N-terminal extensions of 19–40 amino acids, might be involved in phenolic substrate binding specificity as proposed by Hu *et al.*, (1998) but still need to be confirmed experimentally. Most 4CL enzyme favours coumarate acid and rarely can activate other cinnamate derivatives such as ferulic acid or sinapic acid where substrate preferences was control by size exclusion mechanism while substrate preferences for cinnamic acids is mainly dependent on the hydrophobicity of the substrate binding pocket (Weitzel & Petersen, 2010). In addition study by Schneider and co-workers (2003) has shown that 4CL lacking valine^{341} or leucine^{342} could activate sinapic acid to the corresponding Coenzyme A thioester, where the location of valine and leucine was between the conserved box I and II.

Therefore in conclusion, the primary function of *Br4CL* still remains unclear as whether it is involved in flavonoid biosynthesis or monolignols biosynthesis. The substrate preferences of *Br4CL* must be identified prior to its overexpression for flavonoid production in both bacteria and plant systems. This could be done by heterologously expressed *Br4CL* in *E.coli* and subsequently conduct the enzyme assay using various substrates.

In terms of bioactive compound production, a botanical categorisation based on families and genera of the plants producing the bioactive compounds might also be relevant,

as closely related plant species most often produce the same or chemically similar bioactive compounds. However, there are also ranges of examples that species even genetically less related produce similar secondary compounds (Bernhorf et al., 2010)

4CL genes usually exist as a multi-gene family. Four *4CL* genes were detected in the Arabidopsis genome (Raes et al., 2003), two isoform of *4CL* genes in tobacco (Lee and Douglas 1996) and two *4CL* genes in aspen (Hu et al., 1988). Therefore, there is a possibility of the existence of multiple isoforms of *4CL* genes in *B.rotund*a. Determining the relevant isoform of *B.rotunda* responsible in employing the divergence of activated hydroxycinnamoyl-CoA thioesters to the specific branch of phenylpropanoid pathways in plants could be pertinent.

Alignment of amino acids sequence between *B.rotunda* and *Musa acuminata subsp malaccensis* showed unconserved region (repeated sequence) at the beginning of the 4CL protein sequences of *B.rotunda* plant as shown in red boxes in **Figure 5.1**. The chromatogram showed a single peak of the nucleotides sequences in **Figure 5.2** confirming the existence of unique sequences in *4CL B.rotunda* that may have specific function in this species.

5.4 Expression of *4CL* gene in *E.coli* BL21 (DE3)

Further characterization of the gene was done by transforming the coding region of *Br4CL* into *E.coli* cells. The *E.coli* strain of BL21 (DE3) cells was selected for protein expression study as it was widely used for the heterologous expression especially involving the general phenylpropanoid key enzymes such as PAL, CHS and 4CL (Hwang et al., 2003). The uniqueness of *E. coli* strain BL21 (DE3) was the presence of T7 polymerase encoding gene as well as deletion of *Lon* and *OmpT* protease, where this protease will degrade foreign proteins and may stabilize the recombinant proteins (Gopal and Kumar, 2013). The T7 system present in pET vectors is one of the typical strategies that has been proven to be effectively express enzyme (Leonard et al., 2008). Although heterologous expression of certain eukaryotic proteins in *E. coli* may be difficult or even impossible due to lack of ability to add typical posttranslational modifications to eukaryotic proteins, improper folding of heterologously-expressed proteins and possibility of formation of inclusion bodies, nevertheless they were still the best choice because of its ability to secrete the product into the growth medium (through secretion), cytoplasm or periplasm (Khow et al., 2012). Their well-known physiology and characterized genetics information, high growth rates, availability of abundant genetic tools, cost effectiveness and ease of conduct has made microbial cell the first choice to researchers to study protein expression (Du et al., 2011). Therefore, using *E. coli* as a microbial factory has extensively attracted interest in the development of a novel system for modification and efficient production of highly useful compounds.

In this study, *Br4CL* protein was successfully expressed in *E.coli* BL2 (DE3) strain. Li et al, (2014) also showed that 4CL protein could be expressed in *Pueraria lobata* cells. These findings proved that 4CL was not toxic to *E.coli.* SDS-PAGE analysis was performed to verify the expression of *Br4CL* protein in *E.coli*. The purpose of SDS-PAGE is to separate

proteins according to molecular weight where *Br4CL* is about 70kb. In this analysis, sodium dodecyl sulphate (SDS) which is a detergent (soap) was used to dissolve hydrophobic molecules and also has negative charges (sulphate) attached to it. Therefore, when *E.coli* cells were incubated with SDS, the membranes were dissolved; all the proteins were solubilized and covered with many negative charges. Polymerised polyacrylamide was used to separate proteins according to the molecular weight and the migrations of these proteins were influenced by an applied electrical field.

For future studies, substrate preferences and functionality of the 4CL enzyme from *B.rotunda in-vivo* could be examined by supplying different substrates in the recombinant microbial cells and the products were analysed through HPLC. In addition, *in-vitro* enzyme assays could be carried out to determine the substrate preferences by comparing its kinetic activity towards different substrates of cinnamate derivatives. This preliminary information could be used as a basis for further work on the production of targeted bioactive compounds in microbial cells.

There were studies conducted where recombinant microbial cells harbouring multi-genes from various organisms including *4CL*, were able to produce targeted compounds. This was as a proof to the ability of the gene expressions and function in bacteria cells (Hwang et al., 2003; Kaneko et al., 2003; Watts et al., 2004; Katsuyama et al., 2007). *4CL* gene has been used in many heterologous experiments especially in phenylpropanoid biosynthesis as this enzyme catalysed the formation of thioesters which act as key intermediates for many other subsequent targeted compounds. Phenylpropanoid-derived-end compounds were widely produced in heterologous/co-expression study in prokaryote cells as these compounds have important nutritional roles in the human diet (Korkina et al., 2007).

This strategy can be used not only to produce the final target compounds but also to produce intermediates or phenylpropanic precursors. Santos et al., (2011) applied a four-step

heterologous pathway consisting of TAL, 4CL, CHS, and CHI enzymes within an engineered L-tyrosine *E. coli* overproducer to produce the main flavonoid precursor naringenin directly by feeding with glucose. Thus, instead of feeding the media with naringenin produced from chemical synthesis, it was cheaper to use glucose. This could help overcome problems related to the lack of precursor molecules that would eventually lead to lower production of target compounds. Once flavonoid biosynthetic pathway was heterologously introduced and expressed in such hosts, the precursor molecules decreased and were limited within the cell.

Aside from precursors, the lack of co-factors such as malonyl-CoA could relatively result in low production of the target compound and cell growth rate because of the metabolic burden within the cells (Miyahisa et al., 2005). To overcome this, the host could be modified by deleting or blocking some essential genes to enhance the carbon flux through the pathway of interest or by manipulating the carbon pathway of the host cell itself.

CHAPTER 6: CONCLUSION

In this study the full length sequence of *4CL* gene from *B. rotunda* has been successfully isolated from the cell suspension cultures. *4CL* gene that codes for 4-coumarate: coenzyme A Ligase protein converts 4-coumarate and other cinnamate derivatives to respective CoA esters and indirectly has a role in the general phenylpropanoid pathway that produces many secondary metabolites including flavonoids. This protein has been characterized and isolated from many other plants and was conserved in terms of their domains and catalytic active site. Sequence analysis of *Br4CL* showed high similarity with *4CL* gene in other plants and displayed the highest homology with *4CL* gene from *Musa acuminata subsp malaccensis* which is also a monocotyledonous plant. They both share the same Zingiberaceae order with *B. rotunda* and both of them are categorized as herbs. After obtaining the coding region of this gene, an expression profile was conducted in order to determine the transcript level of this gene when supplied with excessive phenylalanine in the *B.rotunda* cell suspension culture media throughout 14 days at the collection time point. The expression levels at various time points were higher in comparison to control and most elevated at day 14. The data showed that phenylalanine could be used as precursors to induce *4CL* gene transcript to be upregulated and may lead to the enhancement of certain flavonoids or any other bioactive secondary metabolites.

The sequence was then further characterized by transforming *Br4CL* gene into *E.coli* to determine the expression of eukaryotic gene in prokaryotes. Many studies have been conducted using *E.coli* as host system to express heterologous combination of genes responsible for the production of desired compound. In this study, the gene was successfully expressed in *E.coli* strains of BL21 (DE3) although the sequence has some unique region and repeated amino acid sequence. Protein modelling, *in-vitro* and *in- vivo* functionality tests of 4CL isoforms from *B.rotunda* could be done in order to understand the substrate preferences

and structural flexibility that might be responsible in determining the carbon flux pathway. To date, only the conserved region and the binding domain of the gene were known but a specific binding of substrates towards the functional group between the amino acids is still yet to be discovered. The position of 4CL at a metabolic branch point connecting general phenylpropanoid metabolism with different end product-specific pathways and as a divergence-point enzyme, connecting general phenylpropanoid metabolism to different subsequent branch pathways has made 4CL a promising target for biotechnological manipulation.

For future studies production of bioactive secondary metabolites in *B.rotunda* using *E.coli* cells as a host could be carried out by combining several biosynthetic enzymes in this plant. The idea of using sets of genes or structural enzymes from only one species could specify the pathway and help in producing true target compounds instead of using genes from different organisms.

In conclusion, this study lays a foundation not only on the role of 4CL in *B.rotunda* plants in flavonoid biosynthesis but also for future research on the phenylalanine metabolic pathway in *B. rotunda* plants. Development of transgenic *B.rotunda* plants overexpressing *4CL* gene could also be developed as an alternative source of enhanced accumulation of selected flavonoids.

REFERENCES

Banerjee, A., RY, K., JM, H., WS, L., & RA, P. (1994). Enzymic preparation of (3R-cis)-3-(acetyloxy)-4-phenyl-2-azetidinone: a taxol side-chain synthon. *Biotechnology and applied biochemistry*, *20*(1), 23-33.

Bauer, N., Kiseljak, D., & Jelaska, S. (2009). The effect of yeast extract and methyl jasmonate on rosmarinic acid accumulation in *Coleus blumei* hairy roots. *Biologia plantarum*, *53*(4), 650-656.

Babbitt, P. C., Kenyon, G. L., Martin, B. M., Charest, H., Slyvestre, M., Scholten, J. D., & Dunaway-Mariano, D. (1992). Ancestry of the 4-chlorobenzoate dehalogenase: analysis of amino acid sequence identities among families of acyl: adenyl ligases, enoyl-CoA hydratases/isomerases, and acyl-CoA thioesterases. *Biochemistry*, *31*(24), 5594-5604.

Becker-Andre´, M., Schulze-Lefert P. and Hahlbrock K. (1991) Structural comparison, modes of expression, and putative *cis*-acting elements of the two 4-coumarate: CoA ligase genes in potato. *J. Biol. Chem*. 266:8551–8559.

Bernhoft, A., Siem, H., Bjertness, E., Meltzer, M., Flaten, T., & Holmsen, E. (2010). Bioactive compounds in plants–benefits and risks for man and animals. In *Proceedings from a Symposium Held at The Norwegian Academy of Science and Letters, Novus forlag, Oslo*.

Cen, D., Gonzalez, R.I., Buckmeier, J.A., Kahlon, R.S., Tohidian, N.B., Meyskens, F.L., Jr. (2002) Disulfiram induces apoptosis in human melanoma cells: a redox-related process. *Mol. Cancer Ther*. 1:197-204.

Chahyadi, A., Hartati, R., & Wirasutisna, K. R. (2014). *Boesenbergia Pandurata Roxb*., An Indonesian Medicinal Plant: Phytochemistry, Biological Activity, Plant Biotechnology. *Procedia Chemistry*, *13*, 13-37.

Cheenpracha, S., Karalai, C., Ponglimanont, C., Subhadhirasakul, S., Tewtrakul, S., (2005) Anti-HIV-1 protease activity of compounds from *Boesenbergia pandurata*. *Bioorganic and Medicinal Chemistry* 16: 1710-1714.

Conti, E., Stachelhaus, T., Marahiel, M. A., & Brick, P. (1997). Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *The EMBO Journal*, *16*(14), 4174-4183.

Cukovica D, Ehlting J, Ziffle J, Douglas C. (2001). Structure and evolution of 4coumarate: coenzyme A ligase (4CL) gene families. *Biological Chemistry* 382: 645–654.

Dixon RA, Paiva NL (1995) Stress-induced phenylpropanoid metabolism. *Plant Cell*. 7:1085-1097

Dobritsa, A. A., Geanconteri, A., Shrestha, J., Carlson, A., Kooyers, N., Coerper, D., ... & Preuss, D. (2011). A large-scale genetic screen in Arabidopsis to identify genes involved in pollen exine production. *Plant physiology*, *157*(2), 947-970.

Du, J., Shao, Z., & Zhao, H. (2011). Engineering microbial factories for synthesis of value-added products. *Journal of industrial microbiology & biotechnology*, 38(8), 873-890.

Edahiro, J. I., Nakamura, M., Seki, M., & Furusaki, S. (2005). Enhanced accumulation of anthocyanin in cultured strawberry cells by repetitive feeding of L-phenylalanine into the medium. *Journal of bioscience and bioengineering*, *99*(1), 43-47.

Effendi Leonard, Yajun Yan, Kok Hong Lim, Mattheos A.G. Koffas (2005) Investigation of Two Distinct Flavone Synthases for Plant-Specific Flavone Biosynthesis in *Saccharomyces cerevisiae. Applied and Environmental microbiology* 8241–8248:71-12

Ehlting, J., Büttner, D., Wang, Q., Douglas, C. J., Somssich, I. E., & Kombrink, E. (1999). Three 4-coumarate: coenzyme A ligases in *Arabidopsis thaliana* represent two evolutionarily divergent classes in angiosperms. *The plant journal*, *19*(1), 9-20.

Emiliani, G., Fondi, M., Fani, R., & Gribaldo, S. (2009). A horizontal gene transfer at the origin of phenylpropanoid metabolism: a key adaptation of plants to land. *Biology Direct*, *4*(1), 1.

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), 356-370.

Fett-Netto, A. G., Melanson, S. J., Nicholson, S. A., Pennington, J. J., & DiCosmo, F. (1994). Improved taxol yield by aromatic carboxylic acid and amino acid feeding to cell cultures of *Taxus cuspidata*. *Biotechnology and bioengineering*, 44(8), 967-971.

Fulda, M., Heinz, E., & Wolter, F. P. (1994). The fadD gene of *Escherichia coli* K12 is located close to rnd at 39.6 min of the chromosomal-map and is a new member of the AMP-binding protein family. *Molecular and General Genetics MGG*, *242*(3), 241-249.

Gopal, G. J., & Kumar, A. (2013). Strategies for the production of recombinant protein in *Escherichia coli*. *The protein journal*, *32*(6), 419-425.

Grand, C., Boudet, A., and Boudet, A.M. (1983) lsoenzymes of hydroxycinnamate: CoA ligase from poplar stems- Properties and tissue distribution. *Planta* 158: 225-229.

Gui, J., Shen, J., & Li, L. (2011). Functional characterization of evolutionarily divergent 4-coumarate: coenzyme A ligases in rice. *Plant Physiology*, 157(2), 574-586.

Gutiérrez-Carbajal, M. G., Monforte-González, M., Miranda-Ham, M. D. L., GodoyHernández, G., & Vázquez-Flota, F. (2010). Induction of capsaicinoid synthesis in Capsicum chinense cell cultures by salicylic acid or methyl jasmonate. *Biologia Plantarum*, *54*(3), 430-434.

Harding, S. A., Leshkevich, J., Chiang, V. L., & Tsai, C. J. (2002). Differential substrate inhibition couples kinetically distinct 4-coumarate: coenzyme A ligases with spatially distinct metabolic roles in quaking aspen.*Plant Physiology*, *128*(2), 428-438.

Harvey, A. L. (2008). Natural products in drug discovery. *Drug discovery today*, *13*(19), 894-901.

Hatfield, R. D., Marita, J. M., Frost, K., Grabber, J., Ralph, J., Lu, F., & Kim, H. (2009). Grass lignin acylation: p-coumaroyl transferase activity and cell wall characteristics of C3 and C4 grasses. *Planta*, *229*(6), 1253-1267.

Herrmann, K. M., & Weaver, L. M. (1999). The shikimate pathway. *Annual review of plant biology*, *50*(1), 473-503.

Hu, W. J., Kawaoka, A., Tsai, C. J., Lung, J., Osakabe, K., Ebinuma, H., & Chiang, V. L. (1998). Compartmentalized expression of two structurally and functionally distinct 4-coumarate: CoA ligase genes in aspen (*Populus tremuloides*). *Proceedings of the National Academy of Sciences*, *95*(9), 5407-5412.

Hussain, M. S., Fareed, S., Ansari, S., Rahman, M. A., Ahmad, I. Z., & Saeed, M. (2012). Current approaches toward production of secondary plant metabolites. *Journal of Pharmacy and Bioallied Sciences*, *4*(1), 10.

Hwang EI, Kaneko M, Ohnishi Y, Horinouchi S (2003) Production of plant-specific flavanones by *Escherichia coli* containing an artificial gene cluster. *Appl Environ Microbiol* 69:2699–2706.

Isa N, Abdelwahab S, Mohan S, Abdul A, Sukari M, Taha M, Syam S, Narrima P, Cheah SC, Ahmad S. (2012) In *vitro* anti-inflammatory, cytotoxic and antioxidant activities of boesenbergin A, a chalcone isolated from *Boesenbergia rotunda* (L.)(fingerroot). *Braz J Med Biol Res* 45(6):524-530.

Jaipetch, T., Kanghae, S., Pancharoen, O., Patrick, V.A., Reutrakul, V., Tuntiwachwuttikul, P., White, A.H., (1982) Constituents of *Boesenbergia pandurata* (syn. *Kaempferia pandurata*): Isolation, crystal structure and synthesis of (±)-Boesenbergin A. *Aust. J. Chem.* 35: 351-361.

Ji, H. F., Li, X. J., & Zhang, H. Y. (2009). Natural products and drug discovery. EMBO reports, 10(3), 194-200.

Kaneko M, Ohnishi Y, Horinouchi S (2003) Cinnamate: coenzyme A ligase from the filamentous bacterium *Streptomyces coelicolor* A3(2). *J Bacteriol* 185:20–27.

Kaneko, M., Hwang, E. I., Ohnishi, Y., & Horinouchi, S. (2003). Heterologous production of flavanones in *Escherichia coli*: potential for combinatorial biosynthesis of flavonoids in bacteria. *Journal of Industrial Microbiology and Biotechnology*, *30*(8), 456-461.

Katsuyama, Y., Funa, N., Miyahisa, I., & Horinouchi, S. (2007). Synthesis of unnatural flavonoids and stilbenes by exploiting the plant biosynthetic pathway in *Escherichia coli*. *Chemistry & biology*, 14(6), 613-621.

Khan, B. M. (2012). Genetic engineering of phenylpropanoid pathway in *L. leucocephala*

Khow, O., & Suntrarachun, S. (2012). Strategies for production of active eukaryotic proteins in bacterial expression system. *Asian Pacific journal of tropical biomedicine*, *2*(2), 159-162.

Knobloch, K. H., Hahlbrock K. (1975) Isoenzymes of *p*-coumarate:CoA ligase from cell suspension cultures of *Glycine max*. *Eur. J. Biochem*. 52:311–320.

Korkina, L. G. (2007). Phenylpropanoids as naturally occurring antioxidants: from plant defense to human health. *Cell Mol Biol*, *53*(1), 15-25.

Kováčik, J., Kron, I., Repčák, M., & Bačkor, M. (2007). Effect of feeding precursors on phenylalanine ammonia-lyase activity and coumarin accumulation in leaves of *Matricaria chamomilla L*. *Plant Growth Regulation*,*52*(1), 9-15.

Kris-Etherton, P. M., Hecker, K. D., Bonanome, A., Coval, S. M., Binkoski, A. E., Hilpert, K. F., ... & Etherton, T. D. (2002). Bioactive compounds in foods: their role in the prevention of cardiovascular disease and cancer. *The American journal of medicine*, 113(9), 71-88.

Kumar, A., & Ellis, B. E. (2003). 4-Coumarate: CoA ligase gene family in *Rubus idaeus*: cDNA structures, evolution, and expression. *Plant molecular biology*, *51*(3), 327-340.

Lall, R. K., Adhami, V. M., & Mukhtar, H. (2016). Dietary flavonoid fisetin for cancer prevention and treatment. *Molecular nutrition & food research*, 60(6), 1396-1405.

Landis, B. H., Mullins, P. B., Mullins, K. E., & Wang, P. T. (2002). Kinetic resolution of β-amino esters by acylation using immobilized penicillin amidohydrolase. *Organic process research & development*, 6(4), 539-546

Lee D, Douglas CJ (1996) Two divergent members of a tobacco 4-coumarate: coenzyme A ligase (4CL) gene family. cDNA structure, gene inheritance and expression, and properties of recombinant proteins. *Plant Physiol* 112(1):193–205

Lee D, Ellard M, Wanner LA, Davis KR, Douglas CJ (1995) The *Arabidopsis thaliana* 4-coumarate: CoA ligase (4CL) gene: stress and developmentally regulated expression and nucleotide sequence of its cDNA. *Plant Mol Biol* 28(5):871–884

Leonard, E., Yan, Y., Fowler, Z. L., Li, Z., Lim, C. G., Lim, K. H., & Koffas, M. A. (2008). Strain improvement of recombinant *Escherichia coli* for efficient production of plant flavonoids. *Molecular pharmaceutics*, *5*(2), 257-265.

Li, Y., Im Kim, J., Pysh, L., & Chapple, C. (2015). Four isoforms of *Arabidopsis thaliana* 4-coumarate: CoA ligase (4CL) have overlapping yet distinct roles in phenylpropanoid metabolism. *Plant physiology*, pp-00838.

Li, Z. B., Li, C. F., Li, J., & Zhang, Y. S. (2014). Molecular cloning and functional characterization of two divergent 4-coumarate: coenzyme A ligases from kudzu (*Pueraria lobata*). *Biological and Pharmaceutical Bulletin*, 37(1), 113-122

Lindermayr, C., Fliegmann, J., & Ebel, J. (2003). Deletion of a single amino acid residue from different 4-coumarate: CoA ligases from soybean results in the generation of new substrate specificities. *Journal of Biological Chemistry*, 278(5), 2781-2786.

Liu J.J., Goh C.J., Loh C.S., Liu P., Pua E.C. (1998) A method for isolation of total RNA from fruit tissues of banana. *Plant Mol. Biol. Reptr.* 16:1-6.

Liu R, Gao M, Zhi-Hong Y, Guan-Hua D. (2008) Pinocembrin protects rat brain against oxidation and apoptosis induced by ischemia-reperfusion both *in vivo* and *in vitro*. *Eur J Pharmacol* 1216:104-15.

Lozoya, E., Hoflmann, H., Douglas, C., Schulz, W., Scheel, D., Hahlbrock, K. (1988) Primary structures and catalytic properties of isoenzymes encoded by two 4-coumarate:CoA ligase genes in parsley, *Eur. J. Biochem*. 176: 661-667.

Mahmoudian, M., Baines, B. S., Drake, C. S., Hale, R. S., Jones, P., Piercey, J. E., ... & Lawrence, G. C. (1993). Enzymatic production of optically pure (2′ R-cis)-2′-deoxy-3′-thiacytidine (3TC, Lamivudine): a potent anti-HIV agent. *Enzyme and microbial technology*, 15(9), 749-755

Manela, N., Oliva, M., Ovadia, R., Sikron-Persi, N., Ayenew, B., Fait, A., ... & Oren-Shamir, M. (2015). Phenylalanine and tyrosine levels are rate-limiting factors in production of health promoting metabolites in *Vitis vinifera* cv. Gamay Red cell suspension. *Frontiers in plant science*, *6*.

Marienhagen, J., & Bott, M. (2013). Metabolic engineering of microorganisms for the synthesis of plant natural products. *Journal of biotechnology*, 163(2), 166-178.

Md-Mustafa ND, Khalid N, Gao H, Peng Z, Alimin MF, Bujang N, Ming WS, Mohd-Yusuf Y, Harikrishna JA, Othman RY (2014) Transcriptome profiling shows gene regulation patterns in a flavonoid pathway in response to exogenous phenylalanine in *Boesenbergia rotunda* cell culture. *BMC Genomics* 15:984

Miyahisa, I., Kaneko, M., Funa, N., Kawasaki, H., Kojima, H., Ohnishi, Y., & Horinouchi, S. (2005). Efficient production of (2S)-flavanones by *Escherichia coli* containing an artificial biosynthetic gene cluster. *Applied microbiology and biotechnology*, *68*(4), 498-504.

Murashige T, Skoog F (1962) A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiol Plant* 15:473–497.

Pandey, R. P., & Sohng, J. K. (2013). Genetics of flavonoids. In *Natural Products* (pp. 1617-1645). Springer Berlin Heidelberg.

Patel, R. N. (2001). Enzymatic synthesis of chiral intermediates for Omapatrilat, an antihypertensive drug. *Biomolecular Engineering*, 17(6), 167-182.

Raes, J., Rohde, A., Christensen, J. H., Van de Peer, Y., & Boerjan, W. (2003). Genome-wide characterization of the lignification toolbox in Arabidopsis. *Plant Physiology*, *133*(3), 1051-1071.

Santos, C. N. S., Koffas, M., & Stephanopoulos, G. (2011). Optimization of a heterologous pathway for the production of flavonoids from glucose. *Metabolic engineering*, *13*(4), 392-400.

Schneider, K., Hövel, K., Witzel, K., Hamberger, B., Schomburg, D., Kombrink, E., & Stuible, H. P. (2003). The substrate specificity-determining amino acid code of 4-coumarate: CoA ligase. *Proceedings of the National Academy of Sciences*, *100*(14), 8601-8606.

Siddiqui, M. S., Thodey, K., Trenchard, I., & Smolke, C. D. (2012). Advancing secondary metabolite biosynthesis in yeast with synthetic biology tools. *FEMS yeast research*, 12(2), 144-170.

Sivakumar, G., Krishnamurthy, K. V., Hao, J., & Paek, K. Y. (2004). Colchicine production in *Gloriosa superba* calluses by feeding precursors. *Chemistry of natural compounds*, *40*(5), 499-502.

Stead, P., Marley, H., Mahmoudian, M., Webb, G., Noble, D., Ip, Y. T., ... & Dawson, M. J. (1996). Efficient procedures for the large-scale preparation of (1S, 2S)-trans-2-methoxycyclohexanol, a key chiral intermediate in the synthesis of tricyclic β-lactam antibiotics. *Tetrahedron: Asymmetry*, *7*(8), 2247-2250.

Stuible, H. P., & Kombrink, E. (2001). Identification of the substrate specificity-conferring amino acid residues of 4-coumarate: coenzyme A ligase allows the rational design of mutant enzymes with new catalytic properties. *Journal of Biological Chemistry*, *276*(29), 26893-26897.

Sullivan, M. L. (2009). Phenylalanine ammonia lyase genes in red clover: expression in whole plants and in response to yeast fungal elicitor. *Biologia Plantarum*, *53*(2), 301-306.

Sun H, Li Y, Feng S, Zou W, Guo K, Fan C, Si S, Peng L (2013) Analysis of five rice 4-coumarate: coenzyme A ligase enzyme activity and stress response for potential roles in lignin and flavonoid biosynthesis in rice. *Biochem Biophys Res Commun* 430:1151–1156

Tamura, K., Dudley, J., Nei, M., & Kumar, S. (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular biology and evolution*, *24*(8), 1596-1599.

Tanaka, T., de Azevedo, M., Durán, N., Alderete, J. B., Epifano, F., Genovese, S., ... & Curini, M. (2010). Colorectal cancer chemoprevention by 2 β-cyclodextrin inclusion compounds of auraptene and 4′-geranyloxyferulic acid. *International Journal of Cancer*, *126*(4), 830-840.

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*, *2015*.

Tan EC, Karsani SA, Foo GT, Wong SM, Rahman NA, Khalid N, Othman S, 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 Tiss Organ Cult* 111:219–229

Trakoontivakorn G, Nakahara K, Shinmoto H, Takenaka M, Onishi-Kameyama M, Ono H, Yoshida M, Nagata T, Tsushida T (2001)Structural analysis of a novel anti mutagenic compound, 4-hydroxypanduratin A, and the antimutagenic activity of flavonoids in a Thai spice, fingerroot (*Boesenbergia pandurata Schult*.) against mutagenic heterocyclic amines. *J Agric FoodChem* 49:3046–305

Van Der Krol, A. R., Mol, J. N., & Stuitje, A. R. (1988). Modulation of eukaryotic gene expression by complementary RNA or DNA sequences.*Biotechniques*, *6*(10), 958-976.

Voo, K.S., Whetten, R., OMalley, D., Sederoff, R.R. (1995). 4-Coumarate: CoA ligase in xylem of loblolly pine. *Plant Physiol*. 108:85-97

Wang, C. H., Yu, J., Cai, Y. X., Zhu, P. P., Liu, C. Y., Zhao, A. C., ... & Yu, M. D. (2016). Characterization and Functional Analysis of 4-Coumarate: CoA Ligase Genes in Mul-berry. *PloS one*, *11*(5), e0155814

. Watts, K. T., Lee, P. C., & Schmidt-Dannert, C. (2004). Exploring recombinant flavonoid biosynthesis in metabolically engineered *Escherichia coli*.*Chembiochem*, *5*(4), 500-507.

Watts, K. T., Lee, P. C., & Schmidt-Dannert, C. (2006). Biosynthesis of plant-specific stilbene polyketides in metabolically engineered *Escherichia coli*. *BMC biotechnology*, *6*(1), 22.

Weitzel, C., & Petersen, M. (2010). Enzymes of phenylpropanoid metabolism in the important medicinal plant *Melissa officinalis L*. *Planta*, *232*(3), 731-742.

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

Xu, B., Escamilla-Treviño, L. L., Sathitsuksanoh, N., Shen, Z., Shen, H., Percival Zhang, Y. H., ... & Zhao, B. (2011). Silencing of 4-coumarate: coenzyme A ligase in switchgrass leads to reduced lignin content and improved fermentable sugar yields for biofuel production. *New phytologist*, *192*(3), 611-625.

Yusuf, N. A., Annuar, M. S. 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.

Zhang, M. M., Wang, Y., Ang, E. L., & Zhao, H. (2016). Engineering microbial hosts for production of bacterial natural products. *Natural product reports*

APPENDICES

Appendix A: Stock solutions, buffers and media of cell suspension cultures

Murashige and Skoog (MS) media powders contained macronutrients, micronutrients, iron source, and vitamins were purchased from Bumi Sains Company.

A1: Stock solutions, buffers and media for *B. rotunda* cell suspensions

(1) Cell suspension media (SM)

MS media powder	4.4g
Maltose	5g
Sucrose	30g
Myo-inositol	0.1g
Malt extracts	0.15g
Biotin (1mg/ml)	1 ml
NAA (1mg/ml)	1 ml
6-BA (1mg/ml)	1 ml
2, 4-D (1mg/ml)	1 ml

The chemical component were dissolved in 800 ml distilled H_2O and stirred continuously while adjusting the pH to 5.8 with 1M NaOH. The volume was adjusted to 1000 ml and sterilized by autoclaving. 4 ml of (1mg/ml) of L-glutamine were added before use.

(2) Callus induction media (CIM)

MS media powder	4.4g
Gel rite	2g
Biotin (1mg/ml)	1 ml
NAA (1mg/ml)	1 ml
2, 4-D (1mg/ml)	1 ml

The chemical component were dissolved in 800 ml distilled H_2O and stirred continuously while adjusting the pH to 5.8 with 1M NaOH. The volume was adjusted to 1000 ml and sterilized by autoclaving. 1 ml of (1mg/ml) of IAA was added before use.

(3) Propogation media (PM)

MS media powder	4.4g

2, 4-D (3mg/ml)	1 ml
Gel rite	3g

The chemical component were dissolved in 800ml distilled H_2O and stirred continuously while adjusting the pH to 5.8 with 1M NaOH. The volume was adjusted to 1000 ml and sterilized by autoclaving.

A2: Stock solutions, buffers and gel for electrophoresis

(1) 5x TBE buffer for DNA use

Tris base	54g
Boric acid	27.5g
0.5M EDTA (pH 8.0)	20 ml
Distilled H_2O to	1000 ml

Tris base, boric acid and EDTA were dissolved in distilled water and adjusted to the final volume of 1000ml before sterilization. Then, the sterile solution was kept in room temperature. The solution was diluted to 1x working concentration (90mM Tris-borate and 2mM EDTA) for agarose gel electrophoresis

(2) 5x TBE buffer for RNA use

The solution was prepared as mentioned above but the distilled water used for topup purpose was autoclaved two times instead of one time

(3) 6X DNA loading dye

1.0 M Tris- Cl (pH 7.6)	100 µl
Bromophenol blue	30mg
Xylene cyanol FF	30mg
0.5M EDTA	1.2 ml
Glycerol	6 ml

The compositions (as stated in Fermentas catalogue) were mixed well and adjusted to final volume of 10ml with sterile distilled water. The loading buffers were kept at room temperature

(4) 2x RNA loading dye

The loading dye was purchased from Fermentas with the composition of 95% formamide, 0.025% SDS, 0.025 % bromophenol blue, 0.025% xylene cyanol FF, 0.025 % ethidium bromide and 0.5mM EDTA. The loading dye was kept at -20^0C.

(5) Molecular weight markers (Ladder)

The Perfect^{TM} 100bp DNA ladder and Perfect Plus^{TM} 1kb DNA ladder were purchased from $EUR_{X;}$ while ssRNA ladder was purchased from NEB. All the ladders were kept at -20^0C

(6) Agarose gel

For preparing 1% agarose gel, 0.2 g of the agarose was added into 20ml of 1x TBE buffer. Approximately 2 µl of ethidium bromide were added into the gel

A3: Stock solutions and buffers of conventional PCR

(1) PCR reaction mixture (7x reaction)

The HotStarTaq *Plus* DNA polymerase kit was purchased from Qiagen. The PCR reaction mixture was prepared as showed in Table A1

Table A1: 7x PCR reaction mixture

Components	Volume/reaction (µl)	Final concentration
10x CoralLoad PCR buffer (with 15mM $MgCl_2$)	10	1x
dNTP mix (5mM of each)	4	200µM of each dNTP
Downstream primer (10µM)	5	0.5 µM
Upstream primer (10µM)	5	0.5 µM
Template	3	$\leq$ 1µg/reaction
HotStarTaq *Plus* DNA polymerase (2.5 U/µl)	0.5	2.5 units/reaction
Sterile distilled H_2O	72.5	-
Total reaction volume	100	-

Note: 10 µl was taken out from the mixture for negative control before adding 3 µl of cDNA as a template and 15 µl were loaded into 6 PCR tubes

A4: Stock solutions and buffers of Reverse-transcriptase PCR

(1) Reverse-transcriptase PCR reaction mixture

The first strand of cDNA was synthesized using a cDNA synthesis Kit (EURx Ltd, Poland). The kit consists of 5x cDNA buffer, Random hexamers primers, 5mM dNTP mix, RNase Inhibitor 12.5 U/µl and dART Reverse Transcriptase. The reactiom mixture was prepared as shown in Table A2

Table A2: 1x Reverse-transcriptase PCR reaction mixture

Component	Volume in 1x reaction (µl)
5x cDNA buffer	4
5mM dNTP mix	4
Random hexamers primers	1
dART Reverse Transcriptase.	1
RNase Inhibitor 12.5 U/µ	1
RNAse-free water	to 13
RNA (10ng- 5µg)	-
Total	20

A5: Stock solutions and buffers of Real-time PCR

(1) Real-time PCR reaction mixture

Power SYBR Green PCR master mix (2X) was purchased from Applied Biosystems. The reaction mixture was prepared as shown in Table A3

Table A3: 1x RT-qPCR reaction mixture

Component	Volume for standard 96 well plate (µl)
Power SYBR green PCR Master mix 2X	6.25
10 µM Forward primer of 4CL(200 nM)	0.625
10 µM Reverse primer of 4CL(200 nM)	0.625
Nuclease free water	4.22
cDNA template (100 ng/µl)	0.78
Total volume	12.5

A6: Agarose gel electrophoresis

(1) Agarose gel electrophoresis for DNA

Agarose gel was prepared as stated in Appendix A2 () and transferred to a gel tank which contained 1x TBE buffer. The samples, each 5 µl, were mixed with 1 µl of 6x loading dye on parafilm and loaded into the wells. After that, the samples were electrophoresed at 120 V for 30 minutes. The gel was visualized under ultraviolet light transilluminator of Gel-Pro IMAGER (MICROlambda) and photograph was taken for result

(2) Agarose gel electrophoresis for RNA

The same protocol was used as stated in Appendix A4 (1) with the 5X TBE buffer of RNA used. However, there are some differences. The RNA samples were mixed well with 2x RNA loading dye in equal volume in 1.5ml microcentrifuged tubes. After that, the samples were loaded into the well. . After that, the samples were electrophoresed at 120 V for 30 minutes. The gel was visualized under ultraviolet light transilluminator of Gel-Pro IMAGER (MICROlambda) and photograph was taken for result

Appendix B: Compositions of Murashige and Skoog (MS)

Essential element	Concentration in medium (mg/l)
Macroelement	
Ammonium nitrate (NH_4NO_3)	1650
Kalium nitrate (KNO_3)	1900
Calcium chloride ($CaCl_2.2H_2O$)	440
Magnesium sulphate ($MgSO_4.7H_2O$)	370
Potassium phosphate (KH_2PO_4)	170
Microelement	
Potassium iodide (KI)	0.83
Boric acid (H_3BO_3)	6.2
Manganase sulphate($MnSO_4.4H_2O$)	22.3
Zinc sulphate($ZnSO_4.\ 7H_2O$)	8.6
Sodium manganate($Na_2MnO_4.2H_2O$)	0.25
Copper sulphate($CuSO_4.\ 5H_2O$)	0.025
Cobalt chloride($CoCl_2.\ 6H_2O$)	0.025
Iron source	
Ferum sulphate ($FeSO_4.7H_2O$)	27.8
Ethylenediamineetetraacetic acid disodium ($Na_2EDTA.2H_2O$)	37.3
Organic supplement	
Myoinositol	100
Nicotinic acid	0.5
Phyridoxine-HCl	0.5
Thiamine-HCl	0.5
Glycine	2
Carbon source	
Sucrose	30000

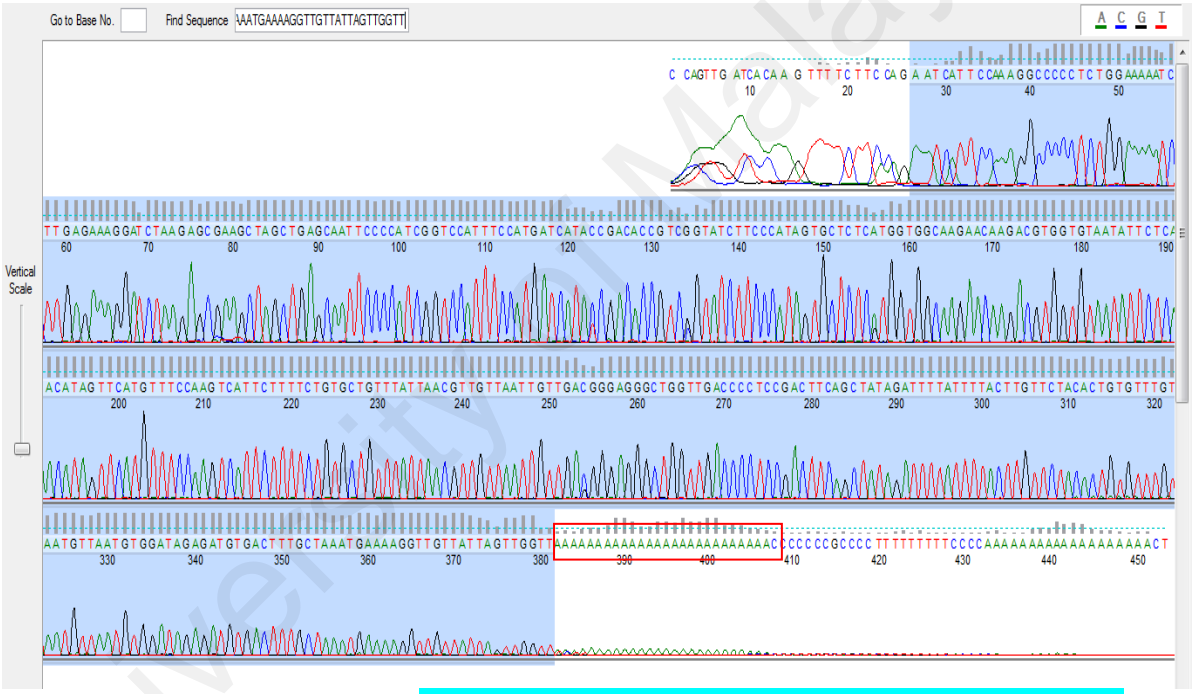
Appendix C: Sequences and Multiple alignment of amino acid sequence

Appendix C1: RACE-PCR method (3' RACE)

4CL - 428bp (Unigene36813_All from transcriptome data)

Blue highlight: Overlap region between forward sequence of 3' RACE PCR product and Unigene36813_All

GAAATCAAGCAGTACATCTCAAGACAGGTGGTTTTCTACAAGAGAATCAACAAAGTTTTCTTCACAGAA
TCATTCCAAAGGCCCCCTCTGGAAAAATCTTGAGAAAGGATCTAAGAGCGAAGCTAGCTGAGCAATTCCC
CATCGGTCCATTTCCATGATCATACCGACACCGTCGGTATCTTCCCATAGTGCTCTCATGGTGGCAAGAA
CAAGACGTGGTGTAATATTCTCAACATAGTTCATGTTTCCAAGTCATTCTTTTCTGTGCTGTTTATTAAC
GTTGTTAATTGTTGACGGGAGGGCTGGTTGACCCCTCCGACTTCAGCTATAGATTTTATTTTACTTGTTC
TACACTGTGTTTGTAATGTTAATGTGGATAGAGATGTGACTTTGCTAAATGAAAAGGTTGTTATTAGTTG
GTTTCTTA



CCAGTTGATCACAAGTTTTCTTCCAGAATCATTCCAAAGGCCCCCTCTGGAAAAATCTTGAGAAAGGATC
TAAGAGCGAAGCTAGCTGAGCAATTCCCCATCGGTCCATTTCCATGATCATACCGACACCGTCGGTATCT
TCCCATAGTGCTCTCATGGTGGCAAGAACAAGACGTGGTGTAATATTCTCAACATAGTTCATGTTTCCAA
GTCATTCTTTTCTGTGCTGTTTATTAACGTTGTTAATTGTTGACGGGAGGGCTGGTTGACCCCTCCGACT
TCAGCTATAGATTTTATTTTACTTGTTCTACACTGTGTTTGTAATGTTAATGTGGATAGAGATGTGACTT
TGCTAAATGAAAAGGTTGTTATTAGTTGGTTAAAAAAAAAAAAAAAAAAAAAAAAAA

C1: Forward sequencing result of 3' RACE PCR product (poly A was detected after the end of 3' region of Unigene36813_All showed in red boxes)

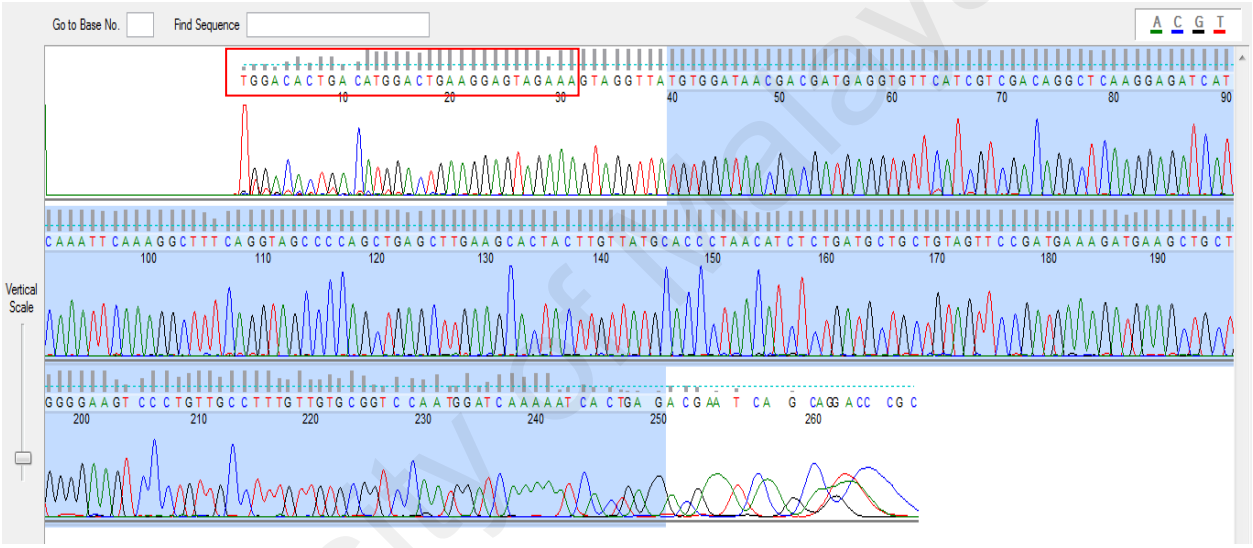
Appendix C2: RACE-PCR method (5' RACE)

4CL -428bp (Unigene36813_All from transcriptome data)

Blue highlight: Overlap region between forward sequence of 5' RACE PCR product and

Unigene36813_All

GAAATCAAGCAGTACATCTCAAGACAGGTGGTTTTCTACAAGAGAATCAACAAAGTTTTCTTCACAGAAA
TCATTCCAAAGGCCCCCTCTGGAAAAATCTTGAGAAAGGATCTAAGAGCGAAGCTAGCTGAGCAATTCCC
CATCGGTCCATTTCCATGATCATACCGACACCGTCGGTATCTTCCCATAGTGCTCTCATGGTGGCAAGAA
CAAGACGTGGTGTAATATTCTCAACATAGTTCATGTTTCCAAGTCATTCTTTTCTGTGCTGTTTATTAAC
GTTGTTAATTGTTGACGGGAGGGCTGGTTGACCCCTCCGACTTCAGCTATAGATTTTATTTTACTTGTTC
TACACTGTGTTTGTAATGTTAATGTGGATAGAGATGTGACTTTGCTAAATGAAAAGGTTGTTATTAGTTG
GTTTCTTA



TGGACACTGACATGGACTGAAGGAGTAGAAAGTAGGTTATGTGGATAACGACGATGAGGTGTTCATCGTC
GACAGGCTCAAGGAGATCATCAAATTCAAAGGCTTTCAGGTAGCCCCAGCTGAGCTTGAAGCACTACTTG
TTATGCACCCTAACATCTCTGATGCTGCTGTAGTTCCGATGAAAGATGAAGCTGCTGGGGAAGTCCCTGT
TGCCTTTGTTGTGCGGTCCAATGGATCAAAAATCACTGAGACGAATCAGCAGGACCCGC

Red highlight: Overlap region between forward and reverse sequence of 5'RACE PCR product

C2: Reverse complement of reverse sequencing result of 5'RACE PCR product (Primer adapter for 5' RACE was detected on the sequences showed in red box)

Appendix C3: Recovered sequences from 5' and 3' RACE

5'TGGACACTGACATGGACTGAAGGAGTAGAAAGTAGGTTATGTGGATAAC
GACGATGAGGTGTTCATCGTCGACAGGCTCAAGGAGATCATCAAATTCAAA
GGCTTTCAGGTAGCCCCAGCTGAGCTTGAAGCACTACTTGTTATGCACCCTA
ACATCTCTGATGCTGCTGTAGTTCCGATGAAAGATGAAGCTGCTGGGGAAG
TCCCTGTTGCCTTTGTTGTGCGGTCCAATGGATCAAAAATCACTGAGGACGA
AATCAAGCAGTACATCTCAAGACAGGTGGTTTTCTACAAGAGAATCAACAA
AGTTTTCTTCACAGAAATCATTCCAAAGGCCCCCTCTGGAAAAATCTTGAGA
AAGGATCTAAGAGCGAAGCTAGCTGAGCAATTCCCCATCGGTCCATTTCCA
TGATCATACCGACACCGTCGGTATCTTCCCATAGTGCTCTCATGGTGGCAAG
AACAAGACGTGGTGTAATATTCTCAACATAGTTCATGTTTCCAAGTCATTCT
TTTCTGTGCTGTTTATTAACGTTGTTAATTGTTGACGGGAGGGCTGGTTGAC
CCCTCCGACTTCAGCTATAGATTTTATTTTACTTGTTCTACACTGTGTTTGTA
ATGTTAATGTGGATAGAGATGTGACTTTGCTAAATGAAAAGGTTGTTATTAG
TTGGTTTCTTA3'

Blue- adapter of 5' gene nested RACE

Red- overlapping region between forward and reverse sequence of 5'RACE PCR product

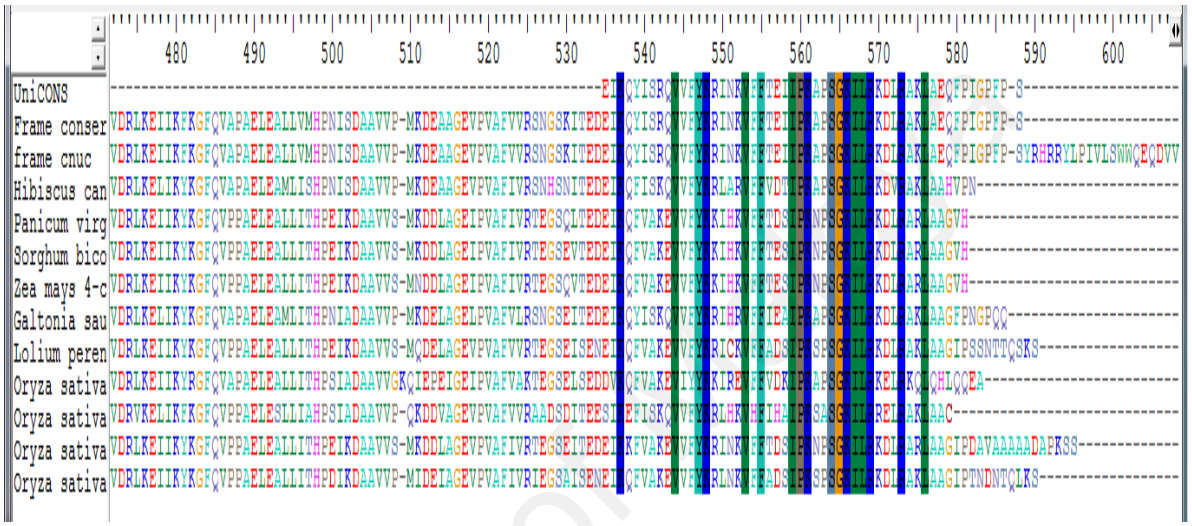
Green- overlapping region between the ends of 5' RACE forward sequence result and conserved region of Unigene36813_All

Purple- overlapping region between the begin of 3' RACE forward sequence result and conserved region of Unigene36813_All

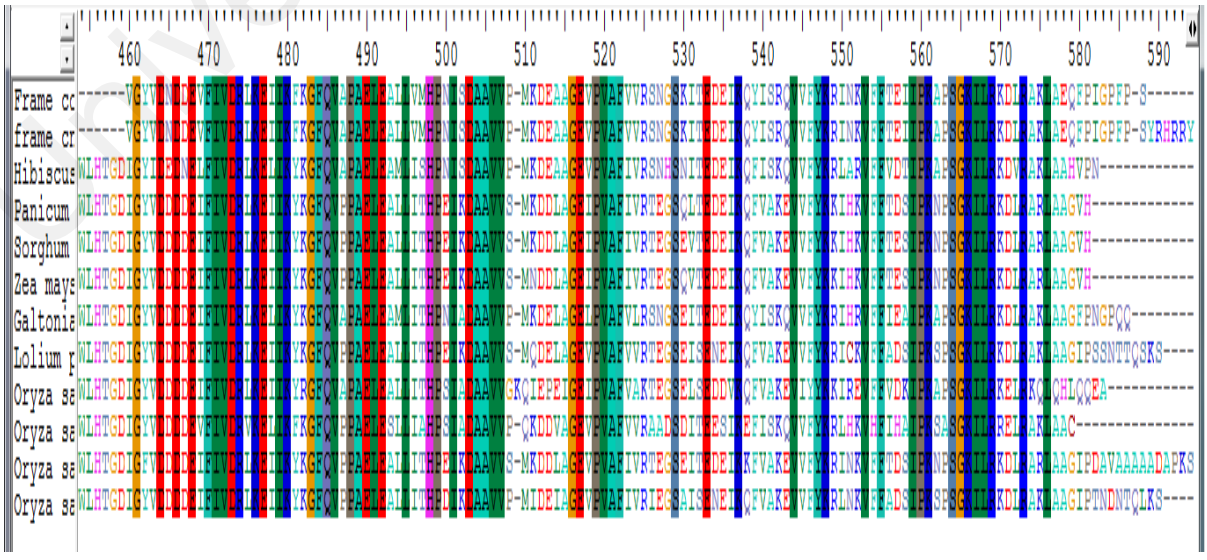
Orange- overlapping region between 3' RACE forward sequence result and unconserved region of Unigene36813_All

Appendix C4: Alignment of amino acid of Unigene36813_All, RACE-PCR product, and 10 of 4CL amino acids from other organism.

(1) Amino acids alignment of Unigene36813_All before RACE experiment with 10 others complete coding 4CL genes namely [*Zea mays*] (NM_001111788.1) [*Panicum virgatum*] (ADZ96250.1) [*Sorghum bicolour*] (XP_002451647.1) [*Lolium perenne*] (AAF37732.1) [*Hibiscus cannabinus*] (ADK24217.1) [*Galtonia saundersiae*] (KF241990.1 and 4 isoform of 4CL from *Oryza sativa* (XP_015650724.1) , (XP_015624111.1), (XP_015625716.1) and (XP_015643415.1)

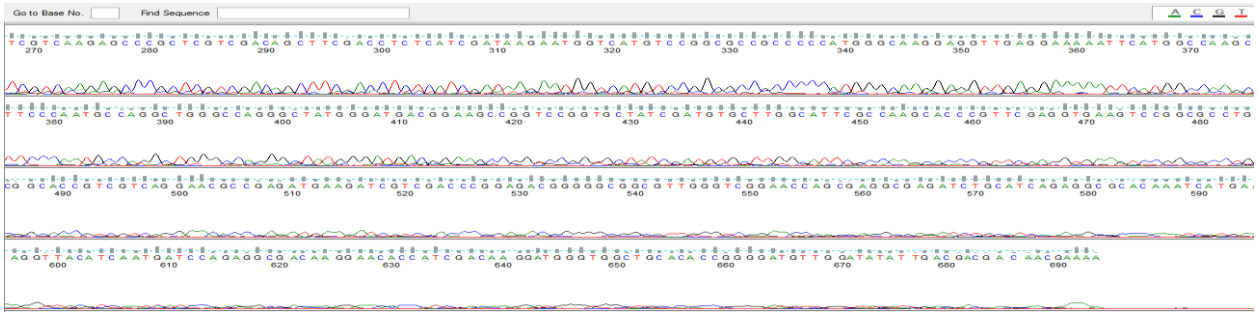


(2) Amino acids alignment of sequence recovered after 5' and 3' RACE with 10 others complete coding 4CL genes namely [*Zea mays*] (NM_001111788.1) [*Panicum virgatum*] (ADZ96250.1) [*Sorghum bicolour*] (XP_002451647.1) [*Lolium perenne*] (AAF37732.1) [*Hibiscus cannabinus*] (ADK24217.1) [*Galtonia saundersiae*] (KF241990.1) and 4 isoform of 4CL from *Oryza sativa* (XP_015650724.1) , (XP_015624111.1), (XP_015625716.1) and (XP_015643415.1)



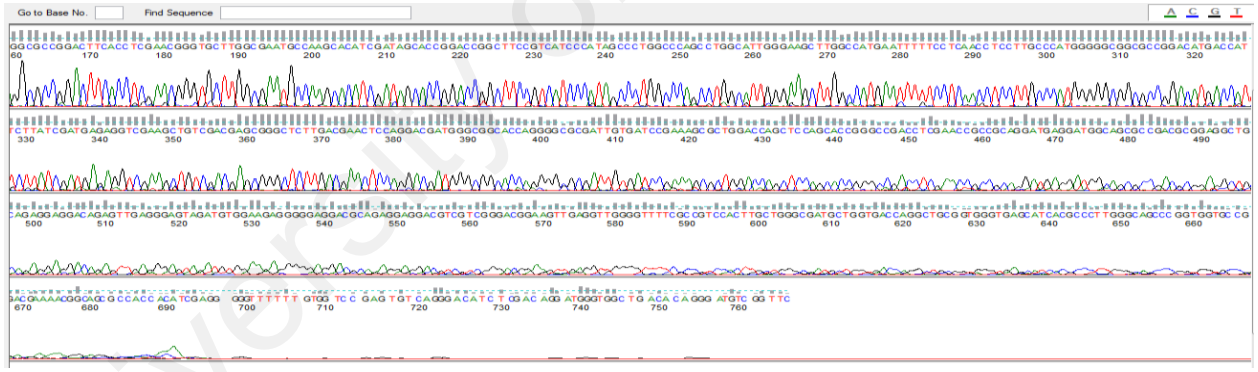
Appendix C5: Sequence (Degenerate primers-PCR method)

(1) Forward sequencing result



AGTCAACCCTCTCGCGTCGGCCTGGTCAGCAGCATCGCCCGGGAAGTGGACGGCGAAAGCCCCAGCCTCG
CTTCCGTCCCGACGACGTCCTCCTCTGCGTCCTCCCCCTCTTCCACATCTACTCCCTCAACTCTGTCCTC
CTCTGCAGCCTCCGCGTCGGCGCTGCCATCCTCATCCTGCGGCGGTTCGAGGTCGGCCCGGTGCTGGAGC
TGGTCCAGCGCTTTCGGATCACAATCGCCCCCCTGGTGCCGCCCATCGTCCTGGAGTTCGTCAAGAGCCC
GCTCGTCGACAGCTTCGACCTCTCATCGATAAGAATGGTCATGTCCGGCGCCGCCCCCATGGGCAAGGAG
GTTGAGGAAAAATTCATGGCCAAGCTTCCCAATGCCAGGCTGGGCCAGGGCTATGGGATGACGGAAGCCG
GTCCGGTGCTATCGATGTGCTTGGCATTCGCCAAGCACCCGTTCGAGGTGAAGTCCGGCGCCTGCGGCAC
CGTCGTCAGGAACGCCGAGATGAAGATCGTCGACCCGGAGACGGGGGCGGCGTTGGGTCGGAACCAGCGA
GGCGAGATCTGCATCAGAGGCGCACAAATCATGAAAGGTTACATCAATGATCCAGAGGCGACAAGGAACA
CCATCGACAAGGATGGGTGGCTGCACACCGGGGATGTTGGATATATTGACGACGACAACGAAAA

(2) Reverse sequencing result



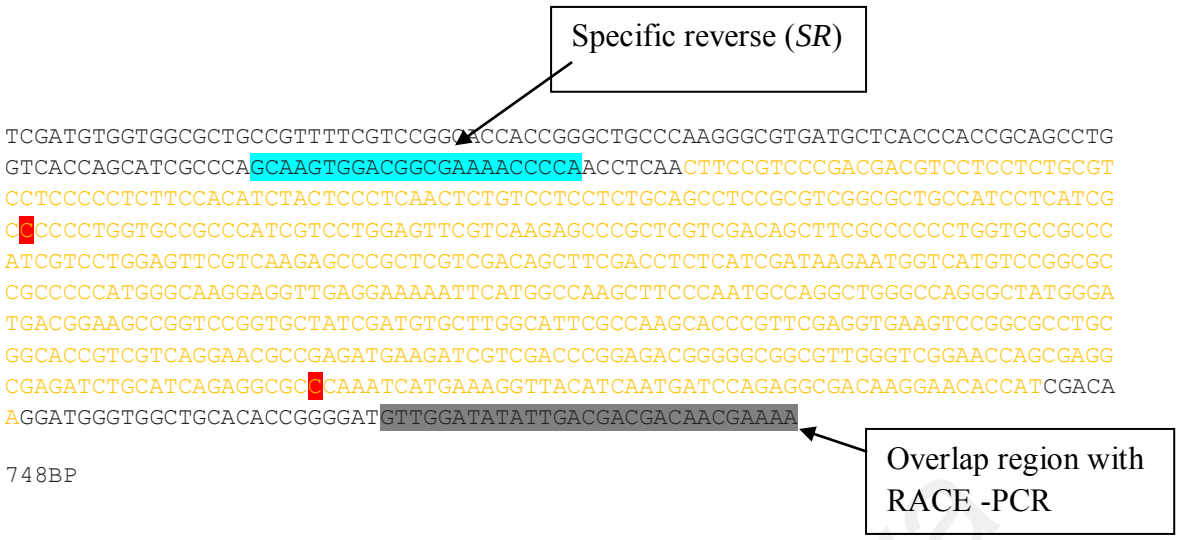
TTCCCGAATGGTGTTCCTTGTCGCCTCTGGATCATTGATGTAACCTTTCATGATTTGGGCGCCTCTGATG
CAGATCTCGCCTCGCTGGTTCCGACCCAACGCCGCCCCCGTCTCCGGGTCGACGATCTTCATCTCGGCGT
TCCTGACGACGGTGCCGCAGGCGCCGGACTTCACCTCGAACGGGTGCTTGGCGAATGCCAAGCACATCGA
TAGCACCGGACCGGCTTCCGTCATCCCATAGCCCTGGCCCAGCCTGGCATTGGGAAGCTTGGCCATGAAT
TTTTCCTCAACCTCCTTGCCCATGGGGGCGGCGCCGGACATGACCATTCTTATCGATGAGAGGTCGAAGC
TGTCGACGAGCGGGCTCTTGACGAACTCCAGGACGATGGGCGGCACCAGGGGCGCGATTGTGATCCGAAA
GCGCTGGACCAGCTCCAGCACCGGGCCGACCTCGAACCGCCGCAGGATGAGGATGGCAGCGCCGACGCGG
AGGCTGCAGAGGAGGACAGAGTTGAGGGAGTAGATGTGGAAGAGGGGGAGGACGCAGAGGAGGACGTCGT
CGGGACGGAAGTTGAGGTTGGGGTTTTCGCCGTCCACTTGCTGGGCGATGCTGGTGACCAGGCTGCGGTG
GGTGAGCATCACGCCCTTGGGCAGCCCGGTGGTGCCGGACGAAAACGGCAGCGCCACCACATCGAGGGGG
TTTTTTGTGGTCCGAGTGTCAGGGACATCTCGACAGGATGGGTGGCTGACACAGGGATGTCGGTTC

Appendix C6: Overlap nucleotides region (Degenerate Primers-PCR method)

(1) CLUSTAL O (1.2.1) - Sequence alignment of forward and reverse complement of reverse sequencing result

```
1       ------------------------------------------------------------
2       GAACCGACATCCCTGTGTCAGCCACCCATCCTGTCGAGATGTCCCTGACACTCGGACCAC


1       ------------------------------------------------------------
2       AAAAAACCCCCTCGATGTGGTGGCGCTGCCGTTTTCGTCCGGCACCACCGGGCTGCCCAA


1       ----AGTCAACCCTCTCGCGTCGGCCTGGTCAGCAGCATCGCCCGGGAAGTGGACGGCGA
2       GGGCGTGATGCTCACCCACCGCAGCCTGGTCACCAGCATCGCCCAGCAAGTGGACGGCGA
                  * * * * *  * ********* *********** * *************

1       AAGCC-CCAGCCTCGCTTCCGTCCCGACGACGTCCTCCTCTGCGTCCTCCCCCTCTTCCA
2       AAACCCCAACCTCAACTTCCGTCCCGACGACGTCCTCCTCTGCGTCCTCCCCCTCTTCCA
        ** ** * * *    ******************************************

1       CATCTACTCCCTCAACTCTGTCCTCCTCTGCAGCCTCCGCGTCGGCGCTGCCATCCTCAT
2       CATCTACTCCCTCAACTCTGTCCTCCTCTGCAGCCTCCGCGTCGGCGCTGCCATCCTCAT
        ************************************************************

1       CCTGCGGCGGTTCGAGGTCGGCCCGGTGCTGGAGCTGGTCCAGCGCTTTCGGATCACAAT
2       CCTGCGGCGGTTCGAGGTCGGCCCGGTGCTGGAGCTGGTCCAGCGCTTTCGGATCACAAT
        ************************************************************

1       CGCCCCCCCTGGTGCCGCCCATCGTCCTGGAGTTCGTCAAGAGCCCGCTCGTCGACAGCTT
2       CGCGCCCCCTGGTGCCGCCCATCGTCCTGGAGTTCGTCAAGAGCCCGCTCGTCGACAGCTT
        *** ********************************************************

1       CGACCTCTCATCGATAAGAATGGTCATGTCCGGCGCCGCCCCCATGGGCAAGGAGGTTGA
2       CGACCTCTCATCGATAAGAATGGTCATGTCCGGCGCCGCCCCCATGGGCAAGGAGGTTGA
        ************************************************************

1       GGAAAAATTCATGGCCAAGCTTCCCAATGCCAGGCTGGGCCAGGGCTATGGGATGACGGA
2       GGAAAAATTCATGGCCAAGCTTCCCAATGCCAGGCTGGGCCAGGGCTATGGGATGACGGA
        ************************************************************

1       AGCCGGTCCGGTGCTATCGATGTGCTTGGCATTCGCCAAGCACCCGTTCGAGGTGAAGTC
2       AGCCGGTCCGGTGCTATCGATGTGCTTGGCATTCGCCAAGCACCCGTTCGAGGTGAAGTC
        ************************************************************

1       CGGCGCCTGCGGCACCGTCGTCAGGAACGCCGAGATGAAGATCGTCGACCCGGAGACGGG
2       CGGCGCCTGCGGCACCGTCGTCAGGAACGCCGAGATGAAGATCGTCGACCCGGAGACGGG
        ************************************************************

1       GGCGGCGTTGGGTCGGAACCAGCGAGGCGAGATCTGCATCAGAGGCGCACAAATCATGAA
2       GGCGGCGTTGGGTCGGAACCAGCGAGGCGAGATCTGCATCAGAGGCGCCCAAATCATGAA
        ********************************************** ************

1       AGGTTACATCAATGATCCAGAGGCGACAAGGAACACCATCGACAAGGATGGGTGGCTGCA
2       AGGTTACATCAATGATCCAGAGGCGACAAGGAACACCATTCGGGAA--------------
        ************************************     *

1       CACCGGGGATGTTGGATATATTGACGACGACAACGAAAA
2       ---------------------------------------
```

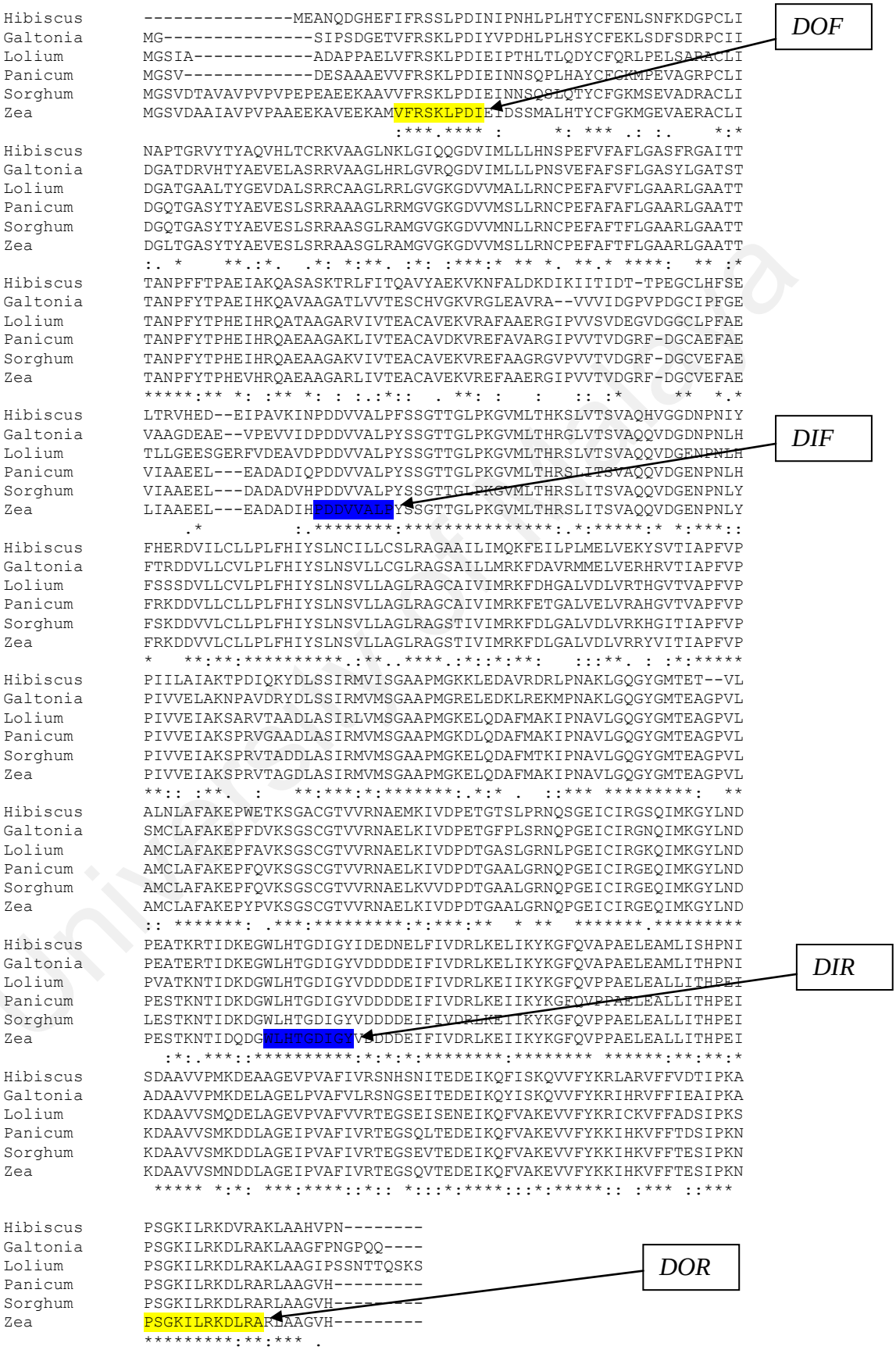
.



The red highlight showed the selected nucleotides based on chromatogram peak intensity

Appendix C7: Multiple alignment of 4CL proteins and position of degenerate primers (*DOF*, *DOR*, *DIF*, *DIR*) (Degenerate primers-PCR method)

(1) CLUSTAL O (1.2.1) multiple sequence alignment of amino acids




```
Hibiscus      ---------------MEANQDGHEFIFRSSLPDINIPNHLPLHTYCFENLSNFKDGPCLI
Galtonia      MG---------------SIPSDGETVFRSKLPDIYVPDHLPLHSYCFEKLSDFSDRPCII
Lolium        MGSIA------------ADAPPAELVFRSKLPDIEIPTHLTLQDYCFQRLPELSARACLI
Panicum       MGSV-------------DESAAAEVVFRSKLPDIEINNSQPLHAYCFGKMPEVAGRPCLI
Sorghum       MGSVDTAVAVPVPVPEPEAEEKAAVVFRSKLPDIEINNSQSLQTYCFGKMSEVADRACLI
Zea           MGSVDAAIAVPVPAAEEKAVEEKAMVFRSKLPDIEIDSSMALHTYCFGKMGEVAERACLI
                                       :***.**** :     *: *** .: :.    *:*
Hibiscus      NAPTGRVYTYAQVHLTCRKVAAGLNKLGIQQGDVIMLLLHNSPEFVFAFLGASFRGAITT
Galtonia      DGATDRVHTYAEVELASRRVAAGLHRLGVRQGDVIMLLLPNSVEFAFSFLGASYLGATST
Lolium        DGATGAALTYGEVDALSRRCAAGLRRLGVGKGDVVMALLRNCPEFAFVFLGAARLGAATT
Panicum       DGQTGASYTYAEVESLSRRAAAGLRRMGVGKGDVVMSLLRNCPEFAFAFLGAARLGAATT
Sorghum       DGQTGASYTYAEVESLSRRAASGLRAMGVGKGDVVMNLLRNCPEFAFTFLGAARLGAATT
Zea           DGLTGASYTYAEVESLSRRAASGLRAMGVGKGDVVMSLLRNCPEFAFTFLGAARLGAATT
              :. *    **.:*.  .*: *:**. :*: :***:* ** *. **.* ****:  ** :*
Hibiscus      TANPFFTPAEIAKQASASKTRLFITQAVYAEKVKNFALDKDIKIITIDT-TPEGCLHFSE
Galtonia      TANPFYTPAEIHKQAVAAGATLVVTESCHVGKVRGLEAVRA--VVVIDGPVPDGCIPFGE
Lolium        TANPFYTPHEIHRQATAAGARVIVTEACAVEKVRAFAAERGIPVVSVDEGVDGGCLPFAE
Panicum       TANPFYTPHEIHRQAEAAGAKLIVTEACAVDKVREFAVARGIPVVTVDGRF-DGCAEFAE
Sorghum       TANPFYTPHEIHRQAEAAGAKVIVTEACAVEKVREFAAGRGVPVVTVDGRF-DGCVEFAE
Zea           TANPFYTPHEVHRQAEAAGARLIVTEACAVEKVREFAAERGIPVVTVDGRF-DGCVEFAE
              *****:** *: :** *: : :.:*::  . **: :   :   :: :*      **  *.*
Hibiscus      LTRVHED--EIPAVKINPDDVVALPFSSGTTGLPKGVMLTHKSLVTSVAQHVGGDNPNIY
Galtonia      VAAGDEAE--VPEVVIDPDDVVALPYSSGTTGLPKGVMLTHRGLVTSVAQQVDGDNPNLH
Lolium        TLLGEESGERFVDEAVDPDDVVALPYSSGTTGLPKGVMLTHRSLVTSVAQQVDGENPNLH
Panicum       VIAAEEL---EADADIQPDDVVALPYSSGTTGLPKGVMLTHRSLITSVAQQVDGENPNLH
Sorghum       VIAAEEL---DADADVHPDDVVALPYSSGTTGLPKGVMLTHRSLITSVAQQVDGENPNLY
Zea           LIAAEEL---EADADIHPDDVVALPYSSGTTGLPKGVMLTHRSLITSVAQQVDGENPNLY
                  .*        :.********:****************:.*:*****:* *:***::
Hibiscus      FHERDVILCLLPLFHIYSLNCILLCSLRAGAAILIMQKFEILPLMELVEKYSVTIAPFVP
Galtonia      FTRDDVLLCVLPLFHIYSLNSVLLCGLRAGSAILLMRKFDAVRMMELVERHRVTIAPFVP
Lolium        FSSSDVLLCVLPLFHIYSLNSVLLAGLRAGCAIVIMRKFDHGALVDLVRTHGVTVAPFVP
Panicum       FRKDDVLLCLLPLFHIYSLNSVLLAGLRAGCAIVIMRKFETGALVELVRAHGVTVAPFVP
Sorghum       FSKDDVVLCLLPLFHIYSLNSVLLAGLRAGSTIVIMRKFDLGALVDLVRKHGITIAPFVP
Zea           FRKDDVVLCLLPLFHIYSLNSVLLAGLRAGSTIVIMRKFDLGALVDLVRRYVITIAPFVP
              *   **:**:**********.:**..****.:*::*:**:   :::**. : :*:*****
Hibiscus      PIILAIAKTPDIQKYDLSSIRMVISGAAPMGKKLEDAVRDRLPNAKLGQGYGMTET--VL
Galtonia      PIVVELAKNPAVDRYDLSSIRMVMSGAAPMGRELEDKLREKMPNAKLGQGYGMTEAGPVL
Lolium        PIVVEIAKSARVTAADLASIRLVMSGAAPMGKELQDAFMAKIPNAVLGQGYGMTEAGPVL
Panicum       PIVVEIAKSPRVGAADLASIRMVMSGAAPMGKDLQDAFMAKIPNAVLGQGYGMTEAGPVL
Sorghum       PIVVEIAKSPRVTADDLASIRMVMSGAAPMGKELQDAFMTKIPNAVLGQGYGMTEAGPVL
Zea           PIVVEIAKSPRVTAGDLASIRMVMSGAAPMGKELQDAFMAKIPNAVLGQGYGMTEAGPVL
              **:: :**.  :   **:***:*:*******:.*:* .  ::*** *********:  **
Hibiscus      ALNLAFAKEPWETKSGACGTVVRNAEMKIVDPETGTSLPRNQSGEICIRGSQIMKGYLND
Galtonia      SMCLAFAKEPFDVKSGSCGTVVRNAELKIVDPETGFPLSRNQPGEICIRGNQIMKGYLND
Lolium        AMCLAFAKEPFAVKSGSCGTVVRNAELKIVDPDTGASLGRNLPGEICIRGKQIMKGYLND
Panicum       AMCLAFAKEPFQVKSGSCGTVVRNAELKIVDPDTGAALGRNQPGEICIRGEQIMKGYLND
Sorghum       AMCLAFAKEPFQVKSGSCGTVVRNAELKVVDPDTGAALGRNQPGEICIRGEQIMKGYLND
Zea           AMCLAFAKEPYPVKSGSCGTVVRNAELKIVDPDTGAALGRNQPGEICIRGEQIMKGYLND
              :: *******: .***:**********:*:***:**  * **  *******.*********
Hibiscus      PEATKRTIDKEGWLHTGDIGYIDEDNELFIVDRLKELIKYKGFQVAPAELEAMLISHPNI
Galtonia      PEATERTIDKEGWLHTGDIGYVDDDDEIFIVDRLKELIKYKGFQVAPAELEAMLITHPNI
Lolium        PVATKNTIDKDGWLHTGDIGYVDDDDEIFIVDRLKEIIKYKGFQVPPAELEALLITHPEI
Panicum       PESTKNTIDKDGWLHTGDIGYVDDDDEIFIVDRLKEIIKYKGFQVPPAELEALLITHPEI
Sorghum       LESTKNTIDKDGWLHTGDIGYVDDDDEIFIVDRLKEIIKYKGFQVPPAELEALLITHPEI
Zea           PESTKNTIDQDGWLHTGDIGYVDDDEIFIVDRLKEIIKYKGFQVPPAELEALLITHPEI
                :*:.***::**********:*:*:*:********:******** ******:**:**:*
Hibiscus      SDAAVVPMKDEAAGEVPVAFIVRSNHSNITEDEIKQFISKQVVFYKRLARVFFVDTIPKA
Galtonia      ADAAVVPMKDELAGELPVAFVLRSNGSEITEDEIKQYISKQVVFYKRIHRVFFIEAIPKA
Lolium        KDAAVVSMQDELAGEVPVAFVVRTEGSEISENEIKQFVAKEVVFYKRICKVFFADSIPKS
Panicum       KDAAVVSMKDDLAGEIPVAFIVRTEGSQLTEDEIKQFVAKEVVFYKKIHKVFFTDSIPKN
Sorghum       KDAAVVSMKDDLAGEIPVAFIVRTEGSEVTEDEIKQFVAKEVVFYKKIHKVFFTESIPKN
Zea           KDAAVVSMNDDLAGEIPVAFIVRTEGSQVTEDEIKQFVAKEVVFYKKIHKVFFTESIPKN
               ***** *:*: ***:****::*:: *:::*:****:::*:*****:: :*** ::***

Hibiscus      PSGKILRKDVRAKLAAHVPN--------
Galtonia      PSGKILRKDLRAKLAAGFPNGPQQ----
Lolium        PSGKILRKDLRAKLAAGIPSSNTTQSKS
Panicum       PSGKILRKDLRARLAAGVH---------
Sorghum       PSGKILRKDLRARLAAGVH---------
Zea           PSGKILRKDLRARLAAGVH---------
              *********:**:*** .
```

Yellow highlight are amino acid regions for degenerate outer primers pair (*DOF*, *DOR*); Blue highlight are amino acid regions for degenerate inner primers pair (*DIF*, *DIR*)

Appendix C8: Sequence position of *DOF* and *SR* primers (Amplification of 5' region of *Br4CL*)

```
Hibiscus      ------------------------------------------ATGGAGGCCAACCAA
Galtonia      ------------------------------------------ATGGGCTCCATCCCG
Lolium        ---------------------------------ATGGGCTCCATCGCGGCGGACGCG
Panicum       ------------------------------------ATGGGTTCCGTGGACGAGTCG
Sorghum       ATGGGTTCCGTGGACACGGCGGTCGCGGTGCCGGTGCCAGTGCCGGAGCCGGAGGCAGAG
Zea           ATGGGTTCCGTAGACGCGGCGATCGCGGTGCCGGTGCCGGCGGCGGAGGAGAAGGCGGTG

Hibiscus      GATGGGCATGAATTCATCTTTCGGTCCAGTCTCCCAGACATAAACATCCCAAACCACCTC
Galtonia      TCGGACGGAGAGACGGTCTTCCGGTCGAAGCTGCCGGACATCTACGTCCCAGACCATCTC
Lolium        CCTCCCGCGGAGCTGGTGTTCCGGTCCAAGCTCCCGGACATCGAGATCCCGAGCCACCTG
Panicum       GCGGCGGCGGAGGTGGTGTTCCGGTCGAAGCTGCCGGACATCGAGATCAACAACAGCCAG
Sorghum       GAGAAGGCGGCGGTGGTGTTCCGTTCCAAGCTTCCCGACATCGAGATCAACAACAGCCAG
Zea           GAGGAGAAGGCGATGGTGTTCCGGTCCAAGCTTCCCGACATCGAGATCGACAGCAGCATG
                       *      * ** ** ** *  ** ** *****  * **     *

Hibiscus      CCTTTGCATACTTACTGTTTTGAGAACCTTTCCAACTTTAAAGATGGTCCTTGCTTAATC
Galtonia      CCCCTCCACTCATACTGCTTCGAGAAGCTCTCCGACTTCTCCGACCGCCCCTGCATCATC
Lolium        ACGCTGCAGGACTACTGCTTCCAGCGCCTGCCGGAGCTCTCCGCGCGCGCCTGCCTCATC
Panicum       CCGCTGCACGCGTACTGCTTCGGCAAGATGCCCGAGGTGGCGGGGCGGCCCTGCCTGATC
Sorghum       TCGCTGCAGACCTACTGCTTCGGGAAGATGAGCGAGGTGGCGGACCGCGCCTGCCTCATC
Zea           GCGCTGCACACCTACTGCTTCGGGAAGATGGGCGAGGTGGCGGAGCGGGCGTGCCTGATC
               *  * **    ***** **        *     *  *    *   *  * *** * ***
```

DOF

- Finalized 4CL sequences recovered from cDNA source of *B. rotunda* using outer and inner degenerate primers from section 4.3.2

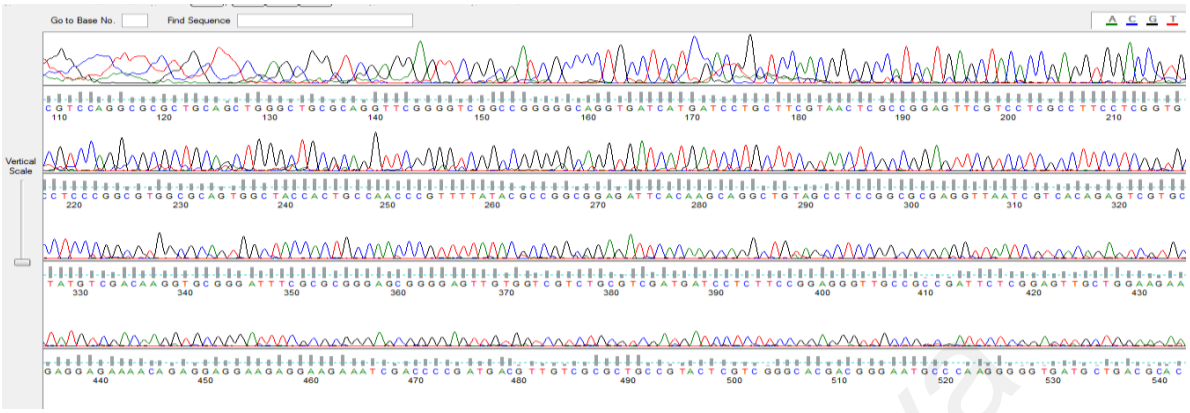
```
TCGATGTGGTGGCGCTGCCGTTTTCGTCCGGCACCACCGGGCTGCCCAAGGGCGTGATGCTCACCCACCG
CAGCCTGGTCACCAGCATCGCCCAGCAAGTGGACGGCGAAAACCCCAACCTCAACTTCCGTCCCGACGAC
GTCCTCCTCTGCGTCCTCCCCCTCTTCCACATCTACTCCCTCAACTCTGTCCTCCTCTGCAGCCTCCGCG
TCGGCGCTGCCATCCTCATCCTGCGGCGGTTCGAGGTCGGCCCGGTGCTGGAGCTGGTCCAGCGCTTTCG
GATCACAATCGCCCCCCTGGTGCCGCCCATCGTCCTGGAGTTCGTCAAGAGCCCGCTCGTCGACAGCTTC
GACCTCTCATCGATAAGAATGGTCATGTCCGGCGCCGCCCCCATGGGCAAGGAGGTTGAGGAAAAATTCA
TGGCCAAGCTTCCCAATGCCAGGCTGGGCCAGGGCTATGGGATGACGGAAGCCGGTCCGGTGCTATCGAT
GTGCTTGGCATTCGCCAAGCACCCGTTCGAGGTGAAGTCCGGCGCCTGCGGCACCGTCGTCAGGAACGCC
GAGATGAAGATCGTCGACCCGGAGACGGGGGCGGCGTTGGGTCGGAACCAGCGAGGCGAGATCTGCATCA
GAGGCGCACAAATCATGAAAGGTTACATCAATGATCCAGAGGCGACAAGGAACACCATCGACAAGGATGG
GTGGCTGCACACCGGGGATGTTGGATATATTGACGACGACAACGAAAA
```

748BP

Specific reverse primer (*SR*) was designed from recovered sequence from section 4.3.2

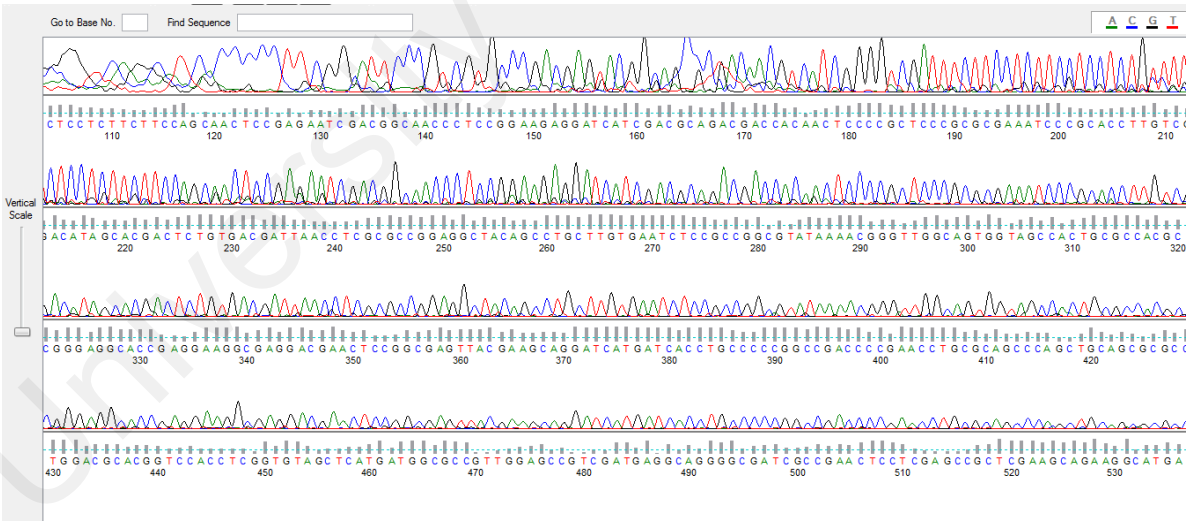
Appendix C9: Amplification of 5' region of *Br4CL*

(1) Forward sequencing result



ATGCGCCTTTTCCTTGCCTTTCGAGCGAGCTCGAGGAGTTCGGCGATCGCCCCTGCCTCATCGACGGCTC
CAACGGCGCCATCATGAGCTACACCGAGGTGGACCGTGCGTCCAGGCGCGCTGCAGCTGGGCTGCGCAGG
TTCGGGGTCGGCCGGGGGCAGGTGATCATGATCCTGCTTCGTAACTCGCCGGAGTTCGTCCTCGCCTTCC
TCGGTGCCTCCCGGCGTGGCGCAGTGGCTACCACTGCCAACCCGTTTTATACGCCGGCGGAGATTCACAA
GCAGGCTGTAGCCTCCGGCGCGAGGTTAATCGTCACAGAGTCGTGCTATGTCGACAAGGTGCGGGATTTC
GCGCGGGAGCGGGGAGTTGTGGTCGTCTGCGTCGATGATCCTCTTCCGGAGGGTTGCCGCCGATTCTCGG
AGTTGCTGGAAGAAGAGGAGAAAACAGAGGAGGAAGAGGAAGAAATCGACCCCGATGACGTTGTCGCGCT
GCCGTACTCGTCGGGCACGACGGGAATGCCCAAGGGGGTGATGCTGACGCACCGGATCCTGATCACCAGC
G

(2) Reverse sequencing result



GGGGTTCAGGCATCACCCCCTTCGGCATTCCCGTCGTGCCCGACGAGTACGGCAGCCCGACAACGTCATC
GGGGTCGATTTCTTCCTCTTCCTCCTCTGTTTTCTCCTCTTCTTCCAGCAACTCCGAGAATCGACGGCAA
CCCTCCGGAAGAGGATCATCGACGCAGACGACCACAACTCCCCGCTCCCGCGCGAAATCCCGCACCTTGT
CGACATAGCACGACTCTGTGACGATTAACCTCGCGCCGGAGGCTACAGCCTGCTTGTGAATCTCCGCCGG
CGTATAAAACGGGTTGGCAGTGGTAGCCACTGCGCCACGCCGGGAGGCACCGAGGAAGGCGAGGACGAAC
TCCGGCGAGTTACGAAGCAGGATCATGATCACCTGCCCCCGGCCGACCCCGAACCTGCGCAGCCCAGCTG
CAGCGCGCCTGGACGCACGGTCCACCTCGGTGTAGCTCATGATGGCGCCGTTGGAGCCGTCGATGAGGCA
GGGGCGATCGCCGAACTCCTCGAGCCGCTCGAAGCAGAAGGCATGAAGCGGCCGGCGATTGTCGATGTGG
ATGTCTGGAAGATTCGACCGAAACAAGGG

Appendix C10: Overlap nucleotide sequences (Amplification of 5' region of *Br4CL*)

(1) CLUSTAL O (1.2.1) - Sequence alignment of forward and reverse complement of reverse sequence

```
1       ---------------------------------------------ATGCGCCTTTTCCT
2       CCCTTGTTTCGGTCGAATCTTCCAGACATCCACATCGACAATCGCCGGCCGCTTCATGCC
                                                      *** *  * *

1       TGCCTTTCGAGCGAGCTCGAGGAGTTCGGCGATCGCCCCTGCCTCATCGACGGCTCCAAC
2       TTCTGCTTCGAGCGGCTCGAGGAGTTCGGCGATCGCCCCTGCCTCATCGACGGCTCCAAC
        * *   *       **********************************************

1       GGCGCCATCATGAGCTACACCGAGGTGGACCGTGCGTCCAGGCGCGCTGCAGCTGGGCTG
2       GGCGCCATCATGAGCTACACCGAGGTGGACCGTGCGTCCAGGCGCGCTGCAGCTGGGCTG
        ************************************************************

1       CGCAGGTTCGGGGTCGGCCGGGGGCAGGTGATCATGATCCTGCTTCGTAACTCGCCGGAG
2       CGCAGGTTCGGGGTCGGCCGGGGGCAGGTGATCATGATCCTGCTTCGTAACTCGCCGGAG
        ************************************************************

1       TTCGTCCTCGCCTTCCTCGGTGCCTCCCGGCGTGGCGCAGTGGCTACCACTGCCAACCCG
2       TTCGTCCTCGCCTTCCTCGGTGCCTCCCGGCGTGGCGCAGTGGCTACCACTGCCAACCCG
        ************************************************************

1       TTTTATACGCCGGCGGAGATTCACAAGCAGGCTGTAGCCTCCGGCGCGAGGTTAATCGTC
2       TTTTATACGCCGGCGGAGATTCACAAGCAGGCTGTAGCCTCCGGCGCGAGGTTAATCGTC
        ************************************************************

1       ACAGAGTCGTGCTATGTCGACAAGGTGCGGGATTTCGCGCGGGAGCGGGGAGTTGTGGTC
2       ACAGAGTCGTGCTATGTCGACAAGGTGCGGGATTTCGCGCGGGAGCGGGGAGTTGTGGTC
        ************************************************************

1       GTCTGCGTCGATGATCCTCTTCCGGAGGGTTGCCGCCGATTCTCGGAGTTGCTGGAAGAA
2       GTCTGCGTCGATGATCCTCTTCCGGAGGGTTGCCGTCGATTCTCGGAGTTGCTGGAAGAA
        ********************************** ************************

1       GAGGAGAAAACAGAGGAGGAAGAGGAAGAAATCGACCCCGATGACGTTGTCGCGCTGCCG
2       GAGGAGAAAACAGAGGAGGAAGAGGAAGAAATCGACCCCGATGACGTTGTCGGGCTGCCG
        ************************************************** *******

1       TACTCGTCGGGCACGACGGGAATGCCCAAGGGGGTGATGCTGACGCACCGGATCCTGATC
2       TACTCGTCGGGCACGACGGGAATGCCGAAGGGGGTGATGCCTGAACCCC-----------
        ************************* *************     * **

1       ACCAGCG
2       -------
```

```
ATGCGCCTTTTCCTTGCCTTTCGAGCGAGCTCGAGGAGTTCGGCGATCGCCCCTGCCTCATCGACGGCTCCAACGGCG
CCATCATGAGCTACACCGAGGTGGACCGTGCGTCCAGGCGCGCTGCAGCTGGGCTGCGCAGGTTCGGGGTCGGCCGGG
GGCAGGTGATCATGATCCTGCTTCGTAACTCGCCGGAGTTCGTCCTCGCCTTCCTCGGTGCCTCCCGGCGTGGCGCAG
TGGCTACCACTGCCAACCCGTTTTATACGCCGGCGGAGATTCACAAGCAGGCTGTAGCCTCCGGCGCGAGGTTAATCG
TCACAGAGTCGTGCTATGTCGACAAGGTGCGGGATTTCGCGCGGGAGCGGGGAGTTGTGGTCGTCTGCGTCGATGATC
CTCTTCCGGAGGGTTGCCGTCGATTCTCGGAGTTGCTGGAAGAAGAGGAGAAAACAGAGGAGGAAGAGGAAGAAATCG
ACCCCGATGACGTTGTCGCGCTGCCGTACTCGTCGGGCACGACGGGAATGCCGAAGGGGGTGATGCTGACGCACCGGA
TCCTGATC ACCAGCG
```

ATG codon was detected (green highlight). The red highlight showed the selected nucleotides based on chromatogram peak intensity.

Appendix C11: Summary for 3 PCR products sequences

Section 4.3.3 Amplification of 5' region of *Br4CL*

ATGCGCCTTTTCCTTGCCTTTCGAGCGAGCTCGAGGAGTTCGGCGATCGCCCCTGCCTCATCG
ACGGCTCCAACGGCGCCATCATGAGCTACACCGAGGTGGACCGTGCGTCCAGGCGCGCTGC
AGCTGGGCTGCGCAGGTTCGGGGTCGGCCGGGGGCAGGTGATCATGATCCTGCTTCGTAACT
CGCCGGAGTTCGTCCTCGCCTTCCTCGGTGCCTCCCGGCGTGGCGCAGTGGCTACCACTGCC
AACCCGTTTTATACGCCGGCGGAGATTCACAAGCAGGCTGTAGCCTCCGGCGCGAGGTTAAT
CGTCACAGAGTCGTGCTATGTCGACAAGGTGCGGGATTTCGCGCGGGAGCGGGGAGTTGTG
GTCGTCTGCGTCGATGATCCTCTTCCGGAGGGTTGCCGTCGATTCTCGGAGTTGCTGGAAGA
AGAGGAGAAAACAGAGGAGGAAGAGGAAGAAATCGACCCCGATGACGTTGTCGCGCTGCC
GTACTCGTCGGGCACGACGGGAATGCCGAAGGGGGTGATGCTGACGCACCGGATCCTGATC
ACCAGCG

Section 4.3.2 Degenerate Primers- PCR method

TCGATGTGGTGGCGCTGCCGTTTTCGTCCGGCACCACCGGGCTGCCCAAGGGCGTGATGCTC
ACCCACCGCAGCCTGGTCACCAGCATCGCCCAGCAAGTGGACGGCGAAAACCCCAACCTCA
ACTTCCGTCCCGACGACGTCCTCCTCTGCGTCCTCCCCCTCTTCCACATCTACTCCCTCAACTC
TGTCCTCCTCTGCAGCCTCCGCGTCGGCGCTGCCATCCTCATCCTGCGGCGGTTCGAGGTCGG
CCCGGTGCTGGAGCTGGTCCAGCGCTTTCGGATCACAATCGCCCCCCTGGTGCCGCCCATCG
TCCTGGAGTTCGTCAAGAGCCCGCTCGTCGACAGCTTCGACCTCTCATCGATAAGAATGGTC
ATGTCCGGCGCCGCCCCCATGGGCAAGGAGGTTGAGGAAAAATTCATGGCCAAGCTTCCCA
ATGCCAGGCTGGGCCAGGGCTATGGGATGACGGAAGCCGGTCCGGTGCTATCGATGTGCTT
GGCATTCGCCAAGCACCCGTTCGAGGTGAAGTCCGGCGCCTGCGGCACCGTCGTCAGGAAC
GCCGAGATGAAGATCGTCGACCCGGAGACGGGGGCGGCGTTGGGTCGGAACCAGCGAGGC
GAGATCTGCATCAGAGGCGCCCAAATCATGAAAGGTTACATCAATGATCCAGAGGCGACAA
GGAACACCATCGACAAGGATGGGTGGCTGCACACCGGGGATGTTGGATATATTGACGACGA
CAACGAAAA

Section 4.3.1 RACE-PCR method

GTAGGTTATGTGGATAACGACGATGAGGTGTTCATCGTCGACAGGCTCAAGGAGATCATCA
AATTCAAAGGCTTTCAGGTAGCCCCAGCTGAGCTTGAAGCACTACTTGTTATGCACCCTAAC
ATCTCTGATGCTGCTGTAGTTCCGATGAAAGATGAAGCTGCTGGGGAAGTCCCTGTTGCCTT
TGTTGTGCGGTCCAATGGATCAAAAATCACTGAGGACGAAATCAAGCAGTACATCTCAAGA
CAGGTGGTTTTCTACAAGAGAATCAACAAAGTTTTCTTCACAGAAATCATTCCAAAGGCCCC
CTCTGGAAAAATCTTGAGAAAGGATCTAAGAGCGAAGCTAGCTGAGCAATTCCCCATCGGT
CCATTTCCATGATCATACCGACACCGTCGGTATCTTCCCATAGTGCTCTCATGGTGGCAAGA
ACAAGACGTGGTGTAATATTCTCAACATAGTTCATGTTTCCAAGTCATTCTTTTCTGTGCTGT
TTATTAACGTTGTTAATTGTTGACGGGAGGGCTGGTTGACCCCTCCGACTTCAGCTATAGATT
TTATTTTACTTGTTCTACACTGTGTTTGTAATGTTAATGTGGATAGAGATGTGACTTTGCTAA
ATGAAAAGGTTGTTATTAGTTGGTTTCTTA

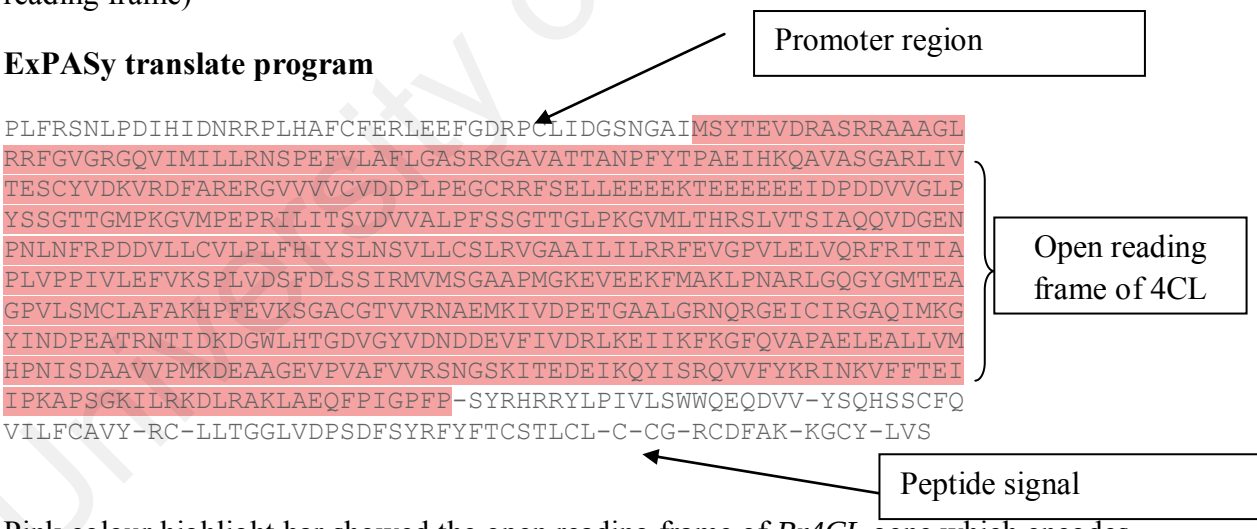
Purple highlight: selected sequences for the synthesized final sequence. Red highlight: Overlapped nucleotides detected between the PCR products from 3 different methods of PCR amplification. Green highlight: start and stop codon.

Appendix C12: Final coding region of *4CL* gene sequence in *B.rotunda* after overlapping sequences from all 3 PCR products

ATGAGCTACACCGAGGTGGACCGTGCGTCCAGGCGCGCTGCAGCTGGGCTGCGCAGGTTCGGGGTCGGCC
GGGGGCAGGTGATCATGATCCTGCTTCGTAACTCGCCGGAGTTCGTCCTCGCCTTCCTCGGTGCCTCCCG
GCGTGGCGCAGTGGCTACCACTGCCAACCCGTTTTATACGCCGGCGGAGATTCACAAGCAGGCTGTAGCC
TCCGGCGCGAGGTTAATCGTCACAGAGTCGTGCTATGTCGACAAGGTGCGGGATTTCGCGCGGGAGCGGG
GAGTTGTGGTCGTCTGCGTCGATGATCCTCTTCCGGAGGGTTGCCGTCGATTCTCGGAGTTGCTGGAAGA
AGAGGAGAAAACAGAGGAGGAAGAGGAAGAAATCGACCCCGATGACGTTGTCGGGCTGCCGTACTCGTCG
GGCACGACGGGAATGCCGAAGGGGGTGATGCCTGAACCCCGGATCCTGATCACCAGCGTCGATGTGGTGG
CGCTGCCGTTTTCGTCCGGCACCACCGGGCTGCCCAAGGGCGTGATGCTCACCCACCGCAGCCTGGTCAC
CAGCATCGCCCAGCAAGTGGACGGCGAAAACCCCAACCTCAACTTCCGTCCCGACGACGTCCTCCTCTGC
GTCCTCCCCCTCTTCCACATCTACTCCCTCAACTCTGTCCTCCTCTGCAGCCTCCGCGTCGGCGCTGCCA
TCCTCATCCTGCGGCGGTTCGAGGTCGGCCCGGTGCTGGAGCTGGTCCAGCGCTTTCGGATCACAATCGC
CCCCCTGGTGCCGCCCATCGTCCTGGAGTTCGTCAAGAGCCCGCTCGTCGACAGCTTCGACCTCTCATCG
ATAAGAATGGTCATGTCCGGCGCCGCCCCCATGGGCAAGGAGGTTGAGGAAAAATTCATGGCCAAGCTTC
CCAATGCCAGGCTGGGCCAGGGCTATGGGATGACGGAAGCCGGTCCGGTGCTATCGATGTGCTTGGCATT
CGCCAAGCACCCGTTCGAGGTGAAGTCCGGCGCCTGCGGCACCGTCGTCAGGAACGCCGAGATGAAGATC
GTCGACCCGGAGACGGGGGCGGCGTTGGGTCGGAACCAGCGAGGCGAGATCTGCATCAGAGGCGCACAAA
TCATGAAAGGTTACATCAATGATCCAGAGGCGACAAGGAACACCATCGACAAGGATGGGTGGCTGCACAC
CGGGGATGTAGGTTATGTGGATAACGACGATGAGGTGTTCATCGTCGACAGGCTCAAGGAGATCATCAAA
TTCAAAGGCTTTCAGGTAGCCCCAGCTGAGCTTGAAGCACTACTTGTTATGCACCCTAACATCTCTGATG
CTGCTGTAGTTCCGATGAAAGATGAAGCTGCTGGGGAAGTCCCTGTTGCCTTTGTTGTGCGGTCCAATGG
ATCAAAAATCACTGAGGACGAAATCAAGCAGTACATCTCAAGACAGGTGGTTTTCTACAAGAGAATCAAC
AAAGTTTTCTTCACAGAAATCATTCCAAAGGCCCCCTCTGGAAAAATCTTGAGAAAGGATCTAAGAGCGA
AGCTAGCTGAGCAATTCCCCATCGGTCCATTTCCATGA

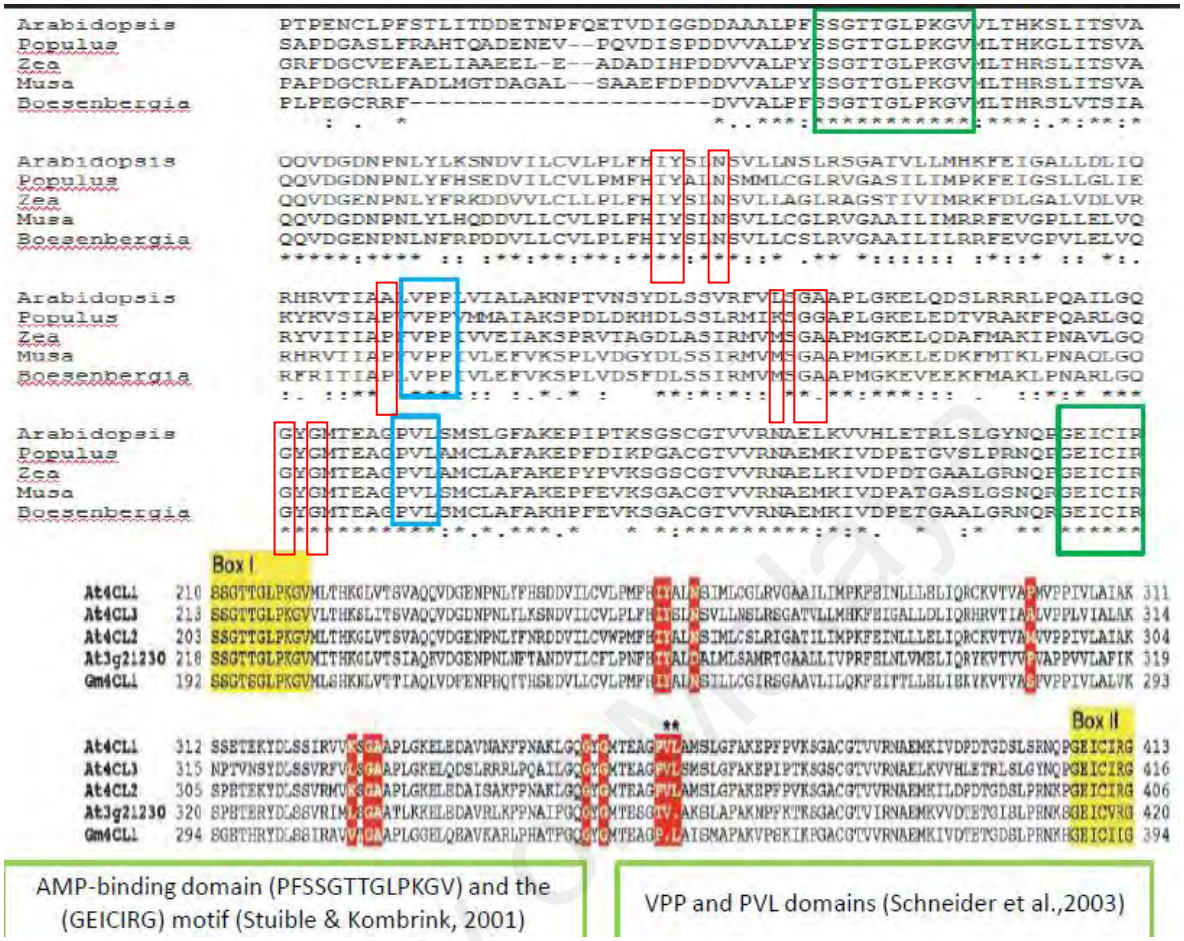
ATG and TGA was detected on the sequence and resulted in 1578bp in size (open reading frame)

ExPASy translate program



Pink colour highlight bar showed the open reading frame of *Br4CL* gene which encodes 526 amino acids.

Appendix C13: Domains and conserved region of *4CL* gene in plants



Appendix C14: *4CL* Unigene36813_All from transcriptome data (428bp)

5’GAAATCAAGCAGTACATCTCAAGACAGGTGGTTTTCTACAAGAGAATCAA
CAAAGTTTTCTTCACAGAAATCATTCCAAAGGCCCCCTCTGGAAAAATCTTG
AGAAAGGATCTAAGAGCGAAGCTAGCTGAGCAATTCCCCATCGGTCCATTT
CCATGATCATACCGACACCGTCGGTATCTTCCCATAGTGCTCTCATGGTGGC
AAGAACAAGACGTGGTGTAATATTCTCAACATAGTTCATGTTTCCAAGTCAT
TCTTTTCTGTGCTGTTTATTAACGTTGTTAATTGTTGACGGGAGGGCTGGTTG
ACCCCTCCGACTTCAGCTATAGATTTTATTTTACTTGTTCTACACTGTGTTTG
TAATGTTAATGTGGATAGAGATGTGACTTTGCTAAATGAAAAGGTTGTTATT
AGTTGGTTTCTTA-3’

5’- GCCCCCTCTGGAAAAATCTT- 3’: *4RTF* Forward primer for real time

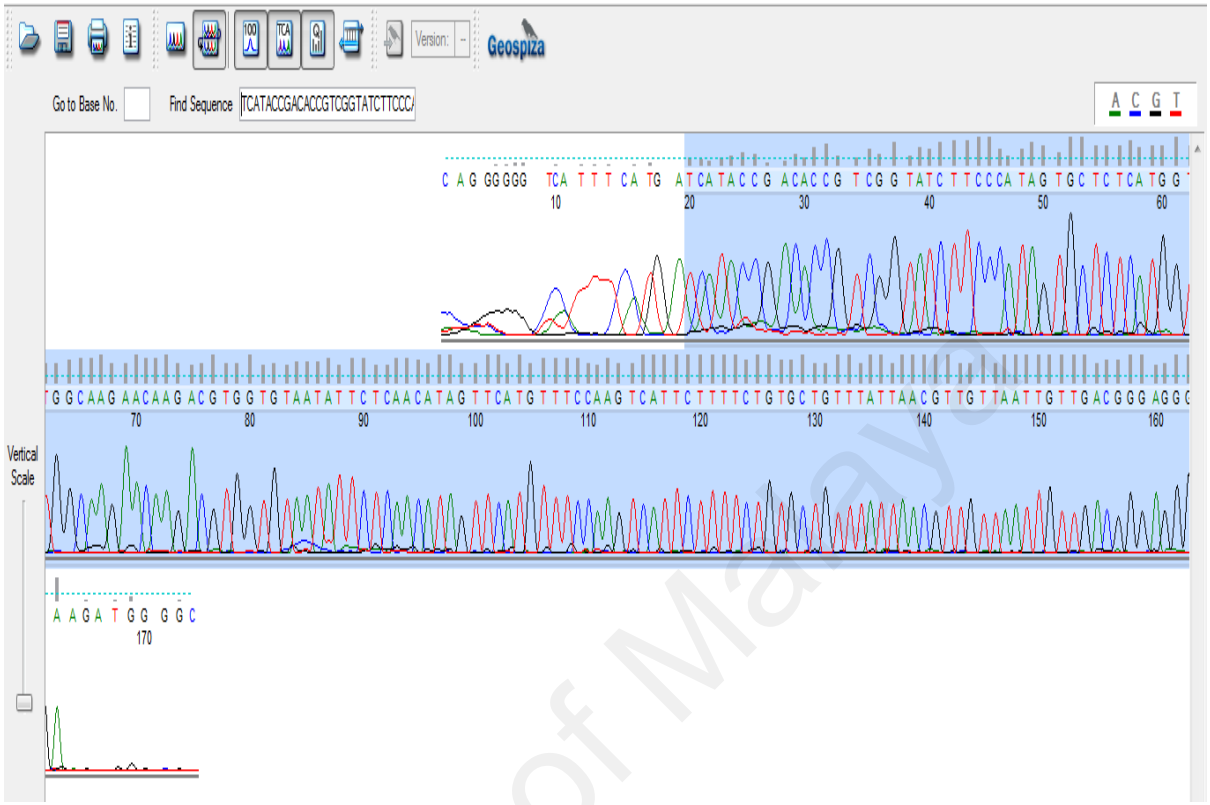
5’-TTG CCACCATGAGAGCACTATG- 3’: *4RTR* Reverse primer for real time

5’-AGCGAAGCTAGCTGAGCAATT- 3’: Forward primer for the existence of 4CL

5’-CCCTCCCGTCAACAATTAAC- 3’: Reverse primer for the existence of 4CL

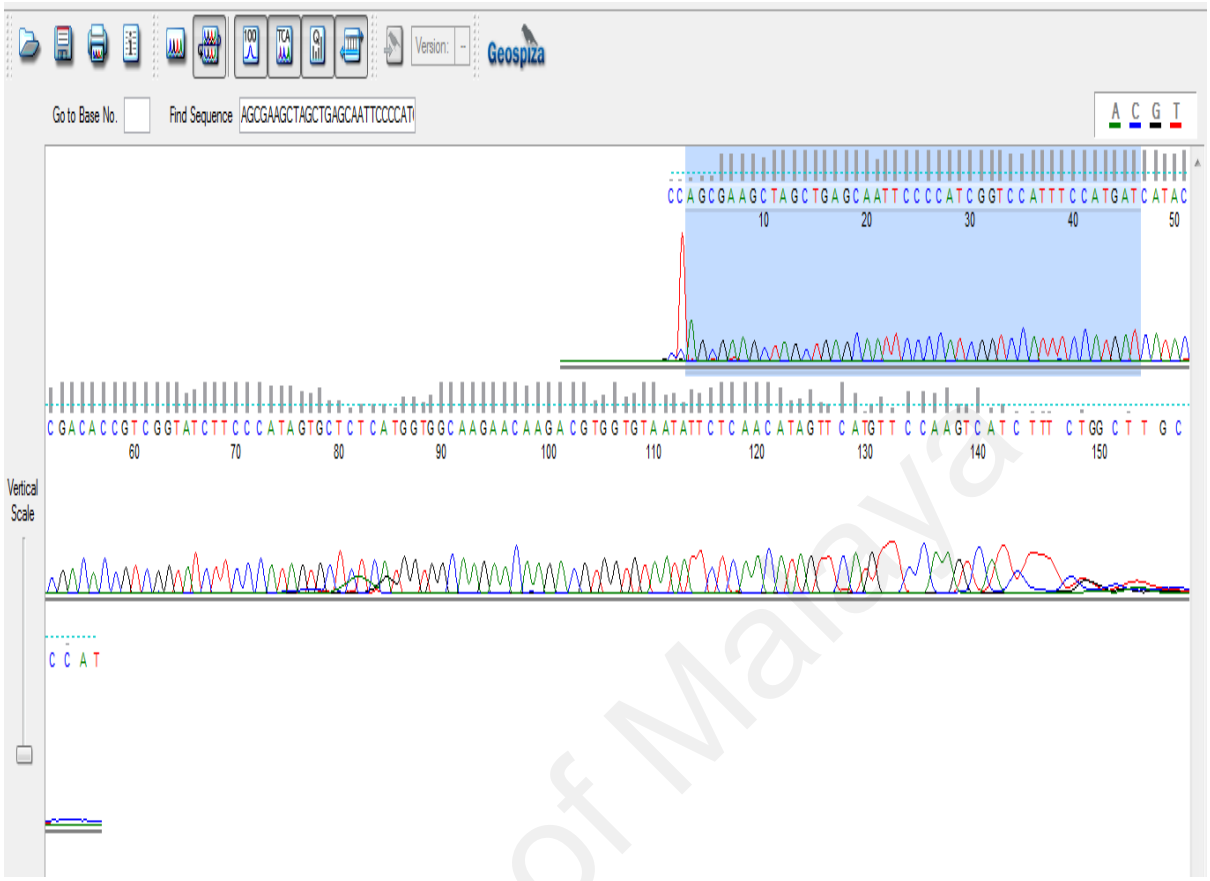
Appendix C15: Verification of the existence of *4CL* Unigene36813_All in *B.rotunda*

(1) Forward sequencing result



GAAATCAAGCAGTACATCTCAAGACAGGTGGTTTTCTACAAGAGAATCAACAAAGTTTTCTTCACAGAAA
TCATTCCAAAGGCCCCCTCTGGAAAAATCTTGAGAAAGGATCTAAGAGCGAAGCTAGCTGAGCAATTCCC
CATCGGTCCATTTCCATGATCATACCGACACCGTCGGTATCTTCCCATAGTGCTCTCATGGTGGCAAGAA
CAAGACGTGGTGTAATATTCTCAACATAGTTCATGTTTCCAAGTCATTCTTTTCTGTGCTGTTTATTAAC
GTTGTTAATTGTTGACGGGAGGGCTGGTTGACCCCTCCGACTTCAGCTATAGATTTTATTTTACTTGTTC
TACACTGTGTTTGTAATGTTAATGTGGATAGAGATGTGACTTTGCTAAATGAAAAGGTTGTTATTAGTTG
GTTTCTTA

(2) Reverse complement of reverse sequencing result



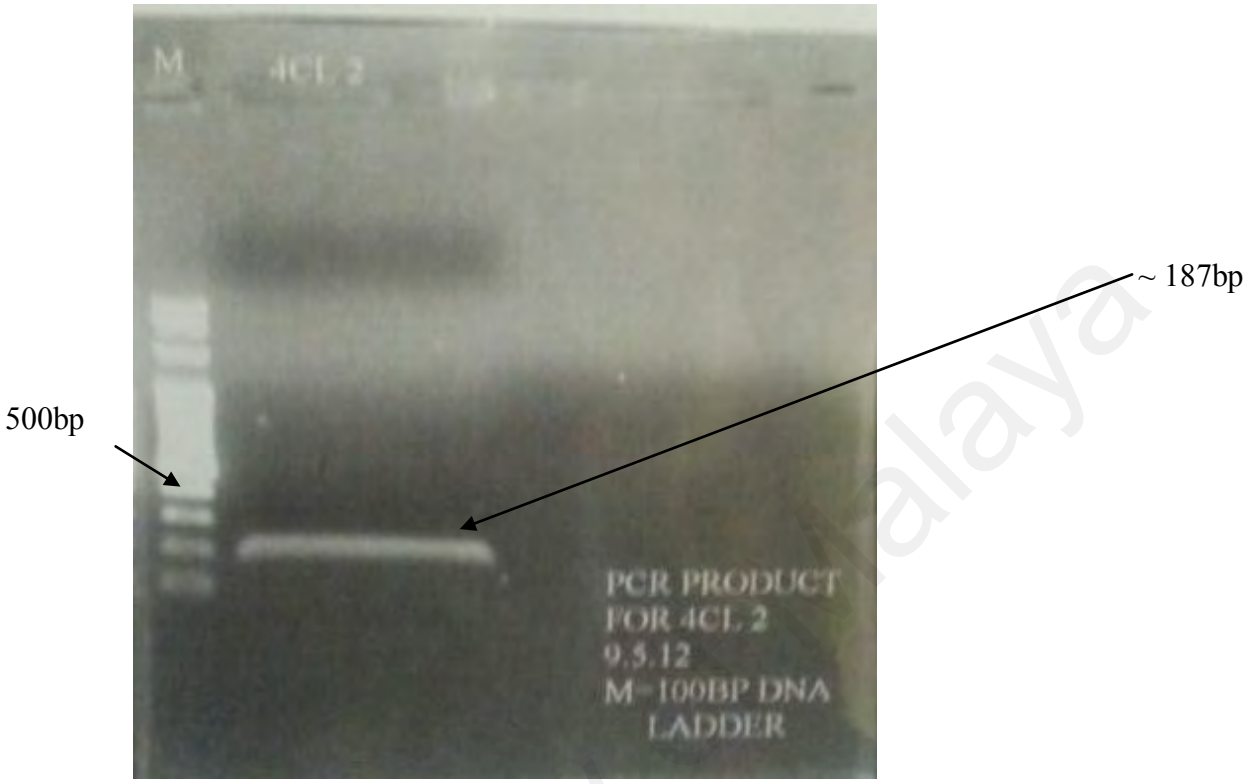
GAAATCAAGCAGTACATCTCAAGACAGGTGGTTTTCTACAAGAGAATCAACAAAGTTTTCTTCACAGAAA
TCATTCCAAAGGCCCCCTCTGGAAAAATCTTGAGAAAGGATCTAAGAGCGAAGCTAGCTGAGCAATTCCC
CATCGGTCCATTTCCATGATCATACCGACACCGTCGGTATCTTCCCATAGTGCTCTCATGGTGGCAAGAA
CAAGACGTGGTGTAATATTCTCAACATAGTTCATGTTTCCAAGTCATTCTTTTCTGTGCTGTTTATTAAC
GTTGTTAATTGTTGACGGGAGGGCTGGTTGACCCCTCCGACTTCAGCTATAGATTTTATTTTACTTGTTC
TACACTGTGTTTGTAATGTTAATGTGGATAGAGATGTGACTTTGCTAAATGAAAAGGTTGTTATTAGTTG
GTTTCTTA

(3) Combined sequence

AGCGAAGCTAGCTGAGCAATTCCCCATCGGTCCATTTCCATGATCATACCGACACCGTCGGTATCTTCCC
ATAGTGCTCTCATGGTGGCAAGAACAAGACGTGGTGTAATATTCTCAACATAGTTCATGTTTCCAAGTCA
TTCTTTTCTGTGCTGTTTATTAACGTTGTTAATTGTTGACGGGAGGG

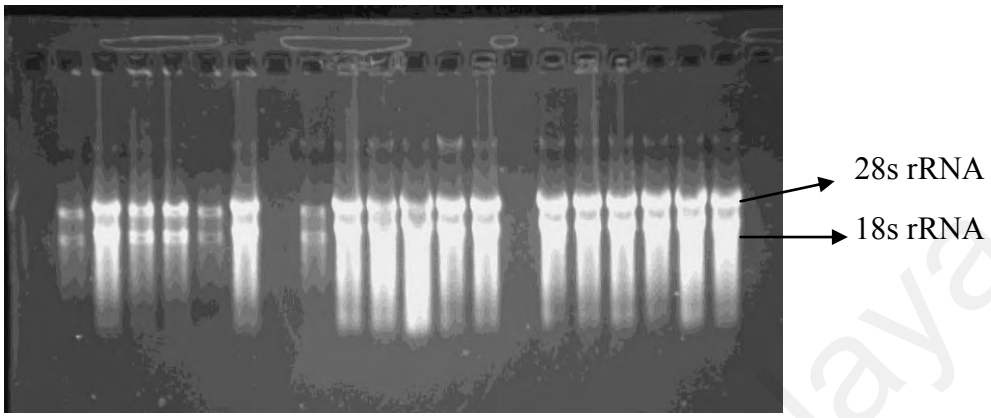
Expected size: 187bp

(4) Typical gel electrophoresis of PCR products of 4CL Unigene36813_All in *B.rotunda* genome

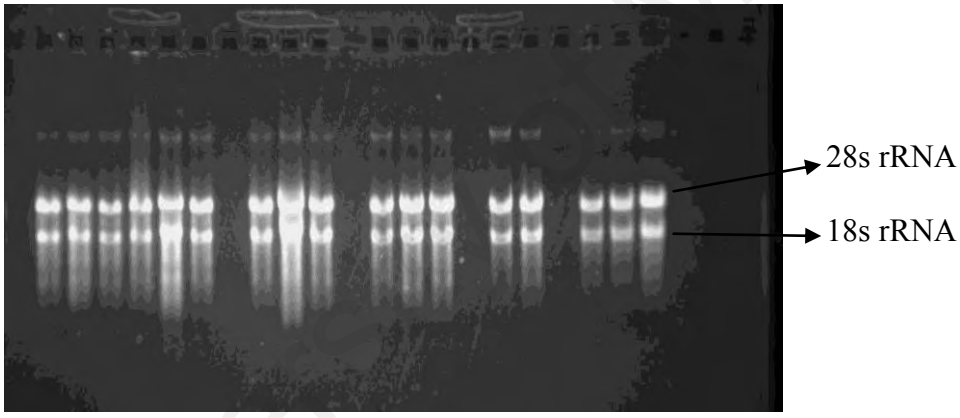


Appendix D: Gel electrophoresis of RNA extraction from treatment and control group of cell suspension of *B.rotunda*

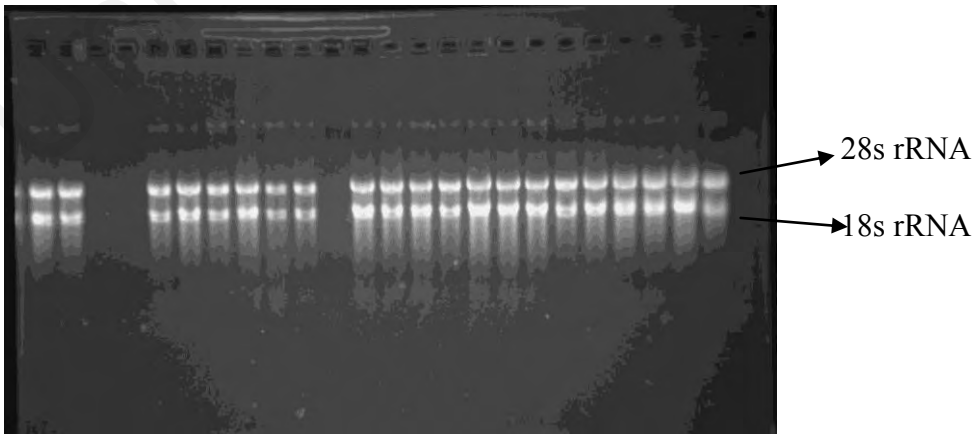
(1) Typical gel electrophoresis of RNA isolated from *B. rotunda* cell suspension cultures RNA samples from C0/1, C0/2, and C0/3



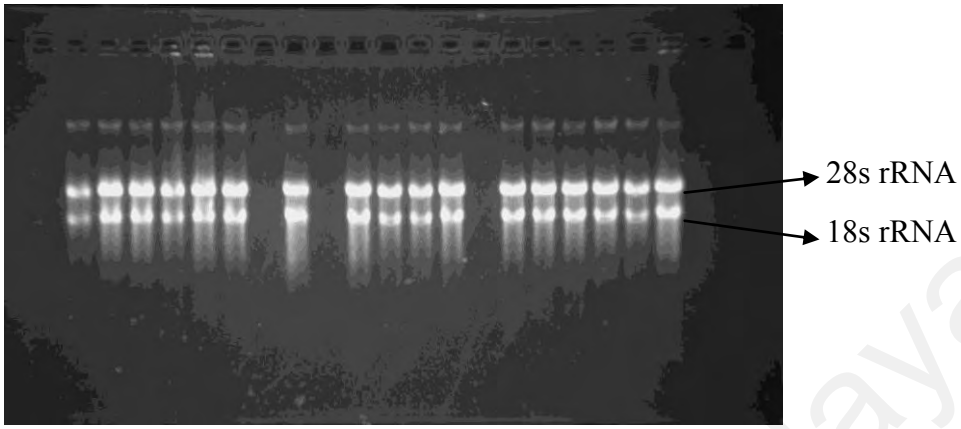
(2) Typical gel electrophoresis of RNA isolated from *B. rotunda* cell suspension cultures RNA samples from C10/1, C10/2, and C10/3



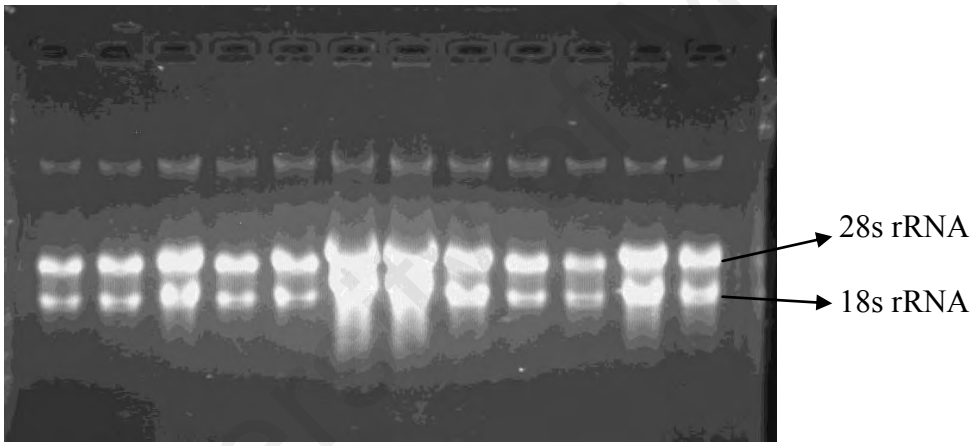
(3) Typical gel electrophoresis of RNA isolated from *B. rotunda* cell suspension cultures RNA samples from T10/1, T10/2, and T10/3



(4) Typical gel electrophoresis of RNA isolated from *B.rotunda* cell suspension cultures RNA samples from C14/1, C14/2, C14/3 and T14/1



(5) Typical gel electrophoresis of RNA isolated from *B.rotunda* cell suspension cultures RNA samples from T14/2, and T14/3



Appendix E: Concentration and purity of selected samples for RT-qPCR

(1) Concentration and purity of RNA samples

Sample Control day 0 (C0)	Concentration (µg/mL)	Purity (260 nm/280nm)	Purity (260 nm/230nm)
C0/1-6	50.7	2.00	2.55
C0/2-3	62.3	1.97	2.47
C0/3-1	42.5	2.06	2.31

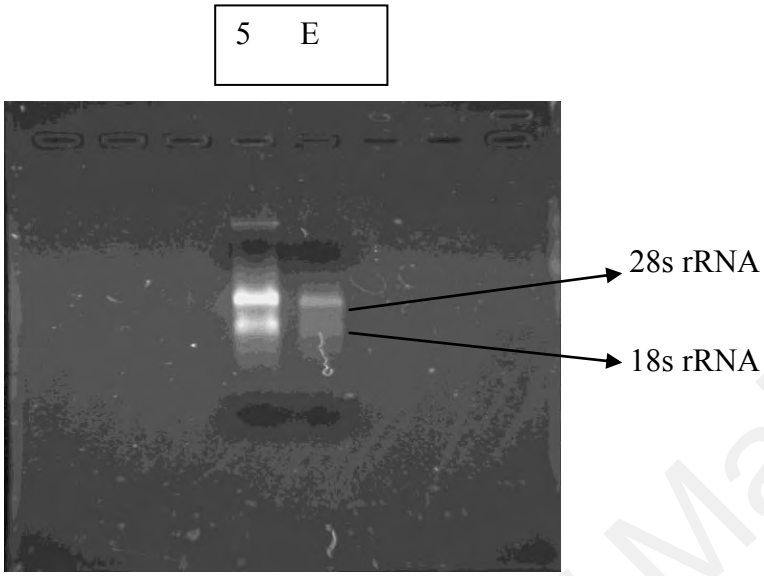
Sample Control day 10 (C10)	Concentration (µg/mL)	Purity (260 nm/280nm)	Purity (260 nm/230nm)
C10/1-5	51.4	2.03	2.71
C10/2-6	28.1	1.83	2.03
C10/3-2	739.3	2.01	2.20

Sample Control day 14 (C14)	Concentration (µg/mL)	Purity (260 nm/280nm)	Purity (260 nm/230nm)
C14/1-5	647.4	1.93	2.18
C14/2-4	544.5	1.88	2.30
C14/3-4	55.2	2.11	2.96

Sample Treatment day 10 (T10)	Concentration (µg/mL)	Purity (260 nm/280nm)	Purity (260 nm/230nm)
T10/1-3	41.8	2.06	2.75
T10/2-1	931.7	1.96	2.47
T10/3-6	56.7	1.95	2.59

Sample Treatment day 14 (T14)	Concentration (µg/mL)	Purity (260 nm/280nm)	Purity (260 nm/230nm)
T14/1-4	33.6	2.00	4.19
T14/2-2	923.4	1.90	2.45
T14/3-1	768.4	1.81	2.05

(2) Typical gel electrophoresis of RNA and DNAse-treated RNA samples from control group at 0 day

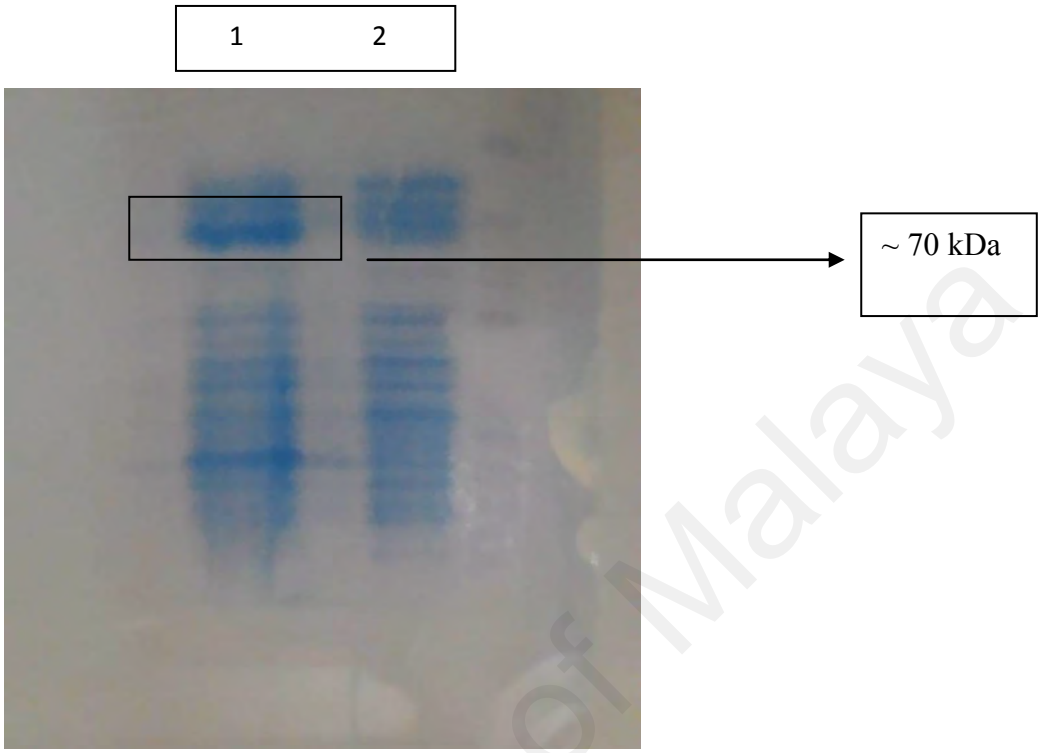


Lane 5-E: sample C0

5 (C0): RNA isolated from control group at 0 day
E (C0): DNAse-treated RNA samples from control group at 0 day

Appendix F: Verification of *Br4CL* protein expression

(1) SDS-PAGE gel of 4CL protein



Lane 1: Cell lysate with induction

Lane 2**:** Cell lysate without induction