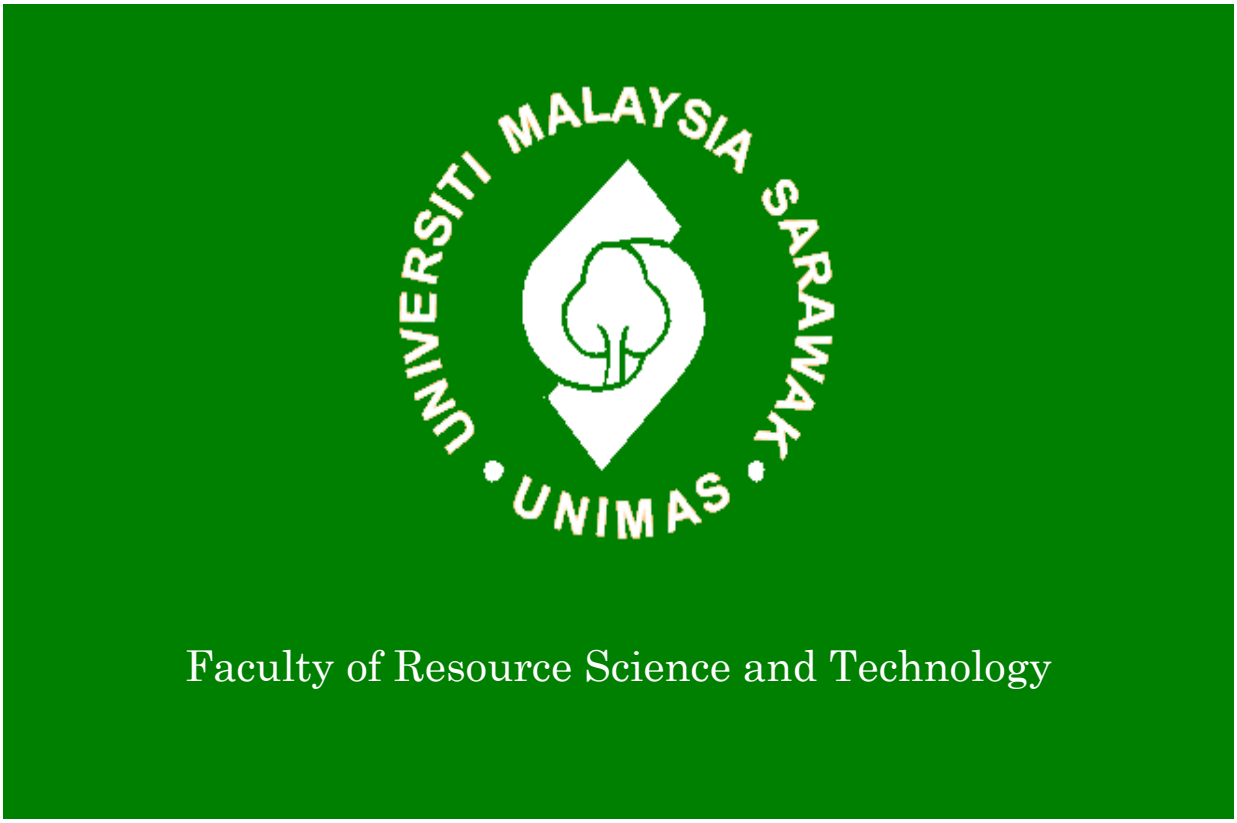

Molecular Cloning and Expression Study of Grb10 Interacting GYF Domain Protein 1 (GIGYF 1) Gene in Zebrafish (*Danio rerio*) Development

How Kim Loong

Bachelor of Science with Honours
(Resource Biotechnology)
2011

Molecular Cloning and Expression Study of Grb10 Interacting GYF Domain Protein 1 (GIGYF 1) Gene in Zebrafish (*Danio rerio*) Development

HOW KIM LOONG (21069)

A thesis submitted in partial fulfilment of the requirement for the degree of Bachelor of Science with Honours (Resource Biotechnology)

Faculty of Resource Science and Technology
UNIVERSITI MALAYSIA SARAWAK
2011

ACKNOWLEDGEMENTS

Above all, I would like to thank God for His blessings upon the completion of this project. My sincere gratitude to my supervisor, Dr Lee Kui Soon for giving me the opportunity to work on this project under his careful guidance, valuable advice, generous suggestions and encouragement. Besides, I would like to dedicate my special thanks to Ms. Ma Xiang Ru (Ph.D) candidate for sharing experiences and knowledge with me, patiently guiding me throughout my project and making valuable comments on this project. I also like to thank Dr Ho Wei Seng and A.P. Dr Edmund Sim Ui Hang for providing essential equipments and reagents to me for my laboratory works. Furthermore, I would like to thank Mr Iskandar, Mrs. Sheela Ungau and other lab assistants for their assistances especially in term of technical problems and provided a most congenital and supportive atmosphere through my project. Finally, not forgetting my friends and family who saw me through all the joys and frustrations of this project, while support, comfort and helping me stay focused on the present.

DECLARATION

I declare that this thesis entitled "Molecular Cloning and Expression Study of Grb10 Interacting GYF Domain Protein 1 (GIGYF 1) Gene in Zebrafish (*Danio rerio*) Development" is the result of my own research except for citation and references which have been duty acknowledged. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

How Kim Loong

Resource Biotechnology

Department of Molecular Biology

Faculty of Resource Science and Technology

University Malaysia Sarawak

Table of Contents

List of Abbreviations

Grb10	Growth Factor Receptor Bound Protein 10
GYF domain	Glycine-Tyrosine-Phenylalanine domain
GIGYF1	Grb10 Interacting GYF Domain Protein 1
GIGYF2	Grb10 Interacting GYF Domain Protein 2
IGF-1	Insulin-like Growth Factor 1
IGF-1R	Insulin-like Growth Factor 1 receptor
CD2BP2	CD2 antigen cytoplasmic tail-binding protein 2
SMY2	Suppressor of Myo2-66 protein
TNRC 15	Trinucleotide Repeat-Containing 15
PERQ	PERQ (Pro-Glu-Arg-Gln)
cDNA	Complimentary DNA
RT-PCR	Reverse Transcription-Polymerase Chain Reaction
PD	Parkinson's Disease
hpf	hour post fertilisation
dpf	day post fertilisation
PRS	Proline-rich sequence
DDC model	Duplication-Degeneration-Complementation model
LG	Linkage group

List of Tables

List of Figures

Molecular Cloning and Expression Study of Grb10 Interacting GYF Domain Protein 1 (GIGYF 1) Gene in Zebrafish (*Danio rerio*) Development

How Kim Loong

Resource Biotechnology
Faculty of Resource Science and Technology
University Malaysia Sarawak

ABSTRACT

Grb10 **I**nteracting **GYF** domain (GIGYF) family genes consist of two members, namely GIGYF1 and GIGYF2 respectively. These two genes were initially identified through their interaction with Grb10, an adapter protein that binds activated Insulin-like Growth Factor I (IGF-1) and insulin receptors and may thus regulate IGF-I-stimulated growth. Recently, a study reported that GIGYF2 may play a causal role in familial Parkinson's disease (PD). However, the function of GIGYF1 is largely unknown. Since both GIGYF1 and GIGYF2 are homologous proteins which share the common features, thus it would be interesting to find out the functional role of GIGYF1. Two GIGYF1 genes namely *GIGYF1a* and *GIGYF1b* were found in the zebrafish genome which consistent with the occurrence of genome duplication in zebrafish. Gene duplication event would enable further mutational study of GIGYF1 gene, especially in determining its relationship with PD which is stated in Duplication-Degenerative-Complementation (DDC) model. Total RNA was extracted from 10 different embryonic developmental stages (4cells, 3hpf, 5.25hpf, 11hpf, 18hpf, 24hpf, 36hpf, 48hpf, 3dpf, 5dpf) of zebrafish. The extracted RNA was then used to synthesize cDNA template and subjected to reverse transcription PCR (RT-PCR) analysis. The *GIGYF1a* and *GIGYF1b* gene were found to be expressed at all selected developmental stages and this might provide some hints to its functional role.

Keywords: *GIGYF1 gene, molecular cloning, gene expression, zebrafish, RT-PCR*

ABSTRAK

***G**rb10 **I**nteracting **GYF** domain (GIGYF) gen keluarga terdiri daripada dua ahli, iaitu GIGYF1 dan GIGYF2. Kedua-dua gen ini awal-awalnya dikenalpasti melalui interaksi mereka dengan Grb10, sebuah protein adaptor yang mengikat dan mengaktifkan insulin-like growth factor (IGF-1) dan reseptor insulin dan dengan demikian dapat mengawal pertumbuhan yang dirangsangkan oleh IGF-1. Sejak kebelakangan ini, sebuah kajian melaporkan bahawa GIGYF2 mungkin memainkan peranan kuasal dalam penyakit Parkinson familial's (PD). Namun, fungsi GIGYF1 masih belum diketahui. Sedangkan kedua-dua GIGYF1 dan GIGYF2 adalah protein homologi yang berkongsi ciri-ciri umum, jadi, amat menarik minat untuk mengetahui peranan GIGYF1. Dua gen GIGYF1 iaitu GIGYF1a dan GIGYF1b telah dijumpai dalam genom zebrafish yang konsisten dengan kejadian duplikasi genom pada zebrafish. Kejadian duplikasi gen membolehkan kajian yang lebih lanjut terhadap mutasi gen GIGYF1, terutamanya dalam menentukan hubungannya dengan PD yang dinyatakan dalam model Duplikasi-Degeneratif-Complementation (DDC). Keseluruhan RNA dari sepuluh tahap perkembangan embrio zebrafish yang berbeza iaitu 4cells, 3hpf, 5.25hpf, 11hpf, 18hpf, 24hpf, 36hpf, 48hpf, 3dpf, 5dpf telah diekstrak dan digunakan untuk mensintesis cDNA supaya dianalisa dengan reverse transcription PCR (RT-PCR). Gen GIGYF1a dan GIGYF1b didapati mempunyai ekspresi pada kesemua tahap perkembangan embrio zebrafish yang dipilih dan ini mungkin memberikan beberapa petunjuk kepada peranan fungsional gen GIGYF1.*

Kata kunci: Gen GIGYF1, pengklonan molekul, ekspresi gen, zebrafish, RT-PCR

1.0 Introduction

Grb10 **I**nteracting **GYF** domain (GIGYF) family genes are GYF domain containing proteins which have the GYF domain that can recognize the proline-rich sequence (Koftler *et al*., 2005). It consists of two members, namely GIGYF1 and GIGYF2 respectively. In mouse, GIGYF1 and GIGYF2 have been found via yeast two-hybrid screening, showing an interaction with the tandem proline-rich region in the N-terminus of Grb10 (Giovannone *et al.*, 2003). Grb10 is an adaptor protein that binds to the intracellular domains of activated tyrosine receptor including insulin-like growth factor (IGF-1) and insulin receptors (Giovannone *et al.*, 2003). It is believed to have a role in regulating the cellular responses to IGF-1, insulin and hormones (Giovannone *et al.*, 2003). However, the function of GIGYF family genes is remained unknown due to limited information. According to Giovannone *et al.* (2003), GIGYF1 protein is proposed to have a particular cooperation role with Grb10 that linked to activate IGF-1 receptors (IGF-1R), and modifies IGF-1 signaling which plays a key role in embryonic growth and development. Recent study carried out by Lautier *et al.* (2008) reported that there is potential genetic linkage between GIGYF2 and familial Parkinson's disease (PD). Parkinson's disease is a kind of neuronal degenerative disease normally caused by the loss of dopamine-producing brain cells and result in uncontrollable movement (Nicholson *et al.*, 2002). The study reported that the mutational effect of GIGYF2 gene in the PARK II locus on chromosome 2q36-37 in human might possess causative effect towards familial PD. However, finding still failed to rule out the role of GIGYF2 in PD onset. Since GIGYF1 shows a high degree of sequence homology with GIGYF2 (Koftler *et al*., 2005), thus it would be interesting to figure out the biological function of GIGYF1 protein. On the other hand, expression study performed by Lee (2008) reported that the GIGYF family genes are strongly expressed

during somite and brain development stage within 4-cell stage until 34 hpf (hour post fertilization) stage. This result might provide an insight to the formation of the brain degenerative disease or muscle disorder by study GIGYF family genes. Moreover, a second copy of GIGYF1 has been identified at the zebrafish linkage group (LG) 10 by Foong in 2010, where as the first copy has been accomplished by Lee in 2008 at the LG7. The presence of these two copies of GIGYF1 genes, designed as *GIGYF1a* and *GIGYF1b* respectively consistent with the occurrence of genome duplication in zebrafish (Woods *et al.*, 2005). This second copy of gene could provide a reliable platform for mutational study of GIGYF1 gene especially in determine its relationship with PD. As stated in the Duplication-Degeneration-Complementation (DDC) models, duplicated gene would accumulates loss-of-function through degeneration mutation and subsequently complement each other to stay viable (Force *et al*., 1999). In this way, a complex mammalian gene function might be shared complimentary by two GIGYF1 copy genes under the DDC model.

Current published expression studies on GIGYF family genes was only restricted to Northern blot analysis on adult mouse tissue. Thus, the main aim of this project is to examine the tissue distribution and expression pattern of GIGYF1 protein in 10 different embryonic developmental stages of zebrafish by performing RT-PCR. Expression study of *GIGYF1a* and *GIGYF1b* might provide some clues to reveal the functional role of GIGYF1 genes through the observation of mRNA expression pattern of the genes.

The main objectives of this study are:

1) To isolate and clone the *GIGYF1a* and *GIGYF1b* gene from zebrafish.

2) To study the expression pattern of *GIGYF1a* and *GIGYF1b*.

2.0 Literature review

2.1 Grb10 Interacting GYF domain (GIGYF) family genes

Two Grb10 interacting proteins, GIGYF1 (Grb10 interacting GYF domain protein 1, or previously known as PERQ amino acid rich with GYF domain 1, PERQ1) and GIGYF2 (or also named PERQ2 or Trinucleotide Repeat-Containing 15, TNRC 15), which contain a GYF domain were identified through yeast two-hybrid screening in mouse using the N-terminal portion of Grb10 delta as bait (Giovannone *et al*., 2003). Deletion analysis revealed that at least two of the three proline-rich motifs of Grb10 are required for binding to both GIGYF1 and GIGYF2. As reported by Giovannone *et al.* (2003), these two proteins shared the common features such as three potential tyrosine phosphorylation sites and candidate binding sites for protein phosphatase-1 (RVPF), 14-3-3 proteins (RSpSXP), and phospholipase C_{γ}(V(F)DDY). The distinguishable characteristics of these molecules are proline-rich region in the proteins, a consensus bipartite nuclear localization motif, a region of clathrin light chain homology, follows by multiple stretches of Gln and Glu residues in the C-terminal of both ends of the proteins. It is discovered that IGF-I stimulation can increase the affinity of GIGYF1 for Grb10 delta and results in a transient binding of both GIGYF1 and Grb10 to IGF-I receptor. Thus GIGYF1 may promote the association of the IGF-I receptor with mGrb10 (Giovannone *et al*., 2003). At later time points, Grb10 remains linked to IGF-I receptors when GIGYF1 was dissociated. Overexpression of the Grb10 binding fragment of GIGYF1 enhances IGF-I receptor-stimulated tyrosine autophosphorylation suggesting that Grb10 and GIGYF1 may act cooperatively to regulate receptor signaling (Giovannone *et al*., 2003). In addition, according to Lautier *et al.* (2008), the mutational of GIGYF2 gene is a frequent cause on familial PD with its mutational effect on PARKII locus. However, this finding has been

debated controversially as the changes of amino acid due to mutations are theoretically would not always be conserved during evolution.

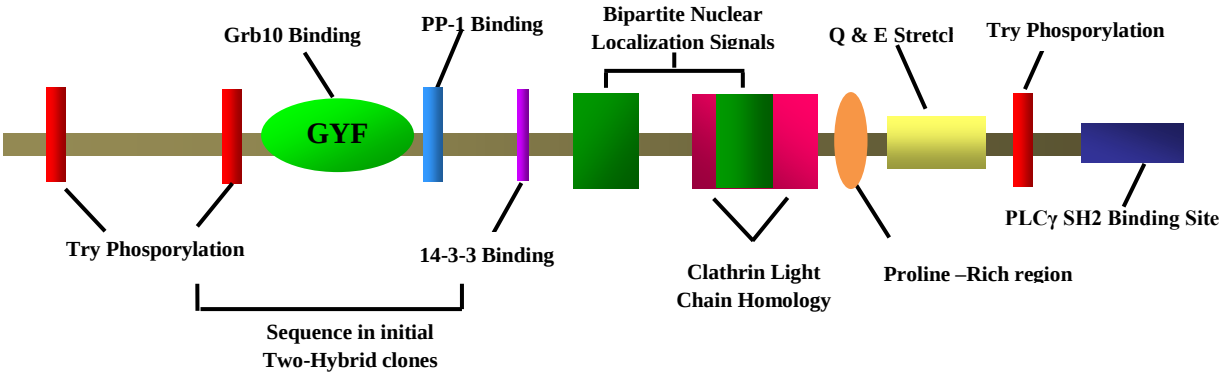

Figure 2.1: Consensus structural motif present in GIGYF1 and GIGYF2 (Adapted from Giovannone *et al*., 2003)

2.2 GYF domain

The Glycine-Tyrosine-Phenylalanine, or GYF domain is first identified in the human intracellular CD2 binding protein (CD2BP2) which binds to a tandem repeat of proline-rich sequence (PRS) in the CD2 tail region (Freund *et al*., 1999). It is a small, versatile proline-rich sequence recognizing adaptors which present in most eukaryotes (Nishizawa *et al*., 1998; Freund *et al*., 1999). It forms a bulge-helix-bulge structure that contributes to the proline sequence binding (Gloeckner *et al.*, 1998). There are two subfamilies of GYF domains: the CD2BP2-type and the SMY2-type. CD2BP2-type is characterized by a tryptophan at position 8 and extended loop between β-strands 1 and 2 whereas for the SMY2-type is characterized by a shorter β1-β2 loop and its recognition of a PPGX motif (Kofler & Freund, 2006). It is hypothesized that CD2BP2 type of GYF domains would function in T-cell signaling and mRNA splicing (Kofler *et al*., 2004), while the SMY2-type of GYF domains would function in transport processes (Lilie & Brown, 1992). Recent

study have demonstrated the critical importance of the GYF tripeptide in ligand binding and allowed identification of novel interaction partners. In mouse, two GIGYF proteins have been shown to interact with Grb10 (Giovannone *et al*., 2003). Consequently, a possible functional role of GYF domain in Grb10-mediated insulin-receptor signaling has been suggested.

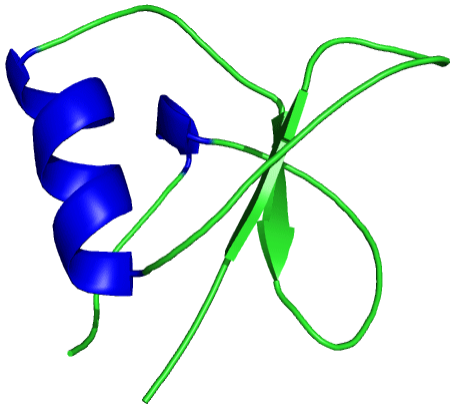

Figure 2.2: GYF domain of CD2

(Adapted from The Pawson Lab: http://pawsonlab.mshri.on.ca/index.php?option=com_content&task=view&id=162&Itemid=64)

2.3 Zebrafish

Zebrafish, also known as *Danio rerio* belongs to the Super-Order Teleostei (bony fish) and the Family Cyprinidae. It becomes widely used as a genetic model to uncover specific functions of unknown proteins (Dooley & Zon, 2000; Rubinstein, 2003) and proven to be an ideal vertebrate model for studying vertebrate embryogenesis. They are small in size, can be economically maintained in large numbers, rapid embryonic development and also have relatively short generation times (Westerfield, 1995). Being transparent during early embryonic stages, easy to manipulate and highly reproductive have makes the zebrafish become popular model organism for molecular studies (Moro *et al*., 2007). Zebrafish breed through external fertilization and this makes it easy to get their eggs for studies (Grunwald

& Eisen, 2002). According to Eisen and Weston, (1993), their embryos are completely transparent during the first 24 hours of development. This gives researchers the opportunity to observe the developing organs and visualize cells directly in the living embryo as well as for *in situ* mRNA hybridization analysis in whole-mount embryos. They also have far lesser cells in their organ to fulfill the equivalent function compared to other higher vertebrates. This gives researcher easier identification of the locations of the migrating cells. Zebrafish and humans share much of the same genetic material than other invertebrate organisms such as *Drosophila* (Barbazuk *et al.,* 2000; Postlethwait *et al.,* 1998). According to Woods *et al.* (2005), the role of human genes in development and disease can be approached by study the zebrafish mutation since gene functions are often conserved in vertebrates. Previous studies suggested that whole-genome duplication had been occurred in the teleost lineage after the divergence from tetrapod lineage. Thus zebrafish is believed to have experience an additional genome wide duplication event during evolution as well (Woods *et al.*, 2005). As the duplicated chromosome segments in zebrafish is always correspond with the specific segments in mammals, especially for human, thus it can serve to suggest functions for human genes through zebrafish mutational study.

2.4 Gene duplication and "Duplication-Degeneration-Complementation (DDC)" model

According to Ohno (1970), gene duplication is a key mechanism in evolution for acquiring new genes and creating genetic novelty in organisms. He proposed that duplicated genes are not constrained to maintain their original function and, hence, they can accumulate divergent mutations and evolve new functions, whereas the ancestral copy maintains its

original functions. Gene duplication typically arise from unequal crossing over, retroposition or chromosomal (or genome) duplication (Watts & Watts, 1968; Zhang, 2003). Unequal crossing-over during meiosis will generates tandem sequence where duplicated genes are usually being located at non identical location in parental genes. Meanwhile, retroposition occurs when the messenger RNA (mRNA) is retrotranscribed to complementary DNA (cDNA) instead of being translated into protein, and inserted in a new genome position (Zhang, 2003). Gene duplication resulted from retroposition is commonly unlinked to the original gene as the insertion is randomized (Ohta, 2003; Zhang, 2003). On the other hand, chromosomal or genome duplication occurs due to lack of disjunction among the daughter chromosomes after DNA replication (Zhang, 2003).

Subsequent to duplication event, three dominant hypotheses have been proposed regarding the rules governing the functional divergence of duplicated genes. The nonfunctionalization hypothesis argues that after duplication, the genes appear to be redundant in function, and thus one of the duplicates would become a pseudogene to be finally lost by degenerative mutation. The neofunctionalization hypothesis on the other hand, serves as an alternative option for both paralogues to be preserved, whereby one daughter gene would acquire new beneficial function, while the other retains the original function (Ohno, 1970). However, this process was believed to be extremely uncommon (Wagner, 1998). Force *et al*. (1999) proposed another alternative hypothesis, called Duplication-Degeneration-Complementation (DDC) or subfunctionalization, whereby both duplicated genes can each take on part of the original functions of the ancestral gene, such that the duplicated genes complement to each other by jointly performing the necessary subfunction of the ancestral gene. DDC process is capable to prevent the loss of duplicated genes by expanding the time period for which genes are expose to natural selection, and hence facilitates the opportunity for the beneficial mutation to evolve new functions.

3.0 Materials and Method

3.1 Primer Design and Sequence

The primers in this project were designed by using two different programs. Primer3 (http://waldo.wi.mit.edu/cgi-bin/primer3.cgi/primer3www.cgi) was used to suggest suitable primer pairs for PCR amplification. All suitable primer pairs suggested by Primer3 were analyzed for hairpin, palindromes, dimmers, and melting temperature (Tm) by using NetPrimer (http://www.premierbiosoft.com/netprimer/netprimer.html). At least a rating of 85 (out of 100; NetPrimer) primers were selected.

3.2 Isolation of Total RNA

Total RNA were extracted from various embryonic developmental stages (4-cell stage until 5dpf) of zebrafish using Tri Reagent. Approximately 400 (less than 2hpf) and 150 (all other stages) embryos were dechorionated using needles and killed by euthanizing in excessive Tricane. The embryos were placed and homogenized in a 1 ml of Tri Reagent contained tube and centrifuged at 12,000 rpm for 10 minutes. After centrifugation, the supernatant was transferred to a new tube and 200 µl of chloroform was added. The samples were shaken for 15 seconds and left at room temperature for approximately 5 minutes. The samples were centrifuged at 12,000 rpm for 15 minutes and the upper phase was transferred to a new tube. 500 µl of isopropanol was added to each sample and incubated at room temperature for 10 minutes. The samples were centrifuged at 12,000 rpm for 10 minutes to precipitate the RNA. The RNA pellet was washed with 70% ethanol and centrifuged again at 7,500 rpm for 5 minutes. The supernatant was removed and the RNA pellet was allowed to dry before adding with 10 µl of RNase free water.

3.3 Agarose Gel Electrophoresis

A 1% (w/v) agarose gel was prepared by using 0.5g of agarose powder into 50ml of 1x TAE buffer. The mixture was heated in microwave for 1 to 2 minutes to dissolve the powder. The gel solution was cooled to about 60 °C and 2µl of Ethidium bromide (10mg/ml) was added and the gel was poured into the moulding tray with the comb in position. After the gel solidified, it was submerged in the 1x TAE buffer to 2 to 5mm depth and ready for use. 2 µl of loading dye (6X) and 6 µl of distilled water were mixed with 2 µl of sample before load to the gel. The gel was run at 120V for approximately 30 minutes and was examined using an UV Transilluminator.

3.4 Quantification of RNA

For spectrophotometric measurement, 2 µl of DNA aliquots was diluted with 68 µl of distilled water in the cuvette and quantified using spectrophotometer. The absorbance values of 260 nm, 280 nm and 320 nm were recorded. The purity of the samples was determined by calculating the ratio of $A_{260\text{-}320/A280\text{-}320}$ and the concentration of RNA was calculated as shown below:

$$\textbf{RNA (µg/µl)} = [A_{260} \times (40\ \text{µg RNA/ml})/(1\ A_{260}\ \text{unit}) \times (\text{dilution factor})] / 1000$$

3.5 cDNA Synthesis

The cDNA synthesis was carried out by using SuperScript™ III First-Strand cDNA Synthesis System (Invitrogen) according to the manufacturer instructions. A mixture of 1 µl of random hexamer primers, 1 µl of RNA, 1µl of 10mM of NTPs mix and 7µl of Nuclease-Free Water was incubated for 5 minutes at 65°C and then at least 1 minute on ice before spun for 10 seconds in a microcentrifuge to collect the condensate. The cDNA Synthesis Mix was prepared according to the manufacturer instructions. 10 µl aliquots of the cDNA Synthesis Mix were added together with 10 µl of RNA and primer mix to each reaction tube on ice for a final volume of 20 µl. Then the tubes were incubated at 25 ºC for 10 minutes and incubated again at 50 ºC for 50 minutes before heat inactivated at 85 ºC for 5 minutes. 1 µl of RNase H was added to each reaction tube and incubated at 37 ºC for 20 minutes. The reaction tubes were stored at -20 ºC prior to PCR amplification.

Table 3.1: cDNA synthesis reaction conditions

Experiment Reaction	**×1**	**Master Mix (×10)**
10X RT Buffer	2.00 µl	20.00 µl
25 mM $MgCl_2$	4.00 µl	40.00 µl
0.1 M DTT	2.00 µl	20.00 µl
RNaseOUT™ (40 U/µl)	1.00 µl	10.00 µl
SuperScript™ III Reverve Transcriptase (200 U/µl)	1.00 µl	10.00 µl
Final volume	10.00 µl	100.00 µl

Table 3.2: Negative control for cDNA synthesis reaction conditions

Experiment Reaction	**×1**	**Master Mix (×10)**
10X RT Buffer	2.00 µl	20.00 µl
25 mM $MgCl_2$	4.00 µl	40.00 µl
0.1 M DTT	2.00 µl	20.00 µl
RNaseOUT™ (40 U/µl)	1.00 µl	10.00 µl
SuperScript™ III Reverve Transcriptase (200 U/µl)	-	-
Nuclease-Free Water (to a final volume of 10 µl)	1.00 µl	10.00 µl
Final volume	10.00 µl	100.00 µl

3.6 Reverse Transcription Polymerase Chain Reaction (RT-PCR)

RT-PCR was carried out in a 25µl reaction per tube by preparing master mix (Table 3.3). All components were added in master mix except for cDNA template. PCR was performed under the following condition: an initial cycle of template denaturation at 95°C for 2 minutes, followed by 35 cycles of amplification at 94°C for 30 seconds, a primer annealing step at a temperature appropriate for the primers for 1 minute, an extension step at 72°C for 1 minutes and a final extension step at 72°C for 5 minutes. To visualise the RT-PCR result, 5 µl of each sample was run on 1% (w/v) agarose gel at 90V for approximately 1 hour with ethidium bromide staining.