CHARACTERIZATION OF THE DYSTROPHIN GENE IN *MACROBRACHIUM ROSENBERGII*

NOOR ANEES FATHIMA

FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2016

CHARACTERIZATION OF THE DYSTROPHIN GENE IN *MACROBRACHIUM ROSENBERGII*

NOOR ANEES FATHIMA

DESSERTATION SUBMITTED IN FULFILMENTOF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE

FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2016

UNIVERSITY OF MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: Noor Anees Fathima

Registration/Matric No: SGR130028

Name of Degree: Master of Science

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

Characterization of the dystrophin gene in *Macrobrachium rosenbergii*

Field of Study:

Genetics and Molecular Biology

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;

(2) This Work is original;

(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;

(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;

(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;

(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature   Date:

Subscribed and solemnly declared before,

Witness's Signature   Date:

Name:

Designation:

ABSTRACT

The dystrophin gene encodes one of the most crucial components of muscle fibers, the dystrophin protein. As the largest gene in humans, it plays an important role in maintaining cytoskeleton structure and normal muscle function. Its absence results in a mortal muscular disease characterized by muscle degeneration and dysfunction. Due to the complexity of dystrophin in humans, the majority of studies into the function of this gene have been conducted in simpler organisms. In this study, the *Macrobrachium rosenbergii* dystrophin was identified and computationally characterized through bioinformatics methods. Its protein domains were identified and compared to domains in homologous dystrophin sequences from other organisms. Differential expression of the dystrophin gene in response to White Spot Syndrome Virus (WSSV) was also studied. An immediate down regulation of dystrophin expression was observed upon WSSV infection, followed by a significant up regulation in the later stages of the infection. Due to the significance of intracellular calcium concentration in dystrophin deficiency studies, it was quantified in healthy *M. rosenbergii* muscle tissues as well as at different time points post WSSV infection. A two-fold increase in intracellular calcium concentration was seen followed by its stabilization 24 hours post infection, supporting existing theories of calcium-dystrophin relations in dystrophin deficient muscles. Similarly, Transmission Electron Microscope (TEM) comparison of healthy and infected *M. rosenbergii* muscle tissues morphology showed distinct deterioration of the muscle fibers and mitochondria. Several possible explanations for these observed changes were postulated and discussed in detail, including possible cleavage of dystrophin due to WSSV induced protease expression.

ABSTRAK

Gen dystrophin mengekod salah satu komponen yang paling penting dalam gentian otot, iaitu protein dystrophin. Sebagai gen terbesar manusia, ia memainkan peranan penting dalam mengekalkan struktur sitoskeleton dan fungsi otot normal. Ketiadaannya pada manusia menyebabkan penyakit otot manusia yang dicirikan oleh kemerosotan otot dan disfungsi. Oleh kerana kerumitan dystrophin pada manusia, majoriti kajian ke atas fungsi gen ini telah dijalankan dalam organisma mudah. Dalam kajian ini, *Macrobrachium rosenbergii* dystrophin telah dikenal pasti dan pengiraannya dicirikan melalui kaedah bioinformatik. Domain protein daripada *M. rosenbergii* dystrophin dikenal pasti dan dibandingkan dengan domain dalam urutan dystrophin homolog daripada organisma lain. Expresi gen dystrophin sebagai tindak balas kepada White Spot Syndrome Virus (WSSV) juga dikaji. Expresi dystrophin yang rendah diperhatikan apabila jangkitan WSSV, diikuti dengan expresi yang tinggi di peringkat akhir jangkitan. Oleh kerana kepentingan kepekatan kalsium intrasel dalam kajian otot kekurangan dystrophin, ia diukur dalam tisu otot sihat *M. rosenbergii* dan juga pada beberapa masa yang berbeza selepas jangkitan WSSV. Peningkatan dua kali ganda dalam kepekatan kalsium intraselular dilihat, diikuti oleh penstabilan 24 jam selepas jangkitan itu, menyokong teori hubungan dengan kalsium dystrophin dalam dystrophin otot kekurangan yang sedia ada. Begitu juga, pengimejan Transmission Electron Microscope (TEM) tisu otot *M. rosenbergii* sihat dan dijangkiti WSSV menunjukkan kemerosotan yang ketara di gentian otot dan mitokondria. Beberapa penjelasan yang mungkin untuk perubahan-perubahan yang diperhatikan telah diandaikan dan dibincangkan secara terperinci, termasuk kemungkinan belahan dystrophin oleh proteinase akibat infeksi WSSV.

ACKNOWLEDGEMENTS

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

LIST OF SYMBOLS AND ABBREVIATIONS

The following is a list of abbreviations used in this manuscript, listed in alphabetical order:

3D	3 dimensional
ANOVA	Analysis of variance
BLAST	Basic Local Alignment Search Tool
BMD	Becker muscular dystrophy
CDD	Conserved domain database
cDNA	Complementary Deoxyribonucleic acid
DMD	Duchenne muscular dystrophy
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotides
EF1-alpha	Elongation Factor 1-alpha
ExPASy	Expert Protein Analysis System
FAO	Food and Agriculture Organization
GMQE	Global Model Quality Estimation
IHHNV	Infectious Hypodermal And Haematopoietic Necrosis Virus
kDa	Kilo Dalton
LAMP	Loop-mediated isothermal amplification

Mb	Mega base
MEGA	Molecular Evolutionary Genetics Analysis
MrNV	*Macrobrachium Rosenbergii* Nodavirus
NCBI	National Center for Biotechnology Information
NMR	Nuclear magnetic resonance
OIE	World Organization for Animal Health
PCR	Polymerase Chain Reaction
PSIPRED	Position Specific Iterated Secondary Structure Prediction
QMEAN	Qualitative Model Energy Analysis
qPCR	Quantitative polymerase chain reaction
RNA	Ribonucleic acid
SR	Sarcoplasmic reticulum
SRA	Short Read Archive
TEM	Transmission Electron Microscope
UV	Ultra violet
WSD	White Spot Disease
WSSV	White Spot Syndrome Virus
Φ	Phi
Ψ	Psi

CHAPTER 1 : INTRODUCTION

Macrobrachium rosenbergii, the giant freshwater prawn is one of the most popular species of prawns in Eastern cuisine. The high demand for giant freshwater prawns globally has resulted in a higher market price and propelled its cultivation worldwide, particularly in Asia. Its cultivation depends highly on factors such as environmental conditions, genetic inheritance and other factors.

M. rosenbergii has been shown to be susceptible to several diseases caused by pathogens such as the infectious hypodermal and haematopoietic necrosis virus (IHHNV), White Spot Syndrome Virus (WSSV) and nodavirus (MrNV). These diseases have caused mass mortalities and economic losses in the aquaculture industry. WSSV has been reported as the cause of the appearance of muscle tissue deterioration symptoms in infected *M. rosenbergii* (Corteel et al., 2012; R B Pramod Kiran, 2002). This raised the question of whether WSSV infection in *M. rosenbergii* significantly affects functional muscle genes. Therefore, the dystrophin gene, one of the most crucial genes in the human muscle tissues was selected as the subject of this study.

The dystrophin gene is the longest known gene in humans and plays a crucial role in maintaining muscle function. Null mutations or absence of this gene activates multiple pathophysiological processes leading to muscular dystrophy, a fatal muscle wasting disease. Its presence in several other invertebrates such as *Caenorhabditis elegans, Danio rerio (*zebrafish*)* and *Drosphila melanogaster* has been reported. However, no studies have been conducted on the crustacean counterparts of this gene.

In this study, the *M. rosenbergii* dystrophin gene is identified, and its properties as well as protein structure characterized. Changes in the expression of this gene in response to WSSV infection are quantified to shed some light on the function of the gene in *M. rosenbergii*. Considering the importance of intracellular calcium concentration in relation to the gene of interest, a comparison of the intracellular calcium concentrations between healthy and WSSV-infected prawns at different stages post-infection were also made. The possibility of WSSV induced dystrophin peptide cleavages were also studied to explore this as a probable reason for the changes seen in dystrophin expression. Additionally, Transmission Electron Microscope (TEM) imaging was conducted on healthy and WSSV infected muscle tissues at 24 hours and 48 hours post infection to study changes in the muscle fibers of *M. rosenbergii*.

Several explanations for the observed changes are explored and discussed; connecting all of the observed dystrophin related changes in WSSV infected samples of *M. rosenbergii*. The findings of this study pave the way for further exploration of the role of this gene and its underlying mechanism, particularly in response to viral infections such as WSSV. It is hoped that the information obtained will contribute to the quest to better unravel the dystrophin gene as well as the mechanism of WSSV.

Objectives

This study aims to:

i)   Identify and computationally characterize the dystrophin gene in *M. rosenbergii*.

ii)   Study the differential expression of the dystrophin gene as well as intracellular calcium concentration and muscle fiber changes in WSSV infected *M. rosenbergii*.

CHAPTER 2 : LITERATURE REVIEW

2.1 *Macrobrachium rosenbergii*

The giant freshwater prawn, *Macrobrachium rosenbergii*, has increasingly become an aquaculture species of major commercial importance, with revenue in Asia alone worth more than 1 billion USD annually (FAO, 2009). It is highly popular, particularly in Asian cuisine due to its large size, taste and aesthetic appeal. As a result of the commercial profitability of this species, it is one of the most extensively cultivated freshwater prawn species globally (M. New, 2002).

The annual yield of freshwater prawns has been steadily increasing over the last decades. Its annual global yield exceeded 444 000 tons worldwide and was valued at 2.2 billion US Dollars in the year 2009. (M. B. New & Nair, 2012). Farmed *M. rosenbergii* makes up 51.7 % of this total production. Although freshwater prawns are also farmed on other continents, Asia is a major industry player in its production, dominating the export markets (FAO, 2011).

However, the upward trend observed in the production of *M. rosenbergii* is not reflected by the production output seen in Malaysian freshwater farms over the past years. This is due to a variety of factors including the fragmented nature of the local freshwater prawn industry, water quality and other technical issues. The main cause of this decline is believed to be the slow growth rate of the species, often further hampered by infections and diseases that result in a large number of mortalities.

2.2 White Spot Syndrome Virus (WSSV)

One of the major diseases that have been recorded as a cause of massive mortalities, and subsequently caused large drops in aquaculture production, is the White Spot Disease (WSD). This disease is caused by the White Spot Syndrome Virus (WSSV), a non-occluded baculovirus capable of infecting a wide range of decapods (C.-F. Lo, S-E Peng, Y-S Chang and G-H Kou, 2005). This virus was first reported more than two decades ago, during a disease outbreak among *Penaeus japonicus* in the early 90s in Southern China. This disease was primarily spread geographically within the same species, before being transmitted to other aquaculture species. The spread of this virus occurred very widely and rapidly within a short duration of time due the popularity of transporting live aquaculture organisms, both for culturing and cuisine purposes. Since its initial discovery, this virus has been proven pathogenic towards a broad variety of hosts ranging from crustaceans to polychaete worms, while its symptoms revolved around muscular weaknesses and deterioration (Bateman et al., 2012; Stentiford, Bonami, & Alday-Sanz, 2009; Vijayan et al., 2005).

Symptoms of WSSV infection in *M. rosenbergii* include visible weakness and lethargy, opaqueness of muscle, muscle necrosis and discoloration of muscle tissues. Infected juvenile samples typically begin showing symptoms of infection around 24 hours post exposure to the virus, while symptoms have been shown to escalate in severity at 48 hours post infection (Corteel et al., 2012). It infects virtually all the tissues of the organism, although some differences in viral prevalence at different infection stages were found in different tissue types (Kou, Peng, Chiu, & Lo, 1998). Typically the stomach, gills, cuticular epidermis and the connective tissue of the hepatopancreas show histological signs of infection at the early infection stages, while the lymphoid organ, antennal gland, muscle

5

tissue, hematopoietic tissue, heart, hindgut and parts of the midgut are infected at later stages. These signs of WSSV infection are similar in most reported hosts of the virus, with infections occurring in all tissues of mesodermal and ectodermal origin.

WSSV is currently listed as a notifiable disease by the Office of International Epiootic (OIE) Aquatic Animal Health Code. Due to the necessity for prompt detection and isolation of WSSV infected organisms in the aquaculture farming industry, multiple detection protocols have been developed to screen suspected organisms for this disease. A nested Polymerase Chain Reaction (PCR) protocol (C. F. Lo et al., 1996) has been recognized by the OIE Manual of Diagnostic Tests for Aquatic Animals as the reference standard for WSSV detection. Since the development of this PCR protocol in 1996, advances in molecular technology has ensued in several reports highlighting the drawbacks of this method. In particular, the annealing temperature used in this PCR protocol has raised concerns regarding the assay specificity (Claydon, Cullen, & Owens, 2004). Although this nested PCR protocol is reported to be the most sensitive (Sritunyalucksana, 2006), there are concerns on the production of false positive results, particularly in the second nested step (Nunan & Lightner, 2011). Therefore, for the purpose of WSSV detection in the present study, a faster and more specific one step WSSV detection PCR protocol based on modifications to the nested probes employed in the OIE assay was used (Nunan & Lightner, 2011).

WSSV infection has also been shown to alter the transcriptional expression of numerous genes in infected aquatic organisms, including the down regulation of kinase1 (F. H. Yuan et al., 2016), upregulation of Calreticulin (Duan et al., 2014) and the expression of Peroxiredoxins in response to WSSV infection (Arockiaraj et al., 2012). In order to better

understand the underlying relationship between this virus and changes in transcriptional gene expression, approaches to identify and quantify its viral load changes have been developed by several studies (Durand & Lightner, 2002; Mekata et al., 2009; Nathiga Nambi et al., 2012).

2.3 Viral quantification

A very effective and precise method to numerically quantify the progression of WSSV is through quantitative Polymerase Chain Reaction (qPCR) of the viral genome. One of the earliest reported methods of WSSV quantification is through florescent qPCR assay. This report of amplification of the WSSV target sequence in infected crayfish, *Cambarus clarkia* was developed for use in screening monoclonal antibody pools (L. Yuan et al., 2007). It then led to the introduction of a simpler, accelerated, loop-mediated isothermal real-time amplification (LAMP) assay for the quantification of WSSV in large samples (Mekata et al., 2009). The WSSV quantification method used in the present study is based on a qPCR protocol targeting more recent fragments of the virus (Mendoza-Cano & Sanchez-Paz, 2013). This quantification utilizes probes targeting the VP28 region of WSSV, an essential region for the survival and propagation of the infection. This highly sensitive assay also costs much less compared to other established methods. The internal control gene, elongation factor-1 alpha (EF1-alpha) was used as a reference gene for the purpose of relative expression calculations in the present study (Dhar, Bowers, Licon, Veazey, & Read, 2009).

WSSV was selected as the pathogen of choice for the present study of the dystrophin gene due to its prevalent and observable effects on the muscle tissues of infected prawns. As dystrophin is one of the most important genes in muscle tissues, studying its relationship

7

with a viral infection that greatly affects the muscle tissue is highly intriguing. This also poses an opportunity for discoveries on the underlying mechanism of WSSV infection, and potentially generates interest in developing the use of *M. rosenbergii* in the study of dystrophin.

2.4 The dystrophin gene

The dystrophin gene is a large and complex gene expressed in muscle tissues. It is located on the X chromosome in multiple isoforms and is essential to maintain the structural integrity of muscle fibers (Francesco Muntoni, Torelli, & Ferlini, 2003). In humans, it spans a length of 2.4 Mb and has a complex structure that is essential to maintain cytoskeleton structure and normal muscle function. This sequence produces a 427 kDa rod-shaped protein that links the cytoskeletal actin and a group of dystrophin associated proteins consisting of dystroglycan and sarcoglycan. Seven different isoforms of the human dystrophin are transcribed from its promoters at the 5' end of the gene. The isoforms vary in size depending on the transcription location. Distantly located promoters expressed in the retina, brain and Schwann cells produce smaller isoforms compared to other transcription sites in the body (Blake, Weir, Newey, & Davies, 2002). Although the precise mechanism of the dystrophin gene is still unclear, collectively with its associated proteins, the functional role of dystrophin is believed to be closely involved in muscle contraction within the cytoskeleton (Rybakova, Patel, & Ervasti, 2000).
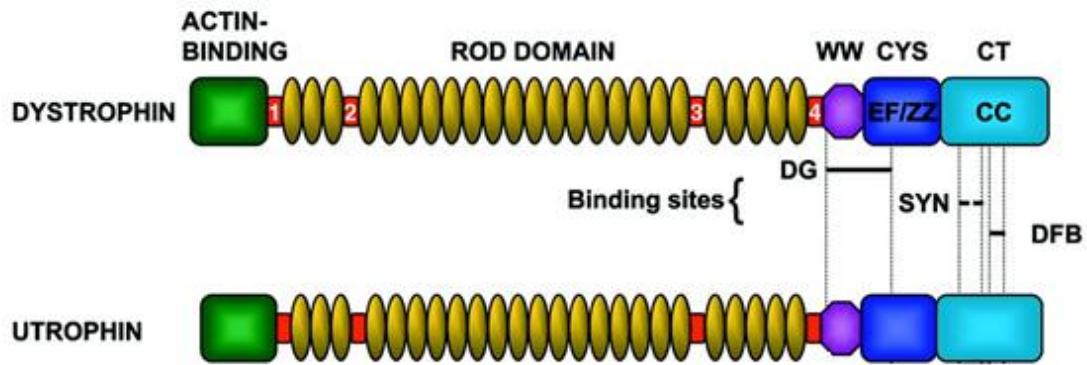
Figure 2.1: Schematic diagram showing the organization of the human dystrophin gene. Four dystrophin regions are shown; the actin binding region, the rod domain region, ~~and~~ the cysteine rich region and the C terminal. The protein domains in the cysteine rich ~~domain~~ region are identified as the WW, ZZ and EF hand domains. The structure of utrophin~~,~~ is shown parallel as comparison.

Figure adapted from (Blake et al., 2002).

These characteristic dystrophin protein domains have been identified and studied in depth by various investigations, particularly in the human dystrophin gene. The cysteine rich region of dystrophin contains the four protein domains, the WW domain, two EF hand domains and the ZZ domain. In addition, the spectrin region of the protein has also been studied in multiple vertebrate and invertebrate homologues of the gene. The WW and EF hand domains of dystrophin directly bind to dystroglycan (Chung & Campanelli, 1999), while the ZZ domains play an indispensable role in this binding of muscle fibres to the dystroglycan (Hnia et al., 2007). The importance of the role of the ZZ domain is reiterated by its additional binding to the EF hand domain, reinforcing the binding between dystroglycan and other dystrophin associated proteins (Ishikawa-Sakurai, Yoshida, Imamura, Davies, & Ozawa, 2004; Vulin et al., 2014).

2.5 Muscular dystrophy

A large number of mutations are possible in the human dystrophin gene, due to its extremely large size. This size causes a complex mutational spectrum (White & den Dunnen, 2006) that leads to a severe muscle wasting disease, namely muscular dystrophy. This disorder is reported to affect as many as 1 in 3500 newborn male babies every year (Finsterer & Stollberger, 2003). Progressive muscle degeneration due to this genetic disorder results in severe pain and eventually death.

The two variations of this disorder, the Duchenne and Becker muscular dystrophies (DMD and DMD) can be distinguished by the appearance of muscular dystrophy symptoms. DMD symptoms typically occur by the age of 12 years old, while BMD symptoms appears after the age of 16 (Mehler, 2000). Approximately 65% of DMD mutations and 85% of BMD mutations is attributed to large deletions in the dystrophin gene. Meanwhile, the remaining mutations are due to slicing, insertion and point mutations (Mendell et al., 2001; Prior & Bridgeman, 2005) As a result of these mutations, the dystrophin gene is unable to fully function in muscle tissues, resulting in the observed symptoms of muscular dystrophy.

Most muscular dystrophy patients die around their twenties of respiratory failure due to lung degradation. Patients also tend to develop cardiac hypertrophy as a result of severe myocardial fibrosis. The quality of life of muscular dystrophy patients are greatly reduced, with the majority of them paralysed and wheelchair bound before reaching adulthood (Blake et al., 2002). The biggest obstacle in treating this disorder is its vast size and the expression of its variants in all types of muscle tissues, including the brain (F. Muntoni & Wells, 2007). Numerous approaches are being developed to better understand this gene and

possible treatment approaches by studying not only the human dystrophin gene, but also in other related dystrophin variants.

2.6 Invertebrate dystrophin

In addition to the paramount study of the dystrophin gene in mammalian skeletal muscle, it has also been investigated in several invertebrates primarily due to the relative ease of studying its function in simpler organisms (Chamberlain & Benian, 2000). Among the invertebrate dystrophin sequences studied, an extraordinary degree of evolutionary sequence conservation was observed throughout the animal kingdom (Roberts & Bobrow, 1998). Although the invertebrate dystrophin sequences were not strictly orthologous to vertebrate dystrophin genes, *cephalochordate amphioxus*, starfish, scallop, *Drosophila melanogaster* and *Caenorhabditis elegans* dystrophin sequences were shown to share a large number of similar characteristics. This discovery is remarkable, especially taking into consideration that the organisms compared are from different phylums (Wang, Pansky, Venuti, Yaffe, & Nudel, 1998). This has generated interest in the scientific community to the potential use of invertebrates as a model organism in studying the dystrophin gene.

In depth investigations on characteristics of invertebrate dystrophin has been conducted almost entirely in *C. elegans* and *D. melanogaster*. In *C. elegans*, mutation in the dystrophin gene causes locomotive dysfunction and progressive muscle degeneration in addition to impaired egg laying abilities (Gieseler, Grisoni, & Segalat, 2000). Reduced muscular capabilities were demonstrated by dystrophin deficient *C. elegans* through uncoordinated movements by moving slowly and leaving irregular tracks on the observation plates. Whereas, mutations in *D. melanogaster* dystrophin result in disruption

of glutamatergic transmissions, progressive muscular degeneration and defects in its wing vein structure (Van der Plas et al., 2007).

Multiple isoforms of dystrophin have also been identified in the sea urchin, of which one shorter transcript show high similarity to vertebrate dystrophin (Wang et al., 1998). It was also shown to contain three major functional domains of dystrophin, namely the EF hand, ZZ and spectrin domains. However, no mention of the WW dystrophin domain in this species was found. Meanwhile, *Pontobdella muricata* (leech) dystrophin protein has been compared to vertebrate dystrophin protein through the specificity of antibodies determined in vertebrates and crude protein extracts of *P. muricata.* In-depth immunoflourescene and immune electron studies of this *P. muricata* dystrophin further reiterated the similarities between vertebrate and invertebrate dystrophin (Royuela et al., 1999). These comparative reports of invertebrate dystrophin homologues highlight the potential of invertebrates in studying the mechanism of the dystrophin gene as well as its use as a model organism for the study of dystrophin deficiency.

2.7 Protein structure prediction and evaluation

Protein structure prediction is an important aspect of bioinformatics and theoretical chemistry that constructs the three-dimensional (3D) structure of a protein from its amino acid sequence. The most commonly used approach in protein structure prediction is comparative or homology modeling which utilizes previously solved protein structures as templates. Due to the limited number of different protein families occurring in nature, the similarity in structure between homologous proteins can be inferred from sequence similarity (Chothia, 1992). This is applied in homology modeling, whereby a 3D protein model of the target amino acid sequence is built by identifying and aligning the target

12

amino acid sequence with a template protein of known structure. The structural frameworks of the aligned regions are then copied or built, satisfying the spatial restraints of the template used (Zhang, 2008).

The present study utilizes the SWISS-MODEL workspace, a web based platform to conduct comparative homology protein structure modeling (Arnold, Bordoli, Kopp, & Schwede, 2006). This workspace allows highly automated modeling procedures with minimal user intervention through the identification of highly similar structural templates. It also conducts preliminary assessments of the protein model constructed, including GMQE and QMEAN composite scores (Benkert, Tosatto, & Schomburg, 2008). This comparative homology modeling method in protein prediction has an 80% accuracy (Pirovano & Heringa, 2010), allowing its use as an important feature of structural genomics to improve protein recognition.

Due to the complex nature of the algorithms used to computationally construct protein models, it is necessary to verify and asses the models produced. As the quality of the protein structure model produced determines it possible applications, estimating its accuracy is absolutely crucial (Baker & Sali, 2001). With the increased usage of computational protein translation and modeling, several checks and assessment methods have been established for this purpose. The present study utilizes two such assessments to evaluate the *M. rosenbergii* dystrophin protein model produced, namely the "WHAT_IF" series of protein evaluations (Vriend, 1990) and the Ramachandran Plot (Ramachandran, Ramakrishnan, & Sasisekharan, 1963). These methods were selected due to their dependability, reproducibility and ease of use. Although advances in computational technology have greatly affected the approaches used in protein modeling, traditional

13

structural biology methods are still widely accepted to be relevant, as implied by the far more prevalent citations of older publications in recent years despite introductions of newer technologies (Chandonia & Brenner, 2006).

The Ramachandran Plot visualizes the dihedral bond angles Phi, Φ and Psi, Ψ of a protein backbone, providing a conformity assessment of computationally produced protein models (J. S. Richardson & David, 2012). The Phi, Φ and Psi, Ψ bond angles are the rotational configuration of bonds between adjacent amino acids in a polypeptide chain. The angle of the bond between nitrogen and the carbon atoms is Phi, Φ and the angle between carbon and carbonyl atoms is Psi, Ψ (Lovell et al., 2003). Combinations of Phi and Psi values are constrained to specific areas on the plot due to steric hindrances within the protein structure. This result in plots for ordered structures, such as alpha helix and beta strands to be clustered within specific regions, commonly presented as contours on the diagram (Ho, Thomas, & Brasseur, 2003; J. Richardson, Keedy, Richardson, Bansal, & Srinivasan, 2013).

This methodology was initially developed to identify permissible values of phi and psi for amino acids and predict theoretically possible conformations of the polypeptide backbone (Zhou, O'Hern, & Regan, 2011). Over the years, numerous redrawing of the plot based on experimental evidences by various researchers confirmed the consistency of the original plot. Recent uses of this plot include the verification of 3D protein structures produced by crystallography, NMR spectroscopy and computer modeling (Kalinowska, Alejster, Salapa, Baster, & Roterman, 2013; Roterman, Konieczny, Banach, & Jurkowski, 2011). As these approaches rely heavily on complicated software algorithms to produce the final protein structure, a Ramachandaran plot is used to identify anomalous phi and psi values, thus

identifying any mistakes that might have occurred in the computationally determined atomic positions within the structure.
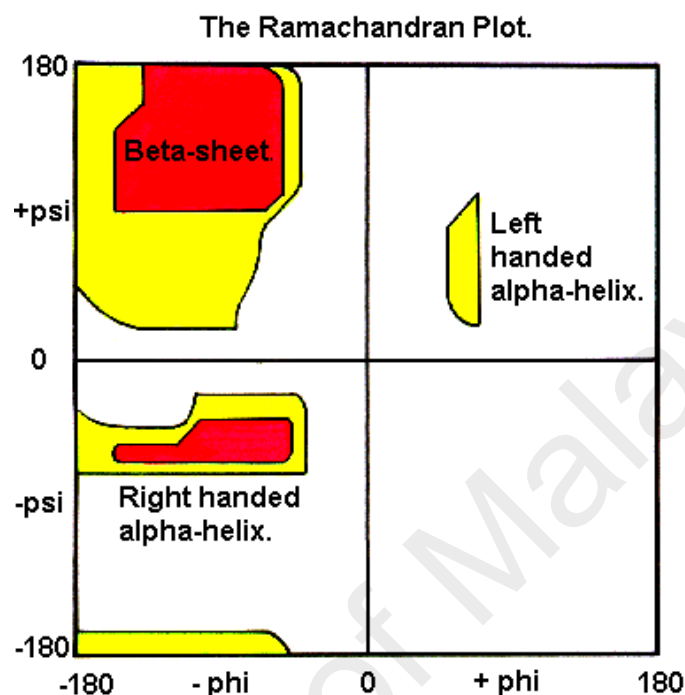
Figure 2.2: A generic map of the Ramachandran Plot showing the accepted locations of Beta-sheet and the two alpha-sheetresidues. The x-axis shows the value of Phi, Φ, while the y-axis shows the value of Psi, Ψ.

Figure adapted from the School of Natural Sciences, Birkbeck College , University of London academic website.

The "WHAT_IF" protein check used to verify the *M. rosenbergii* dystrophin protein structure and Ramachandran plot produced is a freely available web program that conducts a series of comprehensive checks on a predicted protein model, and reports the results of each check in numerical scores (Vriend, 1990). Among assessments commonly conducted on the "WHAT_IF" program are evaluations of the Ramachandran plot appearance, χ_1/χ_2 rotamer normality, backbone conformation and packing quality of the protein structure (Wilson, 1998). The Ramachandran plot evaluation score expresses how well the backbone conformations of tall residues correspond to the allowed contours in the plot. Similarly, the

15

χ_1/χ_2 rotamer normality assesses how well each rotamer fits into the common areas of the χ_1/χ_2 correlation plot. Meanwhile, the packing quality of a protein structure is a representation of a comparison of the average packing environment for all residues in recognized accepted packing files. These checks on the protein structure model evaluate and identify possible errors in the computational construction of the 3D protein structure models through comparative homology modeling.

2.8 Protease activity

In the present study, possible peptide cleavage sites in *M. rosenbergii* are identified and the possibility of cleavage occurrence at each site is calculated to evaluate the possibility of WSSV induced protein cleavage as a reason for the observed changes. Peptide cleavage is the separation of a large protein sequence into smaller groups of amino acid with the help of protein digesting enzymes known as proteases (Creighton, 1993). Protein cleavage typically occurs in muscle tissues to prevent accumulation of unwanted proteins, regulate cellular or physiological processes. Proteases can be broadly divided into two groups based on their mode of action; specific proteases that catalyzes highly specific protein cleavage and non-specific proteases that are involved in a wider range of protein hydrolysis activity (Fuchs et al., 2013). However, disorder of protease regulation in tissues can be damaging and result in diseases.
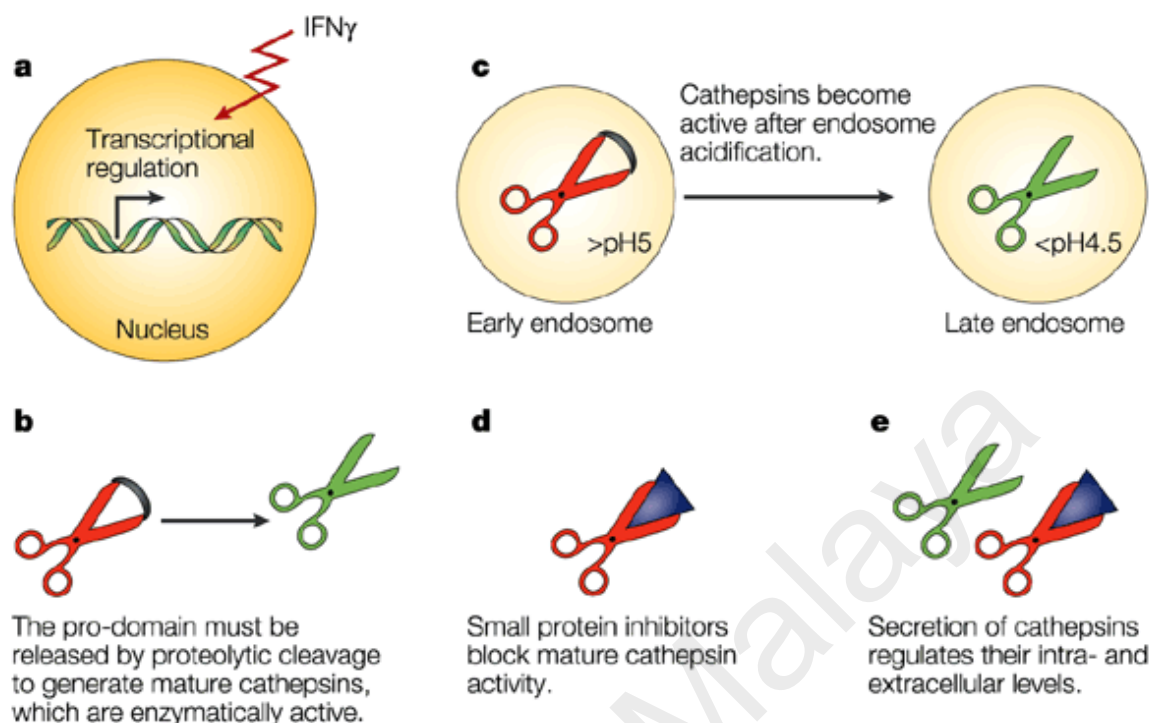
Figure 2.3: A schematic diagram showing the regulation of the cathepsin protease at several different stages: (a) The mRNA level, (b) Enzyme activation, (c) mature cathepsin generation, (d) protein inhibition and (e) cathepsin secretion.

(Figure adapted from (Honey & Rudensky, 2003)

The expression of proteases are greatly affected by the infection of virus or diseases due to their role in immune response against infections (FL Garcia-Carreño, 2014). The importance of proteases in the immune response of crustaceans such as *M. rosenbergii* is greatly amplified by the lack of a conventional adaptive immune system in crustaceans (Welsh, Selin, & Szomolanyi-Tsuda, 2004). Primarily, proteases in the Arthropoda immune system function to regulate peptide synthesis and coagulation (Gorman, Andreeva, & Paskewitz, 2000). Protease up-regulation upon infection, particularly of serine proteases, has been demonstrated in various crustaceans including the oriental shrimp *Fenneropenaeus chinensis,* the tiger shrimp *Penaeus monodon,* the swimming crab *Scylla serrata,* the white leg shrimp *Litopenaeus vannamei* and the freshwater red claw crayfish

17

Cherax quadricarinatus (Jimenez-Vega, Vargas-Albores, & Soderhall, 2005; Liu et al., 2007; Shi, Ren, Zhao, & Wang, 2009; Supungul et al., 2002; Xue, Yang, & Sun, 2013; Zeng & Lu, 2009).

One the most studied protease in the immune response system of invertebrates is Chymotrypsin. Chymotrypsin is a serine residue containing protease that cleaves peptide bonds in amino acid sequence which contain an aromatic amino acid on its amine side (Berg JM, 2002). Its up-regulation has been demonstrated in a number of studies quantifying changes in response to immune challenge in crustaceans (Chai, Yu, Zhao, Zhu, & Wang, 2010; Nayak, Ajay, Ramaiah, Meena, & Sreepada, 2011; Shi et al., 2009; Vaseeharan, Lin, Ko, & Chen, 2006). Due to these reports of Chymotrypsin up-regulation in response to WSSV infection, as well as the high occurrence of possible Chymotrypsin cleavage sites in the studied sequence, the activity of this peptide was investigated further in the present study to gain insights on its possible relationship to the changes in dystrophin expression.

In the present study, computational calculation of cleavage probabilities were conducted to identify potential peptide cleavage sites on the gene of interest through the ExPASy Peptide Cutter tool (Gasteiger et al., 2005). This platform searches through the protein sequence for potential protease cleavage sites for a list of selected enzymes, based on the enzyme properties registered in its database. A more sophisticated model is also available on the same platform to calculate the frequency of cleavage of specific proteases based on probability analysis, taking into consideration cleavage conditions and proteases specificity. This model was utilized in the present study to investigate the probability of occurrence of Chymotrypsin cleavage at each of its identified cleavage sites.

2.9 Intracellular calcium concentration

Since the earliest reports of calcium accumulation in dystrophin-deficient muscle almost four decades ago (Bodensteiner & Engel, 1978), abundant investigations have been made into the role and significance of intracellular calcium concentrations in relation to muscle function and muscular dystrophy (Allen, Gervasio, Yeung, & Whitehead, 2010; Campos et al., 2011; Goonasekera et al., 2014; Mallouk, Jacquemond, & Allard, 2000; Shin, Tajrishi, Ogura, & Kumar, 2013). Large increases in intracellular calcium concentration as well as higher levels of calcium in dystrophin deficient skeletal muscle sarcoplasmic reticulum (SR) were demonstrated in several reports of studies on the myofibres of *mdx* mice (Goonasekera et al., 2014; Mallouk et al., 2000). Total calcium content of dystrophy affected muscle have also been shown to be elevated, even at the very early stages of the disorder (Bertorini et al., 1984). Confirmation that the lack of dystrophin is responsible for this observed increase in calcium concentration comes from the restoration of dystrophin in dystrophin-deficient mutants, resulting in subsequent elimination of the symptom (Cox et al., 1993).

Although the relationship between dystrophin deficiency and increased intracellular calcium concentration is well accepted, several different hypotheses have been offered to explain this increase. The most popular hypothesis predicts that the increase is due to a calcium leakage or influx into the cells through damaged membranes (Bodensteiner & Engel, 1978; Petrof, Shrager, Stedman, Kelly, & Sweeney, 1993). As it is generally accepted that dystrophin and its associated protein complexes have a role in stabilizing the cell membrane, it is predicted that in the absence of dystrophin, membrane damage is more frequent (Miyake & McNeil, 2003). The attribution of this increase in intracellular calcium

concentration to a muscular dystrophy induced disturbance in calcium homeostasis is supported by numerous other reports (Mariol & Segalat, 2001; Ruegg & Gillis, 1999).
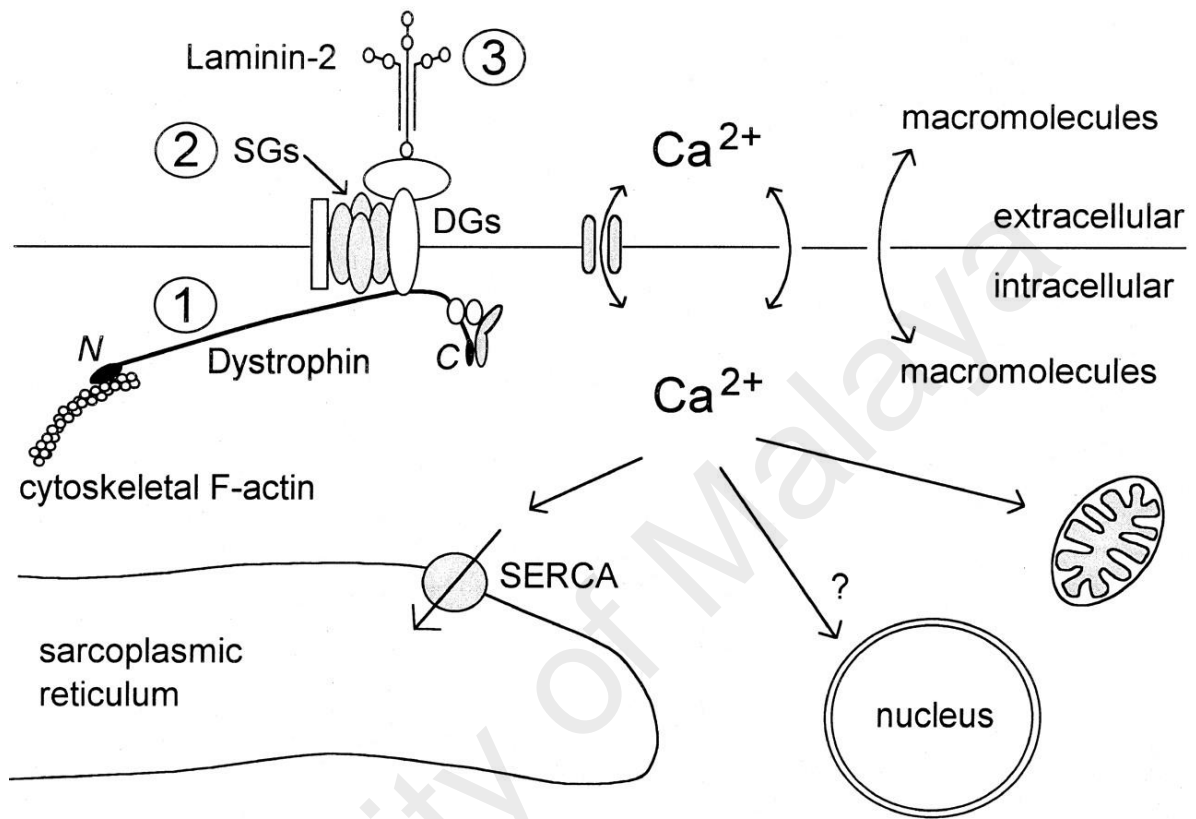
Figure 2.4: Schematic diagram illustrating trans-membrane calcium ion action in healthy muscle fiber cells. Dystrophin (1) is shown linking the cytoskeleton to laminin (3) in the extracellular matrix through the dystrophin associated proteins, dystroglycans and sarcoglycans (2). Calcium ion movement is shown through the calcium channels and distributed to the nucleus, mitochondria and sarcoplasmic reticulum for their organelle specific roles.

Figure adapted from (Berchtold, Brinkmeier, & Muntener, 2000)

Another theory speculates that the absence of dystrophin contributes towards this increase through elevated channel activity, leading to an escalation of intracellular calcium concentration in dystrophin deficient tissues as illustrated in Figure 2.4 (Allen et al., 2010; Berchtold et al., 2000). The absence or mutations of dystrophin in the cytoskeleton effects calcium binding sites such as those found in laminin, distorting calcium handling and

signaling mechanisms in the cytoskeleton. The functional alterations caused by these changes in calcium concentrations are observable in the muscle as part of the pathophysiological conditions of dystrophin deficiency.

Increased calcium presence has been proffered as a possible reason for the muscle weakness observed in dystrophin deficient organisms, mainly due to the activation of calpains at high intracellular calcium concentrations (Verburg, Dutka, & Lamb, 2006). Calpains is a calcium-activated enzyme, that mediates the degradation of individual muscular fiber proteins (Koohmaraie, 1992). It is highly concentrated in the Z-disk region of muscle tissues, where the degradation of muscle tissue begins during muscle wasting (Campos et al., 2011; Ma et al., 2011). Its role in skeletal muscle protein degradation is well documented, both *in vivo* and *in vitro*, when activated by increased calcium concentration (Dahlmann, Kuehn, Reinauer, Kay, & Stauber, 1989; J. Huang & Forsberg, 1998; Sorimachi, Hata, & Ono, 2011).

Due to the impoprtant relationship between intracellular calcium concentration changes and muscle deterioration in dystrophin-deficient tissues, intracellular calcium concentration comparison in WSSV-infected *M. rosenbergii* muscle was conducted in the present study through the use of fluorescent calcium chelators.

2.10 Transmission Electron Microscope (TEM) Imaging

Transmission Electron Microscope (TEM) is an imaging technology that utilizes the transmission of electron beams through an ultra-thin specimen and its interaction to form an image. The application of electron microscopy in biological materials were only made possible by the discovery of the use of osmium tetroxide in sample fixation two decades after the initial introduction of TEM technology (Palade, 1952). Since then, this technology

has also been used to study a variety of muscle tissues in biological specimens in addition to its more common use in material sciences (Delella & Felisbino, 2010; Kan & Nanci, 1988; Komuro, 1999; Polican Ciena et al., 2012; Scothern & Garrod, 2008). Development of these techniques has made the viewing of ultra-structure components of cell, tissue or organs of biological origin possible through modifications of the standard microscopy techniques, taking into consideration the different physical properties of electron microscopy. Similarly, the introduction of Porter Bloom microtome as a technology capable of slicing ultra thin sections of biological tissues (40nm – 60nm histological sections) contributed greatly into the use of TEM in the study of biological tissues (Gordon, 2014).

The ability of TEM to provide higher contrast in biological tissues compared to Scanning Electron Microscopy (SEM) has popularized its use in the study of muscle tissues, notably in relation to dystrophin and muscular dystrophy. The most common use of TEM in the study of dystrophin is in the study of ultra-structural changes in organisms that imitate dystrophin deficiency in muscular dystrophy patients, such as the *mdx* mice (De Aro et al., 2015; Rafael & Brown, 2000). Other reports on the use of TEM in muscle tissue studies include an evaluation of tissues in canine muscular dystrophy models (Lessa et al., 2014), a study on the ultra-structural distribution and relative concentration of iso-enzymes morphologically associated with the M-line of striated muscle (Dankert, Papadi, & Shields, 1992), as well as a mechanical characterization of fully differentiated skeletal muscle fibers (Defranchi et al., 2005).

The use of TEM in viewing dystrophin related muscle morphological changes have been reported in mice models of muscular dystrophy (Goonasekera et al., 2014), whereby clear

disorder in the muscle fiber arrangements were observed in the muscle tissue morphology. The same study also reported swelling and fragmentation of the mitochondria in mouse models that showed signs of muscle degradation, as seen in Figure 2.5 B.
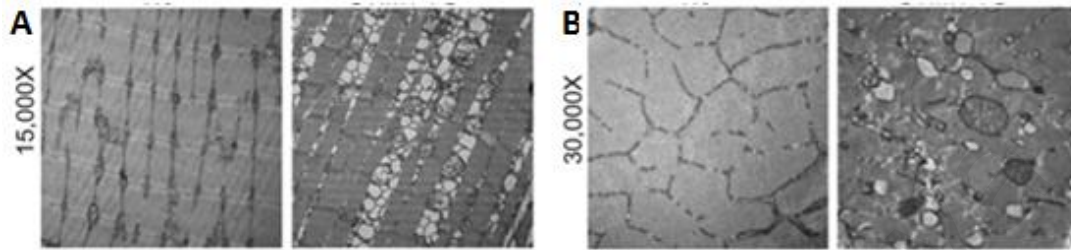
Figure 2.5: TEM imag~~esing~~ of muscle morphology in mouse models of muscular dystrophy. A) The morphological layout of muscle fibers at two different magnifications. B) High magnification TEM images of the mitochondria ~~of~~ in mouse models of muscular dystrophy.

Figure adapted from Goonasekera et al., 2014.

On the other hand, significant changes to the mitochondria due to viral infections have been reported in profile studies of viral mechanisms (Chen et al., 2001). Mitochondria has been reported as a favorite organelle for invading viruses, and many mitochondrial proteins targeted by viruses are relevant to the pathogenesis of the diseases they cause (Anand & Tikoo, 2013). These viral induced effects upon the mitochondria occur on varied timelines, and differ in each infection depending on the host and virus mechanisms. An illustrated example of virus – mitochondria interaction resulting in cell apoptosis is shown in Figure 2.5. Due to this profound effect of mitochondria to viral infections, the organelle was one of the main focus during ultra-structure imaging of WSSV infected *M. rosenbergii* muscle tissues through TEM in the present study.
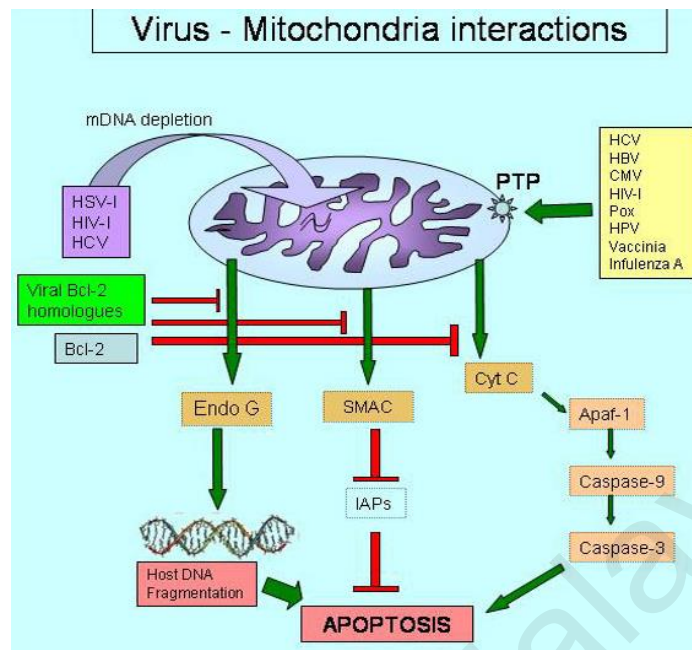
Figure 2.6: A schematic diagram of the cellular events involving the mitochondria following viral infections. Mitochondrial membrane permeabilization is prevented by viral Bcl-2 homologues, resulting in the release of pro-apoptotic factors.

Figure adapted from (Anand & Tikoo, 2013)

2.11 Conclusion of Literature Review

All of these discoveries contribute tremendously towards a better understanding of the dystrophin gene, and this study aspires to pave the way for similar dystrophin studies in *M. rosenbergii*. The present study characterizes properties of the dystrophin gene in *M. rosenbergii* and its expression in response to an immune challenge by WSSV. Several other related changes, namely intracellular calcium concentration, viral load and TEM observation of WSSV infected *M. rosenbergii* muscle tissues were also carried out due to their relationship to the dystrophin gene, as elaborated in the literature review above. The relationship between all these parameters is hoped to shed some light on the significance of the gene in *M. rosenbergii* and the WSSV infection pathway. To date, there has been no previous study on the presence of this gene in crustaceans.

CHAPTER 3 : MATERIALD AND METHODS

3.1 Characterization of *Macrobrachium rosenbergii* dystrophin (MrDys)

3.1.1 Identification of *M. rosenbergii* dystrophin nucleotide sequence

A putative *Macrobrachium rosenbergii* dystrophin nucleotide sequences was retrieved from the Short Read Archive (SRA) of the National Center for Biotechnology Information (NCBI) under the accession number SRR1424572 and verified through BLAST prior to validation.

3.1.2 Total RNA isolation

Total RNA of muscle tissue was isolated from 50 mg of muscle tissue per sample using the TRIZOL reagent (Life Technologies, Carlsbad, CA). In sterile, RNAse-free conditions, 1 ml of trizol was placed into a microcentrifuge tube. Liquid nitrogen was used to freeze and grind the piece of muscle tissue to powder with a sterile mortar and pestle. The powdered tissue was added to 1 ml trizol solution in a microcentrifuge tube and shaken to mix thoroughly. Next, 200 μl of chloroform was added and the tube vigorously shaken before being incubated at room temperature for 10 minutes. The tube was then centrifuged at 12000g for 15 minutes at 4 °C and the aqueous layer transferred to a new tube.

The separated aqueous layer was added to 500 μl of ice-cold isopropanol and vortexed for 10 seconds. The tube was then centrifuged at 12000g for 8 mins at 4 °C and the supernatant discarded. The pellets formed in the microcentrifuge tube were air dried and dissolved with 50 μl of nuclease free water.
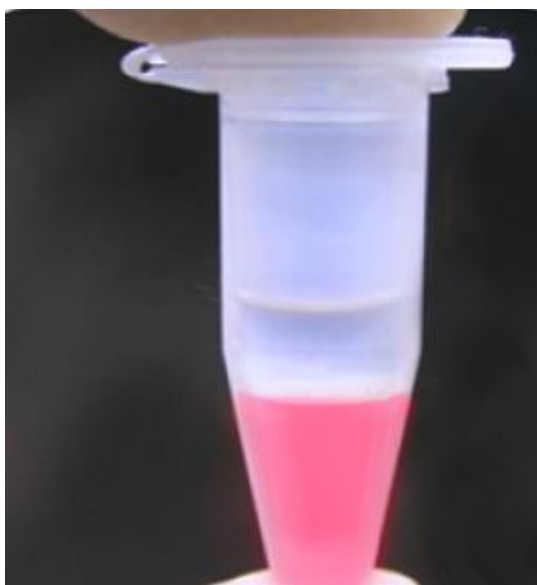
Figure 3.1: A microcentrifuge tube containing the 3 layers formed ~~upon~~ by centrifugation with chloroform. The coloured layer at the bottom is the organic layer, while the thin white layer formed in the middle is a protein layer. The extracted RNA is retrieved from the ~~top~~ clear aqueous layer with extreme care to avoid shaking or contaminating the tube contents.

DNAse treatment was carried out on the extracted RNA (RNAse-free DNAse, Promega) to eliminate DNA contamination prior to reverse transcription to cDNA. The DNAse digestion reaction was set up as shown in Table 3.1.

Table 3.1: Components of the DNAse digestion treatment of the extracted RNA of M. rosenbergii tissues.

Reagent	Volume (µl)
TE buffer	6
RNase-Free DNase 10X Reaction Buffer	1
RNase-Free DNase	1
RNA Nuclease-free water	2
Total	**10**

This DNAse digestion reaction was incubated at 37 $^{\circ}$C for 30 minutes. At the end of the incubation period, 1 µl of DNase Stop Solution was added to terminate the reaction. Next, the mixture was incubated at 65 $^{\circ}$C for 10 minutes to inactivate the activity of the DNase enzyme. The quality and purity of the isolated, DNAse treated RNA was checked on a NanoDrop 2000 Spectrometer (Thermo Scientific, USA). Only RNA samples with an A_{260}/A_{280} ratio of 2.00 was used to synthesize the cDNA samples for qPCR quantification through reverse transcription.

3.1.3 Reverse Transcription of RNA to cDNA

Reverse transcription of the extracted RNA to cDNA was carried out using GoScript reverse transcriptase (Promega, USA) in RNAse free conditions. For each 20 µl of reverse transcriptase reaction, 4 µl of RNA template was used, followed by 1 µl of oligo dt primer. This mixture was incubated at 70 $^{\circ}$C for 5 minutes and immediately chilled at 4 $^{\circ}$C for 5 minutes before being included in the reverse transcription reaction. The reaction was prepared as shown in Table 3.2 and the PCR thermocycler was programmed to the cycles shown in Table 3.3.

The concentration and purity of the cDNA produced was checked using a NanoDrop 2000 Spectrometer (Thermo Scientific, USA) prior to use in quantification of gene expression. Purity of the cDNA was estimated from the ratio of absorbance at 260nm and 280nm. Samples cDNA with an A_{260}/A_{280} ratio of 1.80 was deemed acceptable for use in real-time PCR quantification of the target gene.

Table 3.2: Chemicals and reagents involved in reverse transcriptase of *M. rosenbergii* RNA to cDNA.

Reagent	Volume per reaction (µl)
Nuclease free water (NFC)	6.1
5x reaction buffer	4.0
MgCl2	2.4
dNTP mix (10 mM each dNTP)	1.0
Ribonuclease inhibitor	0.5
reverse transcriptase	1.0
RNA template mixture	5.0
Total	15.0

Table 3.3: Temperature properties of the thermal cycle used for reverse transcriptase of *M. rosenbergii* RNA to cDNA.

Process	Temperature (^{o}C)	Time (minutes)
Annealing	25	5
Extension	42	60
Heat inactivation	70	15

3.1.4 Sequence validation through PCR

Validation of the identified *M. rosenbergii* dystrophin sequence was conducted through Polymerase Chain Reaction (PCR). Probes for the amplification of the *M. rosenbergii* dystrophin sequence was designed on the primer 3 plus software (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi), targeting the *M. rosenbergii* dystrophin transcript. Various parameters including annealing temperature, hair pin loop and GC ratio were taken into consideration in designing the best primers for use in this PCR. Optimizations of the primers designed were conducted by varying the primer concentration, annealing temperature and number of cycles of the PCR reaction. The PCR products obtained were run on 1% agarose gel electrophoresis as described in the following section. The chemicals and PCR regents (GoTaq ® Flexi, Promega, USA**)** used in the PCR is shown in Table 3.4 while the PCR reaction was carried out in a thermal cycler (BioRad, USA) according to the temperature conditions shown in Table 3.5. Oligonucelotide sequences of the gene specific primers designed, as well as its annealing temperatures are shown in Table 3.6.

Table 3.4: PCR reagents used in the amplification of the identified dystrophin sequence, with their respective concentrations and volumes.

Reagent	Final concentration	Volume per reaction (µl)
MgCl$_2$	2.0 mM	1.50
Buffer	1X	3.00
dNTPs mix	0.2 mM each dNTP	1.00

Primer F	0.6 mM	0.40
Primer R	0.6 mM	0.40
DNA Polymerase	1.25 u	0.30
Nuclease free water	-	1.40
Purified DNA	50 ng/µl	2.00
Total		**10.00**

Table 3.5: Temperature properties of the thermal cycles used in PCR amplification of the identified dystrophin sequence.

Temperature (^{o}C)	Time	Cycle
95.0	2 m	1
95.0	30 s	30
57.8	30 s	30
72.0	1 s	30
72.0	5 m	1
4.0	--	--

Table 3.6: Gene specific primers designed to amplify *M. rosenbergii* dystrophin. The oligonucleotide sequences are shown from the 5' end to the 3' end.

Primer	Oligonucleotide sequence	Annealing temperature ($^{\circ}$C)
Dys 1 F	5' - AAGCCCAAAAGACTGCCGAT - 3'	58.60
Dys 1 R	5' - CCCAGTGGGTACTCTCCTGA - 3'	58.60
Dys 2 F	5' - AAATTCCGCGAAGGTTTG - 3'	54.40
Dys 2 R	5' - TCGGCAATTAGCCGGAAC - 3'	54.40
Dys 3 F	5' - GTTAGAAAATTCCGCGAAGG - 3'	57.60
Dys 3 R	5' - GCATGGGGTGAGTGATCTT - 3'	57.60

3.1.5 Verification of *M. rosenbergii* dystrophin sequence

PCR product obtained through the methods described above was run on 1% agarose gel electrophoresis. Powdered agarose (Sigma Aldrich Co Ltd, USA) was weighed and added to the required volume of TAE solution to create a 1% agarose solution. The mixture was then heated in a microwave oven and cooled at room temperature prior to the addition of a nucleic acid stain (RedSafe) to aid illumination for band visualization purposes. The gel was allowed to solidify in a casting mould and immersed with TAE solution as a buffer. An equal volume of PCR product (5 µl) was loaded into each well and electrophoresis of the gel conducted at 160 V and 80 A for 45 minutes. The gel was then viewed under an UV-transilluminator (WiseUV WUV, Germany) and the bands excised using a sterile scalpel.

The excised gel was weighed and placed in a sterile 1.5 ml microcentrifuge tube. In accordance to the manufacturer issued protocol for gel clean up of PCR products (Wizard® SV Gel and PCR Clean-Up System, Promega), 10 µl of Membrane Binding Solution was added per 10 mg of gel slice. The microcentrifuge tube was vigorously vortexed and incubated at 65 $^{\circ}$C until the gel was completely dissolved. For the binding of nucleic acid, the provided mini column was inserted into a collection tube. The dissolved gel mixture was then transferred to the mini column assembly, and left to incubate at room temperature for 1 minute.

The column was then centrifuged at 16000 x g for 1 minute. The flowthrough was discarded, and the mini column reinserted into the collection tube. Next, 700 µl of membrane wash solution was added into the tube and centrifuged at 16000 x g for 1 minute. The flowthrough obtained was discarded and the washing step repeated with 500 µl of membrane wash solution before being centrifuged at 16000 x g for 5 minutes. After discarding the flowthrough, the tube was centrifuged again at 16000 x g for 1 minute with the lid open to allow evaporation of any residual ethanol.

The mini column was then transferred to a sterile 1.5 ml microcentrifuge tube. To elute the nucleic acid retrieved from the gel, 50 µl of nuclease free water was added to the mini column. The tube was then centrifuged 16000 x g for 1 minute before the mini column was removed from the microcentrifuge tube and discarded. The nucleic acid flowthrough was stored at -20 $^{\circ}$C and sequenced through Sanger sequencing on an ABI Applied Biosystems 3730XL DNA Analyzer.

3.1.6 Characterization and structural modeling of *M. rosenbergii* dystrophin

Upon obtaining the sequence of the *Macrobrachium rosenbergii* dystrophin sequence, its corresponding amino acid sequence was deduced by computational translation using the SIB ExPASy Bioformatics Resources Portal translate tool (Artimo et al., 2012). This sequence was then verified through a protein BLAST against the NCBI protein database. Physiochemical properties of the amino acid sequence obtained were then studied using the protein analysis tool ProtParam, also on the ExPASy platform (Gasteiger et al., 2005). Properties of two other homologous arthropod dystrophin sequences, from *Drosophila melanogaster* (fruit fly) and *Danio rerio* (zebrafish) dystrophin, were retrieved from the NCBI database and similarly analyzed. The predicted secondary structure of *M. rosenbergii* dystrophin was conducted using the PSIPRED secondary prediction method on the UCL Bloomsbury Centre for Bioinformatics (Buchan, Minneci, Nugent, Bryson, & Jones, 2013).

Protein modelling to obtain the 3D structure for *M. rosenbergii* dystrophin was conducted through the SWISS-MODEL workspace (Bordoli & Schwede, 2012) . This was done by searching the SWISS-MODEL template library for evolutionary related sequences matching the *M. rosenbergii* dystrophin sequence (Biasini et al., 2014). Each of the template's quality was predicted from the features of the target-template alignment and the templates with the highest quality selected for model building. A 3D model was then built based on the target-template alignment and its quality assessed using the QMEAN score (Benkert, Biasini, & Schwede, 2011). The 3D protein model of *M. rosenbergii* dystrophin was validated by constructing a Ramachandran plot (Gopalakrishnan, Sowmiya, Sheik, & Sekar, 2007). The quality of the protein model obtained was further assessed by conducting a "WHAT_IF" protein model check using the Discovery Studio software (BIOVIA, 2015).

3.1.7 Identification and comparison of conserved domains

All of the conserved protein domains on the *M. rosenbergii* dystrophin sequence were identified through a Conserved Domain Database (CDD) search (http://www.ncbi.nlm.nih.gov/cdd/). The properties of each domain were analyzed and its functional role reported. A similar search was also conducted on 11 other dystrophin sequences from different organisms, identified and retrieved from the NCBI database. The origin of these dystrophin sequences as well as their NCBI accession numbers are shown in table 3.7. The conserved protein domains for each sequence were studied for similarities and a comparison graphically studied.

Table 3.7: The origins of the 11 dystrophin transcripts from different organisms retrieved from the NCBI database, shown with their respective accession numbers.

Species name	Common name	NCBI Accession Number
Asteroidea sp	Starfish	X99737.1
Branchiostoma lanceolatum	Lancelet	X99736.1
Caenorhabditis elegans	Round worm	NM_001306315.1
Canis lupus	Dog	NM_001003343.1
Danio Rerio	Zebrafish	NM_131785.1
Drosophila melanogaster	Fruit Fly	X99757.1
Gallus gallus	Chicken	X13369.1
Homo sapiens	Human	M18533.1
Pectinidae sp	Scallop	X99738.1

Scyliorhinus canicula	Dogfish	X99702.1
Xenopus laevis	African Frog	X99700.1

3.1.8 Homology Analysis

A multiple sequence alignment of these dystrophin nucleotide sequences was conducted using the MEGA 6 software (Tamura, Stecher, Peterson, Filipski, & Kumar, 2013). A phylogenetic analysis of the 12 dystrophin nucleotide sequences (including *M. rosenbergii* dystrophin) was conducted through the Maximum Likelihood method on the same platform. The phylogenetic tree obtained was drawn to scale, with number of substitutions per site used to measure branch lengths.

Multiple sequence alignment of the amino acid sequences for *D.rerio*, *D. melanogaster* and *M. rosenbergii* dystrophin were conducted through Clustal Omega (http://www.ebi.ac.uk/Tools/psa/) on the EMBL-EBI server. A SIM alignment (http://web.expasy.org/sim/) was also conducted on the sequences to compare the locations of sequence similarities in all 3 protein sequences on the ExPASy Bioinformatics platform. The alignments obtained were graphically shown using the LALNVIEW program (Duret, Gasteiger, & Perriere, 1996). This amino acid sequence alignment results are shown both in the sequence form and graphically to demonstrate the similarities.

3.1.9 Protease Cleavage Analysis

The *M. rosenbergii* dystrophin sequence was analyzed for possible protease cleavage sites using the Peptide Cutter tool (http://web.expasy.org/peptide_cutter/). Upon identification of the possible occurrences protease cleavages and potential cleavage locations, in depth analysis of the probability of cleavage by the protease Chymotrypsin was carried out on the same platform by selecting a more sophisticated analysis parameter.

3.2 Immune challenge with White Spot Syndrome Virus (WSSV)

3.2.1 Experimental animals

Specific-pathogen free juvenile *M. rosenbergii* samples were obtained from a freshwater prawn aquaculture farm in the state of Kedah, Peninsular Malaysia. The juvenile prawn samples weighing between 5 g to 8 g each were purchased and transported to the laboratory. All prawn samples were inspected for symptoms of common diseases and acclimatized for 7 days in controlled laboratory conditions prior to the start of experiment. The prawns were placed in 300 L flat-bottomed tanks filled with filtered and aerated freshwater, and the temperature maintained at 27°C throughput the study. A total of 10 juvenile prawns were placed in each tank. Aeration was provided by the means of an electrical air pump. The prawns were fed twice a day with commercially available prawn feed (Red Bee Aquarium Shrimp Feed, China).

Figure 3.2: A juvenile M. rosenbergii sample prior to WSSV injection. A plastic ruler is included in the image taken as reference for its size.

3.2.2 Isolation of Genomic DNA

DNA extraction of muscle tissue was carried out using a total DNA extraction kit (DNeasy Blood and Tissue kit, Qiagen, Malaysia) according to manufacturer's instructions. Muscle tissues of the prawns were dissected; 25 mg was weighed and placed in a 1.5 ml microcentrifuge tube. Into this tube, 180 µl ATL buffer was added, followed by 20 µl of proteinase K. The tube was vortexed vigorously and incubated at 56 °C for 8 hours with periodical vortexing until the tissue had completely lysed. The homogenate produced was added with 200 µl of AL buffer and vortexed thoroughly before the addition of 200 µl of pure ethanol.

The mixture was vortexed again and pipetted into the DNeasy Mini spin-column provided. It was placed into a 2ml collection tube and centrifuged at 6000 x g for 1 minute. The flow-through was discarded and the mini column assembly centrifuged again at 6000 x g for 1

37

minute after the addition of 500 µl AW1 buffer. This step was repeated for centrifugation at 20000 x g for 3 minutes before transferring the column to a new sterile 1.5 ml microcentrifuge tube. The DNA extracted was eluted by the addition of 200 µl elution buffer, and left to incubate at room temperature for 1 minute prior to a final centrifugation of at 6000 x g for 1 minute.

The concentration and purity of the isolated DNA was checked using a NanoDrop 2000 Spectrometer (Thermo Scientific, USA). DNA that reached the desired threshold of an A_{260}/A_{280} purity ratio of 1.8 was stored at $-20\,^{\circ}C$ until further use.

3.2.3 Virus preparation

The White spot syndrome virus (WSSV) used in this study was obtained from a commercial laboratory located in Sabah, East Malaysia (BioSatria). Prior to use in immune challenge, the virus was propagated in the tiger shrimp *Peneus monodon*. The propagated virus was then extracted from muscle tissues of infected *P. monodon* samples through a filtration and homogenization method. Firstly, the dissected tissue was homogenized in 0.1 g/ml TN buffer (20 mM Tris–HCl, 400 mM NaCl, pH 7.4) and centrifuged at 2000g for 10 minutes. Then, the isolate obtained through this centrifugation was diluted to a ratio of 1:5 with 1% NaCl and filtered using a 0.45 µm syringe filter. The filtrate was stored at -80 $^{\circ}C$ prior to use as an inoculum for prawn samples.

To verify the viability of virus filtrate extracted, 3 healthy prawns were injected with 5 ml of the viral filtrate each. DNA from the injected samples was extracted after 24 hours using the DNA extraction methodology elaborated in Section 3.2.2. and tested through a one step PCR WSSV detection protocol (Nunan & Lightner, 2011). The PCR primers used in the one step detection are as shown in Table 3.8 while the chemicals and PCR regents (GoTaq

® Flexi, Promega, USA) used in the PCR reactions are shown in Table 3.9. The PCR reaction was carried out in a thermal cycler (BioRad, USA) using the temperature conditions shown in Table 3.10.

Table 3.8: Oligonucleotide sequence of primers used in the PCR the identification of WSSV infection.

Primer	Oligonucleotide sequence
146 F 2	5' -GTAACTGCCCCTTCCATCTCCA- 3'
146 R 2	5' -TACGGCAGCTGCTGCACCTTG - 3'

Table 3.9: PCR reagents used in the identification of WSSV infection, with their respective concentrations and volumes.

Reagent	Final concentration	Volume per reaction (μl)
MgCl$_2$	2.0 mM	1.50
Buffer	1X	3.00
dNTPs mix	0.2 mM each dNTP	1.00
Primer F	0.6 mM	0.40
Primer R	0.6 mM	0.40
DNA Polymerase	1.25 u	0.30
Nuclease free water	-	1.40
Purified DNA	50 ng/μl	2.00
Total		**10.00**

Table 3.10: Temperature properties of the thermal cycle used for PCR identification of WSSV infection.

Temperature (°C)	Time	Cycle
94	3 m	1
94	20 s	40
62	20 s	40
72	30 s	40
72	3 m	1
4	∞	--

To verify the results of the pCR screening, gel electrophoresis of the PCR product was conducted as described in Section 3.1.5. Positive results, shown by the band appearance of 942 bp sized amplicons in gel electrophoresis, confirm the presence of WSSV infection. Upon confirmation of the viability of the extracted virus, the experimental infection of *M. rosenbergii* with WSSV was started.

3.2.4 Experimental infection of WSSV

A total of 5 healthy *M. rosenbergii* prawns were isolated and kept as the control group. Separately, 20 prawns were infected with WSSV by intramuscularly injecting the fourth abdomen with 5 ml of viral filtrate each, using sterile syringe. A triplicate of prawn samples were anaesthetized and dissected at the following intervals: 3 hours, 6 hours, 12 hours, 24 hours, 36 hours and 48 hours post infection. From the control group, 3 samples were

dissected to be used as control specimens. All dissections were conducted in sterile conditions and the dissected samples individually snap-frozen in liquid nitrogen prior to storage.

3.3 Viral load quantification

3.3.1 Viral load Standard Curve Preparation

Quantification of WSSV was carried out on the DNA isolated from infected *M. rosenbergii* through an absolute quantification SYBR green based method qPCR method, using universal WSSV gene specific primers (Mendoza-Cano & Sanchez-Paz, 2013).

Table 3.11: Oligonucleotide sequence of primers used in the qPCR the quantification of WSSV copy number.

Primer	Oligonucleotide sequence
VP28 – 140 Fw	5' - AGGTGTGGAACAACACATCAAG- 3'
VP28 – 140 Rv	5' -TGCCAACTTCATCCTCATCA- 3'

A standard curve was prepared by conducting this quantitative PCR on viral DNA of known concentrations obtained from BioSatria (Sabah, Malaysia). The WSSV viral standard with a known copy number (6.62×10^6 copies/µl) was diluted through 10-fold serial dilution ranged from 6.62×10^{10} copies to 6.62×10^6 copies of the viral DNA filtrate. All 5 concentrations were quantified using the primers VP28-140Fw and VP28-140Rv through qPCR. Each qPCR reaction contained the chemicals shown in Table 3.12 (SYBR Green, Applied Biosystems, USA) at the shown concentrations and volumes.

Table 3.12: Chemicals and qPCR reagents used in the quantification of WSSV copy numbers in infected *M. rosenbergii*, with their respective concentrations and volumes.

Material	Concentration	Volume (µl)
SYBR Green SuperMix	2X	7.50
10 mM Forward Primer	300 nM	0.45
10 mM Reverse Primer	300 nM	0.45
DNA	50 ng/µl	1.00
H2O	-	5.6
Total		15

The qPCR reactions were carried out in a real time PCR system (7500 Real-Time PCR System, Applied Biosystem, USA) according to the temperature conditions shown in Table 3.13.

Table 3.13: Temperature properties of the thermal cycle used for qPCR quantification of WSSV copy number.

Temperature (°C)	Time	Number of Cycles
50.0 °C	2 m	1
95.0 °C	10 m	40
95.0 °C	15 s	40

60.0 °C	1 m	40

The cycle threshold (ct) values for each concentration of viral copy number was plotted to form a standard curve. This standard curve was later used to analyze the qPCR data for viral load quantification in infected *M. rosenbergii* samples.

3.3.3 Quantification of virus in samples

To quantify the viral load content of infected *M. rosenbergii* samples at different time points post WSSV infection, the DNA of infected *M. rosenbergii* muscle tissues were extracted as described in Section 3.2.2. The extracted DNA from each time point was quantified through qPCR using VP28-140Fw and VP28-140Rv primers as described in Tables 3.13 and 3.14. The ct values obtained through qPCR for each sample was compared with the standard curve constructed in Section 3.3.1 to obtain the copy number of the virus present in each sample. This quantification of the virus copy number was carried out in triplicates for each individual prawn, and the standard error of mean for each sample group calculated. The results obtained is shown in a line graph, with each point representing the mean copy number and error bars representing the standard error of each value.

3.4 Quantification of *Macrobrachium rosenbergii* dystrophin expression

3.4.1 Primer design and optimization

Gene specific primers were designed for the real time quantification of dystrophin expression in muscle tissues of *M. rosenbergii* using the primer design online tool Primer Plus 3 (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi). This primer

43

was designed to target the location of the conserved ZZ Dystrophin domain on the *M. rosenbergii* dystrophin sequence.

Table 3.14: The gene specific primers designed for quantification of *M. rosenbergii* dystrophin.

Primer	Oligonucleotide sequence
MrDysZZ F	5' – TAGCTGTTTTGCATCGTGTTG - 3'
MrDysZZ R	5' – TGGGGTGAGTGATCTTGTGA - 3'

A melt curve analysis was conducted to test the specificity of the designed primers prior to use in dystrophin quantification. The qPCR conditions used for the melting curve analysis are shown in Table 3.15.

Table 3.15: Temperature properties of the thermal cycle used for the melt curve analysis of the designed MrDysZZ F and MrDysZZ R primers.

Temperature (°C)	Time	Cycle
50.0 °C	2 m	1
95.0 °C	10 m	40
95.0 °C	15 s	40
60.0 °C	1 m	40
60.0 °C	7 m	1
95.0 °C	15 s	1
60.0 °C	1 m	1
95.0 °C	2m	1

The dissociation of the DNA was measured at different temperatures and the data obtained plotted into a melt curve graph using the in-built SDS 7500 software (Applied Biosystem, USA). The peak of the melt curves formed were inspected to ensure single peaks before proceeding to the quantification of dystrophin expression in *M. rosenbergii.*

3.4.2 Real-time PCR quantification

Each qPCR reaction for dystrophin quantification contained the following chemicals (SYBR Green, Applied Biosystems, USA) at the concentrations and volumes shown in Table 3.16.

Table 3.16: Chemicals and qPCR reagents used in quantification of *M. rosenbergii* dystrophin, with their respective concentrations and volumes.

Material	Concentration	Volume (µl)
SYBR Green SuperMix	2X	7.50
10 mM Forward Primer	300 nM	0.45
10 mM Reverse Primer	300 nM	0.45
cDNA	50 ng/µl	1.00
H2O	-	5.6
Total		15

The qPCR reactions were carried out in a real time PCR system (7500 Real-Time PCR system, Applied Biosystem, USA) according to the temperature conditions shown in Table 3.17.

Table 3.17: Temperature properties of the thermal cycle used for the quantification of
M. rosenbergii **dystrophin.**

Temperature (°C)	Time	Cycle
50.0 °C	2 m	1
95.0 °C	10 m	40
95.0 °C	15 s	40
60.0 °C	1 m	40

All reactions were carried out using SYBR Green supermix (Applied Biosystems, Germany) on an Applied Biosystems 7500 Real-Time PCR System and the cycle threshold (ct) values were calculated by the in-built ABI 7500 SDS software.

The internal control gene, elongation factor 1-alpha (ELF-1) (Dhar et al., 2009), was also similarly quantified. The relative gene expression of *M. rosenbergii* dystrophin was analyzed in accordance to the Livak method (Livak & Schmittgen, 2001) to study the changes in gene expression at different time points post WSSV infection.

Table 3.18: Oligonucleotide sequence of internal control primers used as reference gene quantification in the qPCR the quantification of *M. rosenbergii* **dystrophin**

Primer	Oligonucleotide sequence
EF-1 F	5' - TCGCCGAACTGCTGACCAAGA- 3'
EF-1 R	5' - CCGGCTTCCAGTTCCTTACC - 3'

Results are graphically presented as the mean expression of the *M. rosenbergii* dystrophin gene relative to the internal reference gene, ELF-1 of three replicates using the GraphPad Prism 6 software. The ANOVA statistical analysis was performed was performed to test the significance of the results using the same software.

3.5 Quantification of intracellular calcium concentration

3.5.1 Sample preparation

Muscle tissues from prawn samples were thawed and 50 mg of muscle tissue was weighed. The tissue was placed in a sterile 1.5 ml tube and immersed with 3% Trichloroacetic Acid solution. The tissue was crushed using a sterile micro-centrifuge tube homogenizer and vortexed for 1 minute prior to incubation at 4 ^{O}C for 30 minutes. The mixture was then centrifuged at 6 000 rpm for 15 minutes. The supernatant obtained was collected and used in the assay.

3.5.2 Calcium quantification

A colormetric quantification kit (Calcium Assay kit, Metallogenics, Japan) was used to quantify the amount of intracellular calcium ions present in muscle tissues of the studied prawn samples (Takahashi, Camacho, Lechleiter, & Herman, 1999) . On a 96-well plate, 190 µl of the R-A buffer was added into each well followed by 50 µl of chelate buffer. Each sample well contains 5 µl of the supernatant obtained through the homogenization process described in section 3.5.1, while the blank well contains 5 µl of distilled water, and 5 µl of standard calcium solution (10 mg/dL) was used as standard for calibration purposes. A summary of the contents of each well is provided in Table 3.19.

Table 3.19: The content of all blank, standard and sample wells in the intracellular calcium concentration assay.

Solutions	Assay samples (µl)		
	Blank	Standard	Sample
	OD_{blank}	OD_{std}	OD_{sample}
Distilled water	5	-	-
Standard calcium solution	-	5	-
Assay sample	-	-	5
R-A buffer	190	190	190
Chelate buffer	50	50	50

The 96 well plate was incubated at room temperature for 5 minutes, before the absorbance of each well was read at 570 nm and 700 nm using a microplate reader (TECAN M200 PRO, USA).
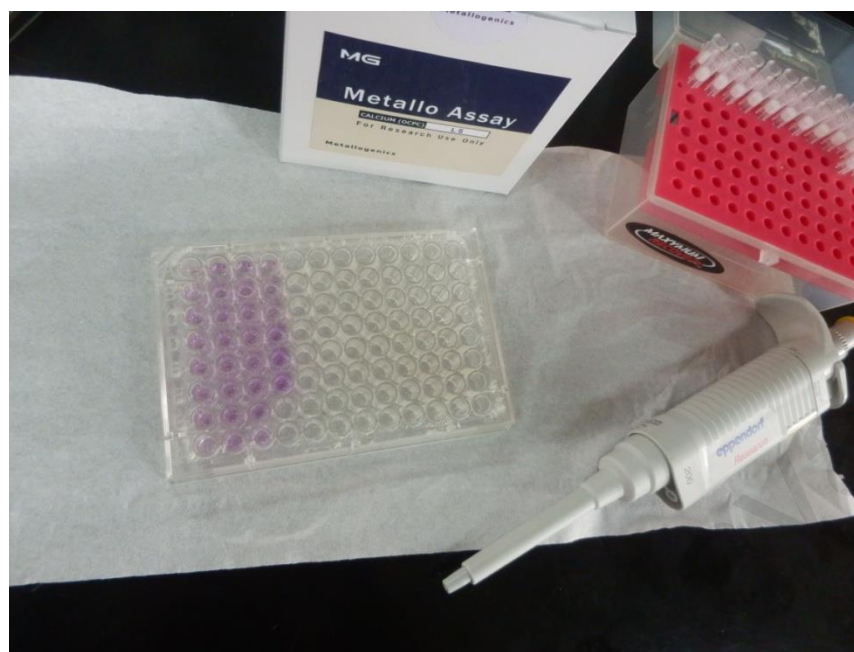
Figure 3.3: A 96 well plate containing the assayed tissues samples and florescent reagents described in Table 3.19.

The intracellular calcium concentration present in healthy prawn muscle tissues, as well as *M. rosenbergii* samples from each time point post WSSV infection was calculated using the formula:

$$\text{Calcium concentration (mg/dL)} = \frac{OD_{sample} - OD_{blank}}{OD_{std} - OD_{blank}} \times 10.$$

This quantification was conducted in three biological replicates with three technical replicates each. All numerical data obtained was analyzed through one way ANOVA and Dunnett's multiple comparison tests at a significance level of $P < 0.05$ using GraphPad Prism 6 software (California, USA). Data is represented as mean ± standard error mean (S.E.M.) for each set of readings.

3.6 Transmission Electron Microscope (TEM) Imaging of Muscle Tissues

3.6.1 Sample preparation

In order to view and compare the muscle morphology conditions of the muscle tissues of healthy and WSSV infected *M. rosenbergii*, tissue samples of infected *M. rosenbergii* at 24 hours and 48 hours post WSSV infection were prepared in through a gradient dehydration method. The dissected muscle tissues were trimmed into cubes with sides measuring 3 mm and fixed in 4 % glutaraldehyde (buffered in 0.1 M cacodylate buffer) for 4 hours at 4 ^{O}C. The cubes were then gently washed in 0.1 M cacodylate buffer, and this process repeated twice.

The washed tissue cubes were immersed in 1 % osmium tetroxide (buffered in 0.1 M cacodylate buffer) for 2 hours at 4 ^{O}C for fixation. Next, the tissue cubes were washed with double distilled water to remove any traces of the osmium tetroxide fixing reagent prior to gradual dehydration. Gradual dehydration of the sample tissues were carried out by soaking them in a series of ethanol solution with increasing concentrations. Ethanol solutions of the following percentages of concentration; 35 %, 50 %, 70 % and 95 % were prepared by mixing with ultra pure distilled water. The tissue cubes were dehydrated by immersion in the series of ethanol solutions, in ascending order, for 10 minutes at each concentration. Finally, the tissue cubes were dehydrated 3 times in pure ethanol (100 %) for 15 minutes each.

The dehydrated tissue cubes were placed in propylene oxide solution for 15 minutes, and this step was repeated twice. Next, the tissues were infiltrated with a propylene oxide and resin solution at the ratio of 1:1 for 1 hour. This step was repeated with propylene oxide and

resin infiltration at a ratio of 1:3 for 2 hours, followed by overnight soaking of tissues in pure resin. The fixated samples were then gently placed in specimen vials and rotated on a rotator at room temperature. Subsequently, the tissue cubes were embedded in capsules filled with resin and allowed to polymerize overnight at 60 ^{O}C in a heated chamber.

The polymerized blocks were then trimmed and cut into semi-thin sectioning using a glass knife and stained with methylene blue. Next, ultra-thin sections with a thickness of approximately 70 nm were cut using a diamond knife in a RMC PT-PC Powertome Ultramicrotome. The ultra-thin sections produced were stored in a vacuum container until transmission electron microscope (TEM) viewing.

3.6.2 Sample imaging

To obtain high resolution images of the muscle morphology, the prepared ultra-thin sections of the tissue were placed gently and precisely on a copper grid. The samples were then viewed in a HT7700 Transmission Electron Microscope (Hitachi, Japan) at 100 kV accelerating voltage in High Contrast Mode. Appropriate zooming and focusing options were selected for each viewing, and images of infected and healthy *M. rosenbergii* muscle fibers were captured using a charge-coupled device camera.

4.1 Characterization of Macrobrachium rosenbergii dystrophin

4.1.1 Amplification and verification *M. rosenbergii* dystrophin sequence

Several Polymerase Chain Reaction (PCR) primers were designed to target different regions of the *Macrobrachium rosenbergii* dystrophin transcript obtained from its transcriptome and the sequence was amplified through PCR. Gel electrophoresis was conducted to inspect PCR products obtained from all three sets of the designed gene specific primers. Gel electrophoresis of the PCR products obtained show that amplification was successful in all three sets of gene specific primers designed (Figure 4.1). To further verify contents of the PCR products, the PCR products obtained using the primers Dys1 (381 bp), Dys2 (718 bp), and Dys3 (1118 bp) were sequenced through Sanger sequencing. Complete sequences obtained from each PCR product and their alignment is shown in the Appendix Table 1 and Figure 1 respectively.
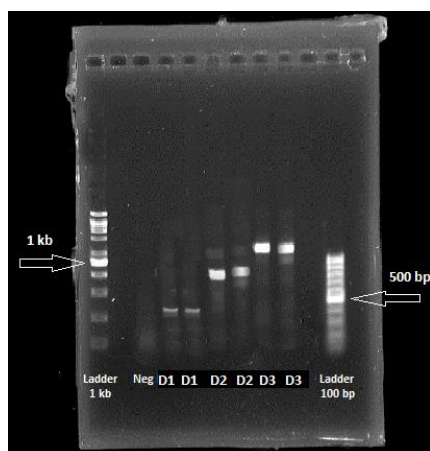
Figure 4.1: Agarose gel electrophoresis of the amplified products of primers Dys1, Dys2 and Dys3. The contents of the gel electrophoresis wells are 1kb DNA ladder, PCR products of Dys1, Dys2 and Dys3 primers as well a 100 bp DNA ladder.

```
1     CGGTAGAGGA AGTGAGCGGG CGGTTAGACG AAGCCCAAAA GACTGCCGAT GGGTGGGGCC   60
61    CCATCTCCCA AGACCCGAAC CTCGCAGATC AACACGCCCA CAATGTTAGA AAATTCCGCG   120
121   AAGGTTTGGC GGAATTGCAG CGCGGAGTGG ACGACGTCAA CGACCAAGCC GCTCGGTTTT   180
181   CGGCTCATAA TGTGCCACTC ACCCCTGCTA ACTCGGCCAA GCTCCATGAC CTCAATAATA   240
241   GATGGAAAAA TTTAGAAACA GCAGTCAATG ACAGATGGCG ACAAGTGGCC AGCAGATCTC   300
301   GAGAGGTCAC ACCACTCACC CCTGCCCAGT TGGCATCTTC AGTGTCACCA CCATGGGAGA   360
361   GGGCTACAAC TAGTAATAAA GTGCCTTATT ACATCAATCA TGATCAGGAG AGTACCCACT   420
421   GGGATCATCC AATCATGATG GACTTACTAG ACTCTATGAA TGAATTCAAC CATATCAAGT   480
481   TCTCTGCCTA CAGAACAGCT ACAAAACTTA GAATGTTACA AAAGAAGCTG TCATTGGATC   540
541   TTGCCAAACT GCAAATGGCA ACAGATGTAT TTGATGAGCA TGGTCTCAGG GGACAAAATG   600
601   ACCGGCTAAT TGATGTAGGT GATATGGTCG TGGTTTTGTC AGCTTTATAT GCAAATATAT   660
61    CTGCTGATCA TCCAGAAGTT AATACCACCC TTGCCATAGA CCTCTGTCTC AATTGGCTTT   720
721   TAAATGTATA TGACAGCCAA AGGACAGGGC AGATGAGAGT TCTTTCTTTC AAGATCGGTC   780
781   TCGTCTGCCT TTGCTGCGGT CATTTAGAAG AAAAATACAG ATACATGTTC CGGCTAATTG   840
41    CCGACCCAAA TCGCCTAGTT GATCAAAGAA AATTAGGGTT ATTGTTACAC GACTGTGTAC   900
901   AAGTTCCACG CCAGTTGGGT GAGGTTGCAG CATTTGGTGG ATCAAACATT GAACCATCAG   960
961   TTCGTAGCTG TTTCACAAAA GCTGGAAAGG ACAGAGAAAC TATTGAGGCT GTGCATTTCT   1020
1021  TGGCATGGGT GCAACAAGAA CCCCAGTCAC TAGTGTGGTT AGCTGTTTTG CATCGTGTTG   1080
1081  CAGCCTCAGA AAATATTCAG CATCAGGTCA AATGCAATAT CTGCAAGGCA TATCCTATTG   1140
1141  TAGGCTTACG GTACCGTTGC CTCAAGTGCC TCAGCTTTGA TATGTGTCAG CGGTGCTTCT   1200
1201  TTGATGGAAG AAGCGGCAAG AGTCACAAGA TCACTCACCC CATGCG                 1246
```

Figure 4.2: The 1246 bp long ucleotide sequence of *M. rosenbergii* Dystrophin obtained from M. *rosenbergii* muscle tissue, numbered from the 5' end. This sequence was submitted to the NCBI database with the accession number KU198993.

Upon obtaining experimental confirmation through PCR and computational structural prediction, the 1246 base pair dystrophin sequence (Figure 4.2) obtained was submitted to the NCBI GenBank database with the accession number KU198993. A homology analysis though BLAST on the NCBI database revealed high similarity of the *M. rosenbergii* dystrophin sequence with predicted dystrophin sequences from various other organisms

including arthropods such as *Acromyrmex echinatior*, *Cerapachys biroi*, *Harpegnathos saltator*, *Megachile rotundata*, *Wasmannia auropunctata* and *Apis florea*. No previous records of the dystrophin gene in prawn or shrimp species were found through the homology search.

4.1.2 Characterization of *M. rosenbergii* dystrophin

The nucleotide sequence of *M. rosenbergii* dystrophin was computationally translated through ExPASy bioinformatics resources and the deduced amino acid sequence of *M. rosenbergii* dystrophin is shown in Figure 4.3. This computational protein translation of the sequence produces a 414 amino acids long protein sequence.
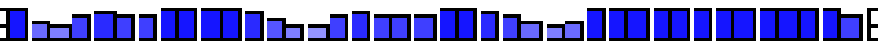
A physiochemical analysis was conducted on the amino acid sequence using the protein analysis tool ProtParam, also on the ExPASy platform. The data obtained through this analysis showed a predicted molecular weight of 47 kDa and a theoretical isoelectric point of 7.69 for *M. rosenbergii* dystrophin. A similar analysis was also conducted on homologous *Drosophila melanogaster* and *Danio rerio* dystrophin nucleotide sequences that have been retrieved from the NCBI database, and computationally translated into protein sequences. A comparison of the physiochemical properties of all 3 amino acid sequences are shown in Table 4.1.

```
1    V E E V S    G R L D E    A Q K T A    D G W G P    20
21   I S Q D P    N L A D Q    H A H N V    R K F R E    40
41   G L A E L    Q R G V D    D V N D Q    A A R F S    60
61   A H N V P    L T P A N    S A K L H    D L N N R    80
81   W K N L E    T A V N D    R W R Q V    A S R S R    100
101  E V T P L    T P A Q L    A S S V S    P P W E R    120
121  A T T S N    K V P Y Y    I N H D Q    E S T H W    140
141  D H P I M    M D L L D    S M N E F    N H I K F    160
161  S A Y R T    A T K L R    M L Q K K    L S L D L    180
181  A K L Q M    A T D V F    D E H G L    R G Q N D    200
201  R L I D V    G D M V V    V L S A L    Y A N I S    220
221  A D H P E    V N T T L    A I D L C    L N W L L    240
241  N V Y D S    Q R T G Q    M R V L S    F K I G L    260
261  V C L C C    G H L E E    K Y R Y M    F R L I A    280
281  D P N R L    V D Q R K    L G L L L    H D C V Q    300
301  V P R Q L    G E V A A    F G G S N    I E P S V    320
21   R S C F T    K A G K D    R E T I E    A V H F L    340
341  A W V Q Q    E P Q S L    V W L A V    L H R V A    360
361  A S E N I    Q H Q V K    C N I C K    A Y P I V    380
381  G L R Y R    C L K C L    S F D M C    Q R C F F    400
401  D G R S G    K S H K I    T H P M                   414
```

Figure 4.3: The deduced _M. rosenbergii_ dystrophin amino acid sequence obtained through computational translation, numbered from the 5' end.

Table 4.1: Predicted physicochemical properties of dystrophin from *H. sapiens, M. rosenbergii, D. melanogaster* and *D. rerio*.

Properties	*Homo sapien* dystrophin	*Macrobrachium rosenbergii* dystrophin	*Drosophila melanogaster* dystrophin	*Danio rerio* dystrophin
Number of Amino Acids	3690	414	951	634
Molecular Weight (Dalton)	427329.3	47022.6	108196.1	72027.6
Theoretical pI	5.65	7.69	5.94	6.67
Total number of negatively charged residues (Asp+Glu)	580	48	133	79
Total number of positively charged residues (Arg+Lys)	492	49	116	75
Total number of atoms	60223	6565	15005	10023
Computed Instability Index	56.34	40.73	56.97	58.09
Aliphatic Index	87.63	86.91	76.87	76.80
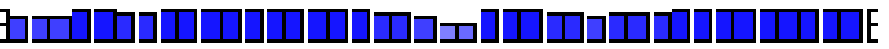
```
Conf: [confidence bar graph]
Pred: [helix] [helix]
Pred: CHHHHHHHHHHHCCCCCCCCCCCHHHHHHHHHHHHHHHH
  AA: VEEVSGRLDEAQKTADGWGPISQDPNLADQHAHNVRKFRE
              10        20        30        40

Conf: [confidence bar graph]
Pred: [helix] [helix]
Pred: HHHHHHHHHHHHHHHHHHHHCCCCCCCCCCHHHHHHHHHHH
  AA: GLAELQRGVDDVNDQAARFSAHNVPLTPANSAKLHDLNNR
              50        60        70        80

Conf: [confidence bar graph]
Pred: [helix] [helix]
Pred: HHHHHHHHHHHHHHHHHHHHCCCCCCCCCCCCCCCCCCHHHH
  AA: WKNLETAVNDRWRQVASRSREVIPLTPAQLASSVSPPWER
              90       100       110       120

Conf: [confidence bar graph]
Pred: [helix] [arrow] [helix]
Pred: HCCCCCCCCCEBCCCCCCCCCCCCCHHHHHHHHHHHHHHCCCH
  AA: ATTSNKVPYYINHDQESTHWDHPIMMDLLDSMNEFNHIKF
             130       140       150       160

Conf: [confidence bar graph]
Pred: [helix] [helix]
Pred: HHHHHHHHHHHHHHHCCCCCHHHHHHHHHHHCCCCCCCC
  AA: SAYRTATKLRMLQKKLSLDLAKLQMATDVFDEHGLRGQND
             170       180       190       200

Conf: [confidence bar graph]
Pred: [helix] [helix]
Pred: CCCCHHHHHHHHHHHHHHHHHHHHCCCCCCHHHHHHHHHHHH
  AA: RLIDVGDMVVVLSALYANISADHPEVNTTLAIDLCLNWLL
             210       220       230       240

Conf: [confidence bar graph]
Pred: [helix] [arrow] [helix] [helix]
Pred: HHHCCCCCCCCEEEEHHHHHHHHHHCCOCHHHHHHHHHHHHHHCC
  AA: NVYDSQRTGQMRVLSFKIGLVCLCCGHLEEKYRYMFRLIA
             250       260       270       280
```

57

Figure 4.4: The secondary structure of *M. rosenbergii* dystrophin as predicted through the PSIPRED secondary prediction method. The legend displayed below the structure shows the symbols representing the helixes, strands and coils.

The secondary structure of *M. rosenbergii* dystrophin was predicted through the PSIPRED secondary prediction method to obtain a better understanding of the structural content of the protein sequence. The secondary structure obtained shows several helixes, coils and strands, as well as their respective locations on the amino acid sequence (Figure 4.4). The confidence level of the prediction represented by histograms throughout the sequence, also indicate high confidence throughout the sequence.
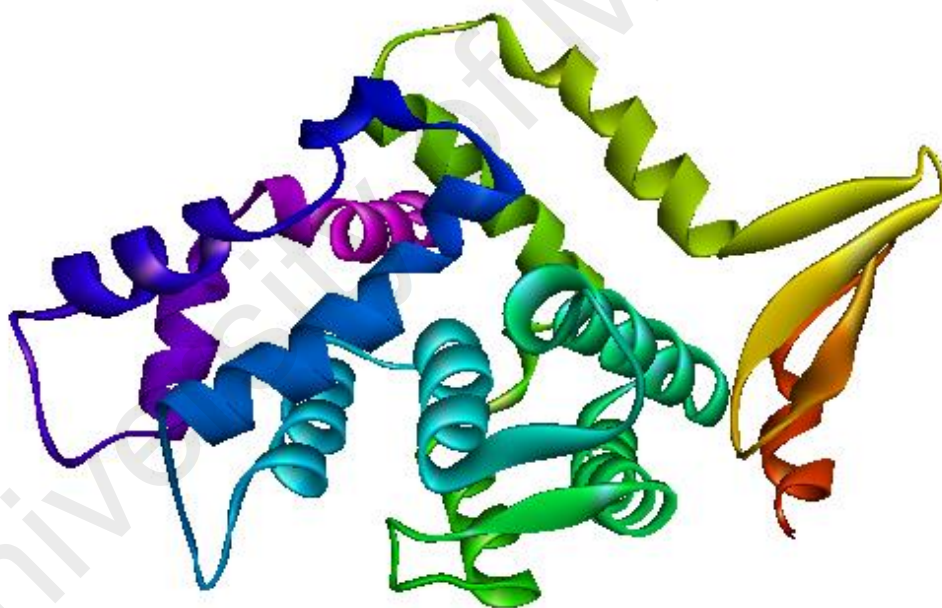
Figure 4.5: The 3D protein model of *M. rosenbergii* dystrophin constructed by target-template alignment with the best matched protein template. This model obtained a QMEAN4 score of -2.18 and a GMQE score of 0.56.

A 3D model of the *M. rosenbergii* dystrophin protein was produced using the SWISS-MODEL workspace. Templates matching the *M. rosenbergii* dystrophin sequence were identified in the Protein Data Bank template library and the quality of each template was studied through target-template alignment. The target-template alignment with the highest quality is shown in Appendix Figure 3. A 3D model was then built based on this target-template alignment, and its quality assessed using QMEAN and GMCE scores (Figure 4.5). The model obtained for the monomer *M. rosenbergii* dystrophin did not contain any ligands.

Verification of the accuracy of the *M. rosenbergii* dystrophin 3D protein model was carried out by constructing a Ramachandran Plot (Figure 4.6). In this plot, the green circle signs represent glycine residues, squares represent prolines and triangles represent all other residues. The plot obtained indicates that the amino acid residues in the predicted protein structure have good confirmation to the accepted bond angle values, as well as a favourable empirical distribution of the residues.
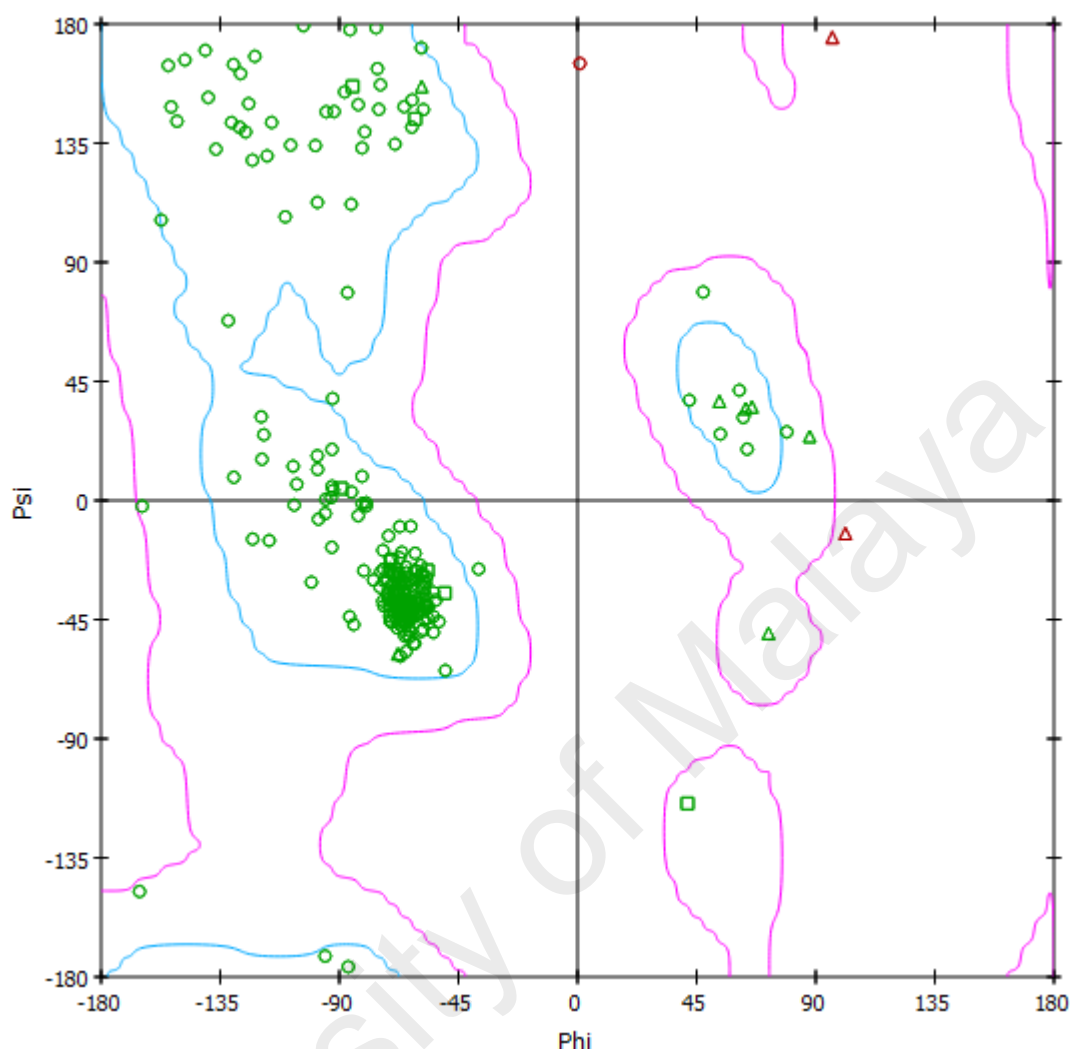
Figure 4.6: A Ramachandran plot of the predicted *M. rosenbergii* dystrophin protein model. The X axis shows the Phi angle of each residue, while the Y axis shows the Psi angle of each residue. The blue contours indicate the preferred region for helixes, red for strand residues, while all residues are shown either as green circles, squares or triangles.

The Ramachandran plot of the predicted *M. rosenbergii* dystrophin protein, as well as the

3D model was then analyzed through the protein checking program "WHAT_IF" to

numerically assess the suitability and accuracy of the prediction (Table 4.2). The z-scores

obtained evaluated the Ramachandran plot as falling within the accepted range for a well-

defined protein structure with a score of -0.170. Similarly, the protein model obtained a

backbone conformation Z-score of 0.697, indicating a well refined protein structure.

Table 4.2: **Evaluation scores of the *M. rosenbergii* dystrophin protein structural model obtained through a series of protein check conducted by the WHAT_IF program.**

Evaluation scores	Dystrophin *(M. rosenbergii)*
Z-scores	
Ramachandran plot appearance	-0.170
χ_1/χ_2 Rotamer normality	-1.549
Backbone conformation	0.697
1st generation packing quality	-0.321
2nd generation packing quality	-1.964
RMS Z-scores	
Bond lengths	0.681
Bond angles	1.242
Omega angle restraints	0.864
Side-chain planarity	1.342
Improper dihedral distribution	1.994 (loose)
Inside/Outside distribution	1.015

The *M. rosenbergii* protein dystrophin sequence was then studied for conserved protein

domains and four conserved domains were identified; the WW, EF-Hand-2, EF-Hand-3 and

ZZ Dystrophin domains, as well as a Spectrin region (Figure 4.7). All five of these protein

domains are well studied characteristic domains of the human dystrophin protein, each playing a crucial role in the structure of dystrophin. The discovery of all four of these dystrophin domains, as well as the spectrin region within the *M. rosenbergii* dystrophin sequence is remarkable due its much smaller size compared to the human dystrophin gene. The function, size and structure of each one of these domains are explored in detail in the discussions section of this dissertation.
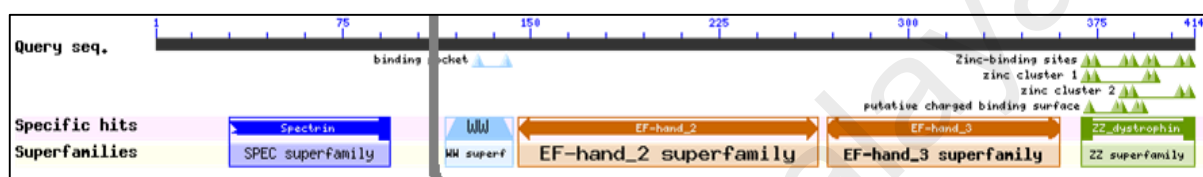
Figure 4.7: Conserved protein domains indentified in the deduced *M. rosenbergii* dystrophin amino acid sequence. Each domain is shown in a different color and labeled with its protein domain name.

4.1.3 Phylogenetic Analysis

The NCBI database was searched for previously identified dystrophin sequences, and 11 characterized sequences were found, each from a different organism. A full list of the organism of origin for each sequence and their NCBI accession number has been included in the earlier Table 3.4. Meanwhile, the nucleotide sequence of each of the 11 dystrophin sequences is included in the Appendix Table 8.

A phylogenetic tree was constructed to further investigate the relationship between *M. rosenbergii* dystrophin and the twelve different dystrophin sequences (Figure 4.8). The phylogenetic tree based on maximum likelihood shows *M. rosenbergii* dystrophin clustered together with dystrophin homologues from *C. elegans.* Another dystrophin homologue studied, from *D. melanogaster*, is also shown to be closely related.

```
                                                          0.07
                                                    ┌──────────── Gallus gallus dystrophin
                                              0.02   │
                                          ┌─────┐     │              0.02
                                          │     └─────┤          ┌──────── Homo sapiens dystrophin
                                    0.05   │            │    0.07  │
                                ┌─────────┤            └─────────┤  0.03
                          0.06   │          │                      └──────── Canis lupus familiaris dystrophin
                    ┌───────────┤          │            0.10
              0.10  │            │          └──────────────────── Xenopus laevis dystrophin
          ┌────────┤            │                   0.13
    0.06  │         │            └────────────────────────────── Scyliorhinus caniculua dystrophin
  ┌──────┤         │                    0.12
  │       │         └─────────────────────────────────────────── Danio rerio dystrophin
  │  0.08 │                            0.16
  │       └─────────────────────────────────────────────────── Branchiostoma lanceolatum dystrophin
  │                          0.20
┌─┤ ────────────────────────────────────────────────────────── Asteroidea sp.dystrophin
│ │                     0.18
│ └──────────────────────────────────────────────────────────── Pectinidae sp. dystrophin
│                                               0.26
│                                     ┌──────────────────── Drosophila melanogaster dystrophin
│  0.01                               │         0.26
└─────────────────────────────────────┤    ┌─────────────── Macrobrachium rosenbergii dystrophin (MrDys)
                                       └────┤  0.46
                                        0.09 └─────────── Caenorhabditis elegans Dystrophin related
```

Figure 4.8: Phylogenetic tree of *M. rosenbergii* dystrophin with dystrophin nucleotide sequences from 11 different organisms inferred by the Maximum Likelihood method using MEGA 6. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site.

4.1.4 Conserved domain homology analysis

Conserved protein domains on all of the identified dystrophin sequences from twelve different organisms reveal that similar conserved domains are found in all twelve sequences (Figure 4.9). All of the sequences contain at least 4 out of the 5 dystrophin characteristic protein regions identified; Spectrin, WW, EF-Hand-2, EF-Hand-3 and ZZ Dystrophin. A protein sequence alignment of 3 dystrophin sequences from *M. rosenbergii, D. melanogaster* and *D.rerio* was conducted through amino acid sequence alignment and is shown in the Appendix Figure 2. The location of each aligned region in comparison to its position in *M. rosenbergii* is graphically presented in a schematic diagram (Figure 4.9).

64
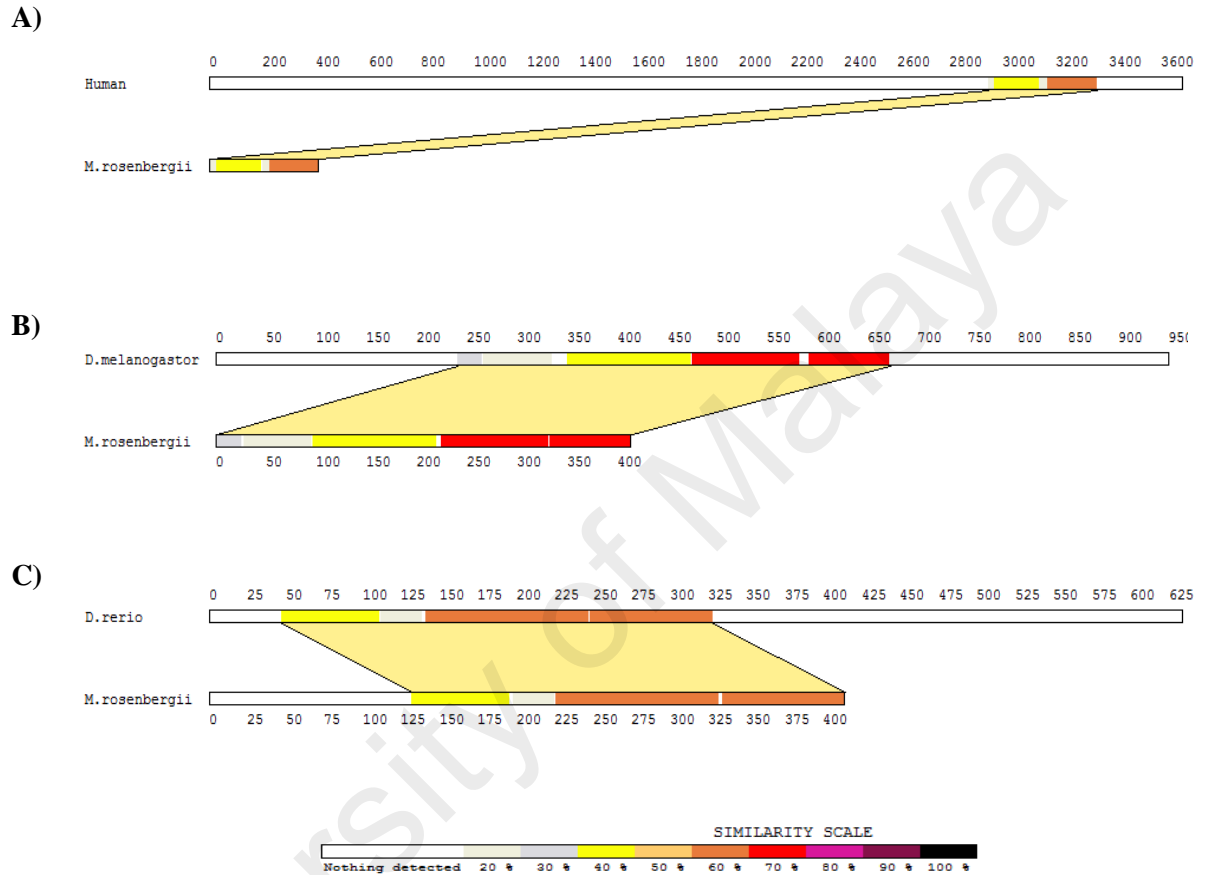
Figure 4.9: A schematic diagram showing the alignments and positions of sequence similarities of the *M. rosenbergii* dystrophin amino acid sequence and with it counterparts in humans, *D. melanogaster* and *D. rerio*. A) The location of similarity of the 414 amino acids long *M. rosenbergii* dystrophin sequence with the 3690 amino acids long human dystrophy sequence. **B)** The alignment of *M. rosenbergii* dystrophin protein sequence with its homolog sequence in *D. melanogaster.* **C)** The alignment of *M. rosenbergii* dystrophin protein sequence with its homolog sequence in *D. rerio.* The level of similarity indicated by each color is shown in the Similarity Scale legend at the bottom of the diagram.
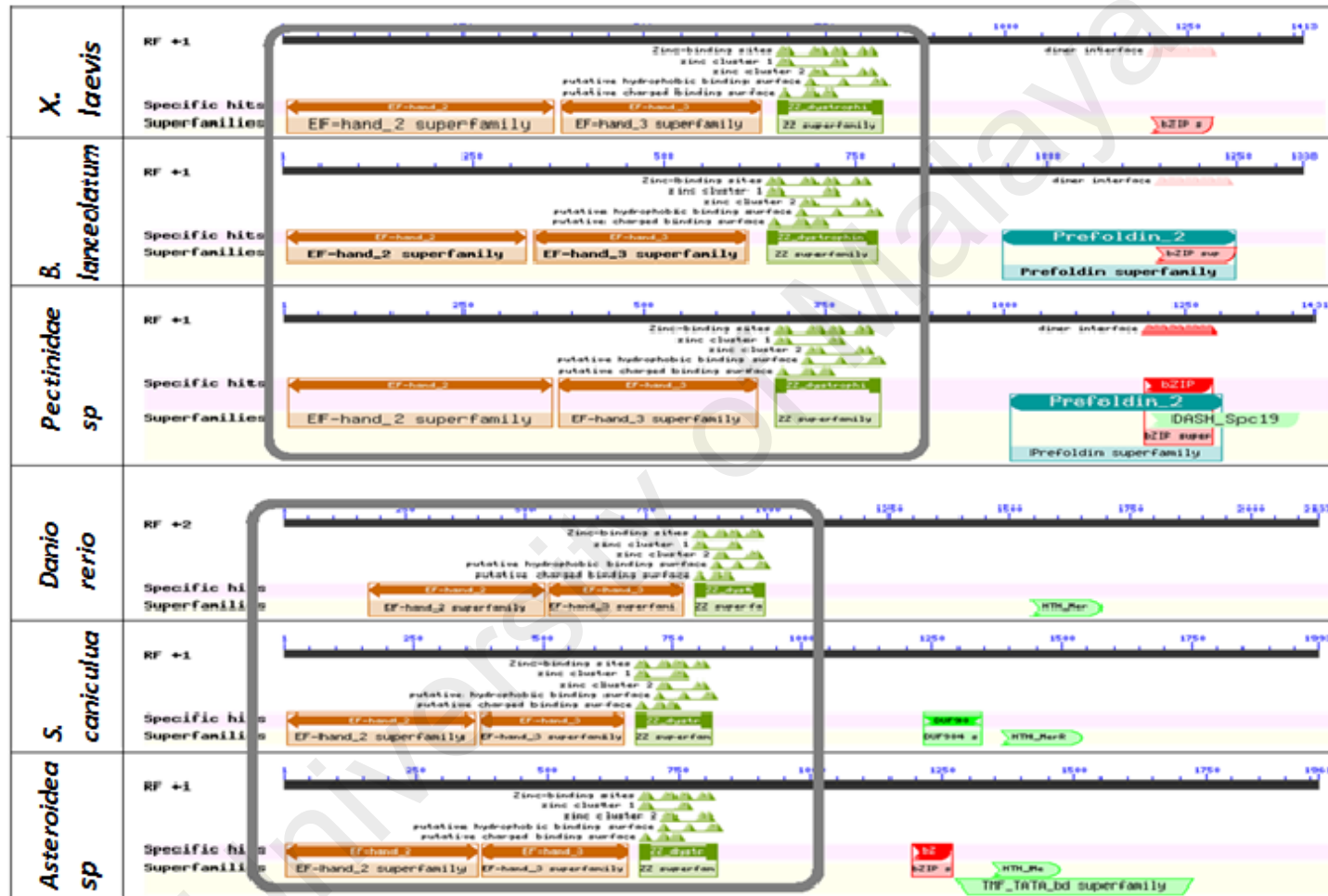
Figure 4.10: A graphical illustration of the conserved protein domains found in 12 dystrophin sequences retrieved from the NCBI nucleotide database, with the accession number for each sequence. Similar repeated domain patterns are grouped together and marked.

Table 4.3: The Positions of potential cleavage sites in the amino acid sequence of *M. rosenbergii* dystrophin.

Name of enzyme	No. of cleavages	Positions of cleavage sites
Proteinase K	192	1 2 3 4 8 10 11 14 15 18 21 27 28 32 35 38 40 42 43 44 45 49 52 56 57 59 61 64 66 67 69 72 74 77 81 84 85 86 87 88 92 95 96 101 102 103 105 106 108 110 111 114 118 119 121 122 123 127 129 130 131 136 138 140 144 148 149 154 155 158 160 162 163 165 166 167 169 172 176 178 180 181 183 186 187 189 190 192 195 202 203 205 209 210 211 212 214 215 216 217 219 221 225 226 228 229 230 231 232 234 236 238 239 240 242 243 248 253 254 256 258 260 261 263 268 269 270 272 274 276 278 279 280 285 286 291 293 294 295 299 301 305 307 308 309 310 311 316 317 320 324 325 327 332 333 334 335 336 337 339 340 341 342 343 346 350 351 352 353 354 355 356 359 360 361 363 365 369 373 376 377 379 380 382 384 387 390 392 399 400 410 411
Thermolysin	127	7 14 20 26 27 31 34 37 41 42 48 55 56 58 60 65 68 71 73 83 86 87 94 95 104 107 109 110 113 120 130 143 144 145 148 151 157 159 161 165 168 170 171 175 177 180 182 184 185 189 194 201 202 208 209 210 211 213 214 216 218 220 229 230 231 235 238 239 241 250 252 253 255 257 259 260 262 267 274 275 277 278 279 284 285 290 292 293 294 298 304 308 309 310 315 319 323 326 333 336 338 339 340 342 349 350 352 353 354 355 358 359 360 364 368 372 375 378 379 381 386 389 391 398 399 409 413
Pepsin	101	8 17 27 37 42 44 45 59 65 73 74 76 81 83 92 104 109 110 130 139 140 147 148 149 154 155 160 162 163 168 169 171 175 178 179 180 182 183 189 190 194 202 211 212 214 215 216 229 230 233 234 235 236 237 238 239 240 242 243 253 255 256 260 262 263 267 268 271 272 274 276 278 285 290 293 294 295 304 310 311 324 338 339 341 342 349 350 351 352 353 355 356 381 382 384 386 389 391 392 398 400
Chymotrypsin	94	8 18 27 31 33 38 42 45 59 62 66 74 77 81 84 92 105 110 118 129 130 140 145 146 148 149 152 155 157 160 163 169 171 172 176 178 180 183 185 190 193 195 202 208 212 215 216 230 234 236 238 239 240 243 251 254 256 260 263 267 268 272 274 275 276 278 285 291 293 294 295 305 311 324 338 339 340 342 350 352 353 356 357 367 382 384 387 390 392 394 399 400 408 414
Trypsin	49	7 13 36 37 39 47 58 73 80 82 91 93 98 100 120 126 159 164 168 170 174 175 182 196 201 247 252 257 271 273 277 284 289 290 303 321 326 329 331 358 370 375 383 385 388 397 403 406 409

4.1.5 Protease Cleavage analysis

A protein cleavage analysis was carried out on the protein translation of *M. rosenbergii* dystrophin with a total of 23 enzymes and chemicals. The results obtained show possible protein cleavage by a variety of enzymes including Chymotrypsin, Proteinase-K, Thermolysin, Pepsin, and several other endopeptidases.

The enzymes with the 5 highest probabilities of protein cleavage on the sequence; Proteinase K, Thermolysin, Pepsin, Chymotrypsin and Trypsin are shown in Table 4.3. Based on these findings, as well as the significance of Chymotrypsin in regards to WSSV infection, Chymotrypsin cleavage probabilities were looked deeper into. An in-depth calculation of the probability of cleavage at each potential site revealed 52 potential cleavage sites with a Chymotrypsin cleavage probability of 51% or more. The probability of Chymotrypsin cleavage at each one of these potential cleavage sites on the *M. rosenbergii* dystrophin is reported in percentages the Appendix Table 7.

4.2 Immune challenge with White Spot Syndrome Virus (WSSV)

4.2.1 Experimental infection of WSSV

Viral inoculum of WSSV was prepared and the experimental infection of *M. rosenbergii* was conducted by injecting the prawn specimens. DNA was extracted from the injected prawns 24 hours and 48 hours post injection and tested for WSSV injection through PCR screening to check the viability of the inoculated virus. The PCR products obtained were viewed on an agarose gel for evaluation.

Gel electrophoresis of the PCR products of the WSSV infection testing showed distinct bands at 24 hours post injection (Figure 4.10), as well as at 48 hours post injection (Figure 4.11). The positive results obtained confirmed successful infection of *M. rosenbergii* as well as the viability of the inoculated viral extract. Thus, immune challenge of *M. rosenbergii* for the purpose of studying dystrophin expression was continued using the same viral extract.
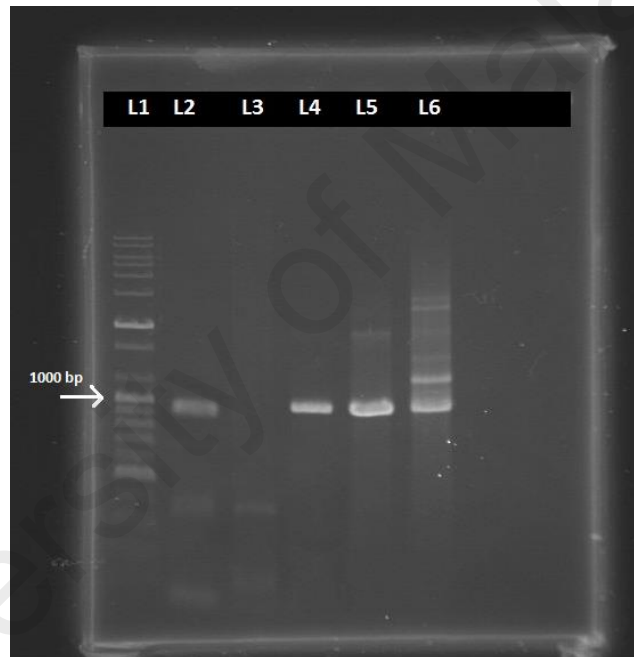
Figure 4.11: The gel image of PCR product of WSSV screening of infected *M. rosenbergii* at 24 hours post injection. The contents of the gel electrophoresis wells are L1: 1000 bp DNA ladder, L2: Viral positive control, L3: Viral negative control, L4 – L6: PCR products of WSSV viral screening. Positive results are shown by the presence of 942 bp bands.
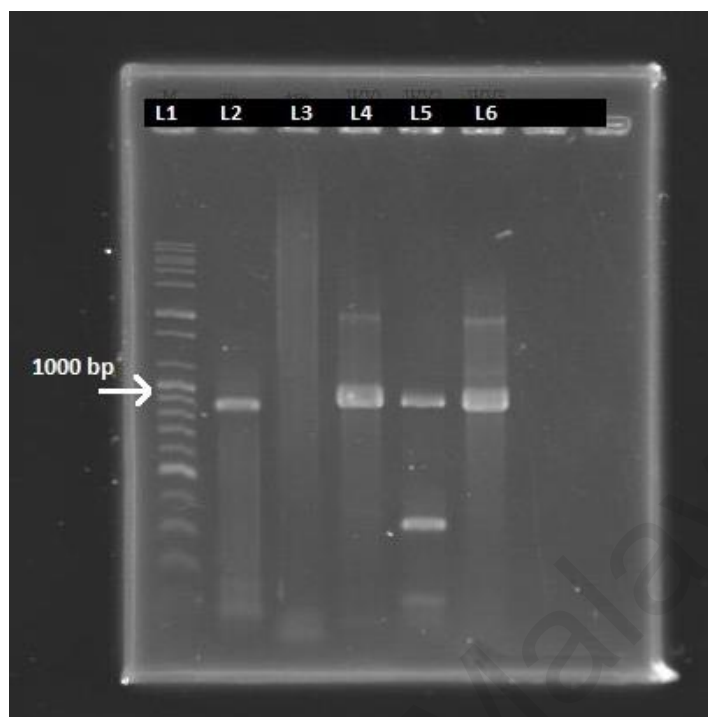
Figure 4.12: The gel image of PCR product of WSSV screening of infected *M. rosenbergii* at 48 hours post injection. The contents of the gel electrophoresis wells are L1: 1000 bp DNA ladder, L2: Viral positive control, L3: Viral negative control, L4 – L6: PCR products of WSSV viral screening. Positive results are shown by the presence of 942 bp bands.

Upon injection of WSSV into the appendage of healthy *M. rosenbergii*, infection symptoms were not observable in the early stages (0-12 hours post injection). At 12 hours post injection, the prawns began to gradually grow lethargic and exhibited a decline in appetite. Other clinical symptoms (discoloration of muscle and loose cuticles) caused by WSSV were detected at 24 hours post injection. The infection symptoms continued to deteriorate until 48 hours post infection. No mortalities were observed in the prawns although *M. rosenbergii* at 48 hours post WSSV infection appeared highly moribund, by lying at the bottom of the tank with debilitated swimming abilities.

4.2.2 Viral load quantification Standard Curve

In order to quantify the viral content of WSSV in *M. rosenbergii* samples at different time points post WSSV infection, absolute quantification through qPCR was carried out. A viral load standard curve was constructed by serially diluting the viral extract of WSSV virus to five different concentrations: 10 ng/µl, 1 ng/µl, 0.1 ng/µl, 0.01 ng/µl and 0.001 ng/µl. The ct value of virus content at each concentration was obtained through qPCR and a standard curve was plotted using these ct values and concentrations (Figure 4.13). The best fit line was constructed and its equation y = -0.38x + 33.393 was obtained.
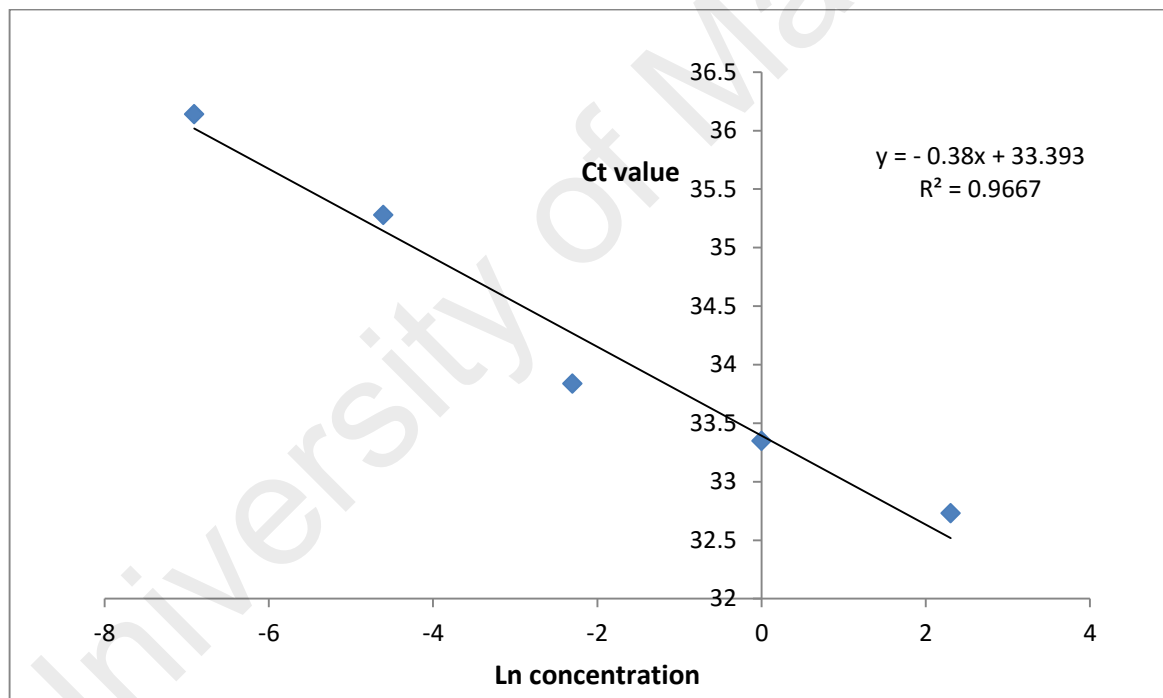
Figure 4.13: Standard curve of ct values for quantitative PCR of the VP28 region of the WSSV virus constructed through qPCR of viral filtrate with ten-fold serial dilution. The line of best fit has an equation of y = -0.38x + 33.393 with an efficiency of 96.67%.

4.2.3 Viral quantification of infected samples

DNA extracted from *M. rosenbergii* muscle tissue samples at different time points post infection was used in the absolute quantification of viral content in the infected samples. Using the equation obtained from the standard curve (y = -0.38x + 33.393), the DNA concentration was calculated for ct values obtained at each time point post infection. The viral copy number was calculated using the formula:

$$\text{Copy number} = \frac{(mass \times 6.022 \times 10^{23})}{(length \times 10^9 \times 650)}$$

The results obtained were plotted in a line graph with error bars at each point representing the standard error of mean for each sample group (Figure 4.14). Quantification of WSSV copy numbers in infected *M. rosenbergii* muscle tissues showed no significant changes until 24 hours post injection. The copy numbers of WSSV at 3, 6 and 24 hours post injection were 5.47×10^4, 2.95×10^3 and 4.99×10^5 copies/µl, respectively. However, the copy number of WSSV increased drastically at 36 hours to 5.4×10^7 copies and continued to increase exponentially until it reached 2.1×10^9 copies/µl at 48 hours post WSSV infection.
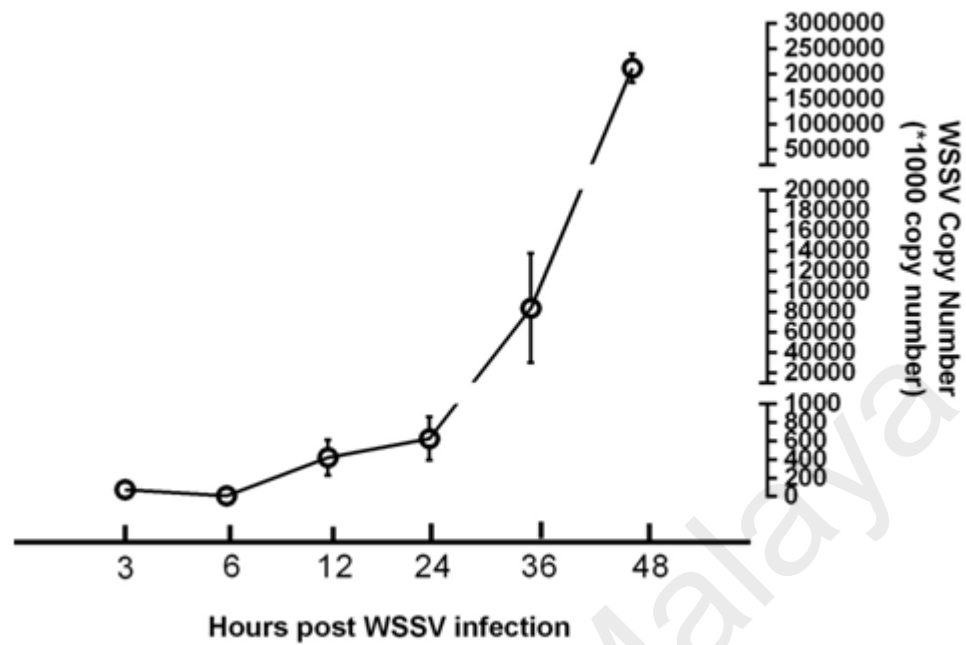
Figure 4.14: WSSV Viral copy numbers in the DNA of infected *M. rosenbergii* muscle tissues at different time points post infection. The quantification was carried out in triplicates, and the standard error of mean for the viral copy number of each sample group is shown by error bars at each point.

4.3 Quantification of M. rosenbergii dystrophin through qPCR

4.3.1 Primer optimization

Gene specific primers were designed to quantify *M. rosenbergii* dystrophin through qPCR, targeting an essential dystrophin protein domain, the ZZ domain. The synthesized oligonucleotide primers were tested for target specificity through a melt curve analysis study of the gene specific qPCR primers designed. Melt curve analysis show a single peak, indicating a single product intercalated with the SYBR green dye used, as a result of a successful reaction (Figure 4.15).
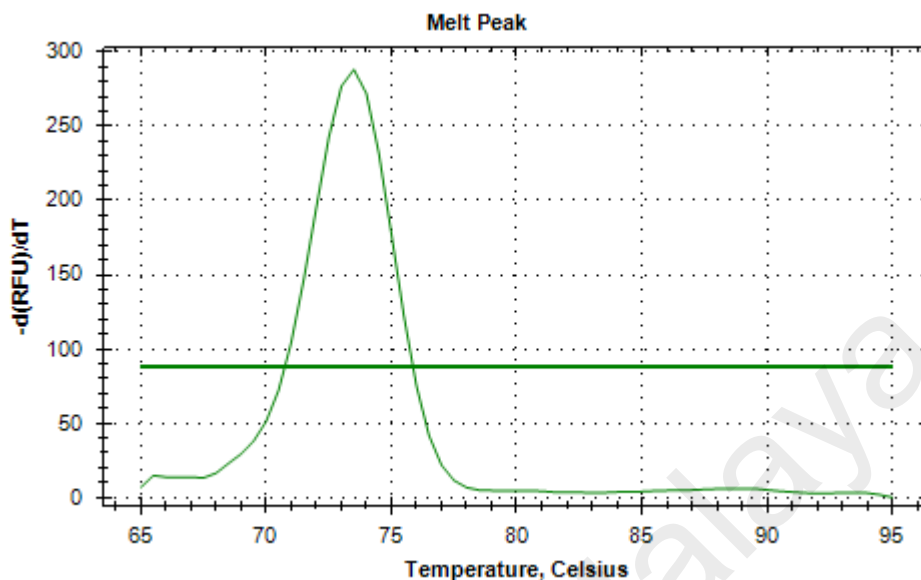
Figure 4.15: Melt curve analysis of gene specific qPCR primers designed to quantify the expression of *M. rosenbergii* **dystrophin in muscle tissue cDNA. A single peak was produced, indicating the intercalation of a single PCR product with the SYBRgreen dye used, as a result of successful reaction.**

The qPCR amplicon produced was also sequenced to further verify the specificity of the target gene prior to use in quantification of the expression of *M. rosenbergii* dystrophin. The sequence obtained is shown in Figure 4.16. Through computational amino acid translation, a 50 amino acids long sequence coding for the target ZZ protein domain of *M. rosenbergii* dystrophin was obtained from the sequenced amplicon.

A) CCCTATTTCCGCATCAGGTCAATGCAATATCTGCAAGGCATATCCTATTGTA
GGCTTACGGTACCGTTGCCTCAAGTGCCTCAGCTTTGATATGTGTCAGCGGT
GCTTCTTTGATGGAAGAAGCGGCAAGAGTCCAAGATCACTCACCCCAAT

B) PISASGQCNI CKAYPIVGLR YRCLKCLSFD MCQRCFFDGR SGKSPRSLTP

Figure 4.16: A) The nucleotide sequence of the amplicon produced through qPCR of *M. rosenbergii* **cDNA targeting the dystrophin gene. B) The translated amino acid sequence of the amplicon produced through qPCR of** *M. rosenbergii* **cDNA.**

4.3.2 Quantification of *M. rosenbergii* dystrophin in infected samples

R e l a t i v e   e x p r e s s i o n   o f   D y s t r o p h i n   i n
r e s p o n s e   t o   i m m u n e   c h a l l e n g e   w i t h   W S S V
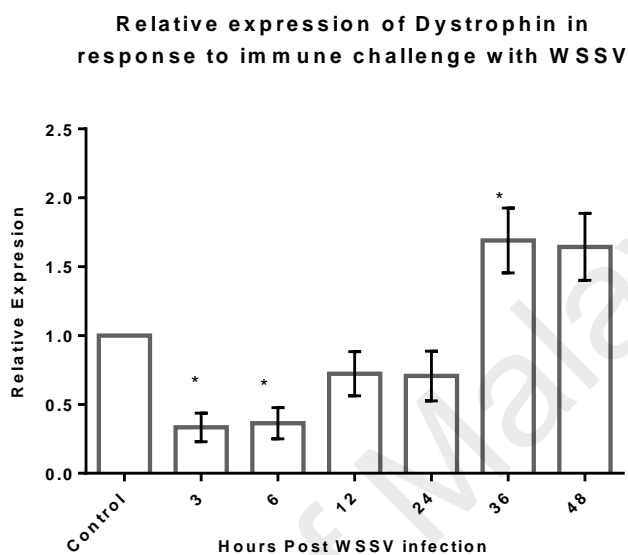
Figure 4.17: Relative quantification of *M. rosenbergii* dystrophin in WSSV infected samples. All quantifications were conducted in triplicates and presented as a mean of the numerical data obtained. Error bars represent Standard Error of Mean (S.E.M) of the three values. Asterisks represent statistical significance of P > 0.005 through ANOVA analysis.

Quantification of *M. rosenbergii* dystrophin expression in response to WSSV infection was conducted through qPCR quantification of the cDNA of infected *M. rosenbergii* at different time points post infection (Figure 4.17). The elongation factor ELF-1 was used as the internal control gene for this relative quantification (Dhar et al., 2009).

Relative expression of *M. rosenbergii* dystrophin in muscle tissues at 3 and 6 hours post injection was decreased by 0.33 fold and by 0.36 fold respectively in comparison to the control sample group. Subsequently, its expression gradually increased from 12 hours

onwards, particularly at 36 hours of post injection where it showed an increase of 1.6 fold relative to the control.

4.4 Quantification of intracellular calcium concentration

intracellular calcium concentration quantification in muscle tissues of *M. rosenbergii*
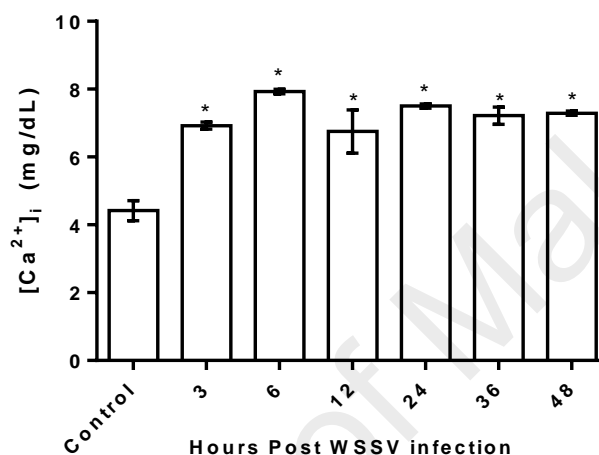
Figure 4.18: Bar chart showing ~~the~~ intracellular calcium concentration in infected *M. rosenbergii* muscle tissues at different time points post infection. All quantifications were conducted in triplicates and presented as a mean of the numerical data obtained. Error bars represent Standard Error of Mean (S.E.M) of the three values. Asterisks represent statistical significance of P > 0.005 through ANOVA analysis.

Quantification of intracellular calcium concentration in the muscle tissue of WSSV infected *M. rosenbergii* showed significant increase between the control samples and at 3 hours post WSSV infection, from 4.5 mg/dL to 6.9 mg/dL (Figure 4.18). A further increase in intracellular calcium concentration was observed until its peak at 7.9 mg/dL at 6 hours post infection This decreased slightly at 12 hours post injection to 6.75 mg/dL, before being maintained at the elevated concentrations of 7.5 mg/dL, 7.22 and 7.29 mg/dL at 24 hours, 36 hours and 48 hours of post injection respectively.

4.5 Transmission Electron Microscope (TEM) imaging of muscle tissues

4.5.1 Muscle tissue imaging

M. rosenbergii muscle sample tissues from the following time points of the WSSV immune challenge were used in TEM imaging of muscle morphology: 0 hours (control), 24 hours and 48 hours post infection. Three samples of muscle tissues were dissected and processed as elaborated in the methodology for each time point. The images were taken at each time point is shown in Figures 4.19 - 4.22. Qualitative comparison of the images obtained was carried out to compare the muscle morphology of the sample tissues, particularly of muscle fiber and mitochondria conditions. The comparisons showed that tissues obtained from the control samples were the least intact, while samples at 48 hours post WSSV infection were affected the most.

This was seen through the conditions of the muscle fibers observed, the alignment of the muscle disks, particularly the z–disk, which is most affected by dystrophin. Disorder in the disk alignments and a degraded state of muscle fibers were seen in the samples at 48 hours post infection, with noticeable breakage in the fiber arrangement. On the other hand, equal and orderly alignment was seen in the control samples.

Similarly, the mitochondria organelles in the tissue samples of healthy M. rosenbergii were intact in the ultra-structural TEM images. Abnormal swellings of the cristae were observed in the mitochondria at 24 hours post infection when compared to healthy samples. Meanwhile, mitochondria cristae at samples 48 hours post WSSV infection were absent, presumably damaged through excessive swelling.
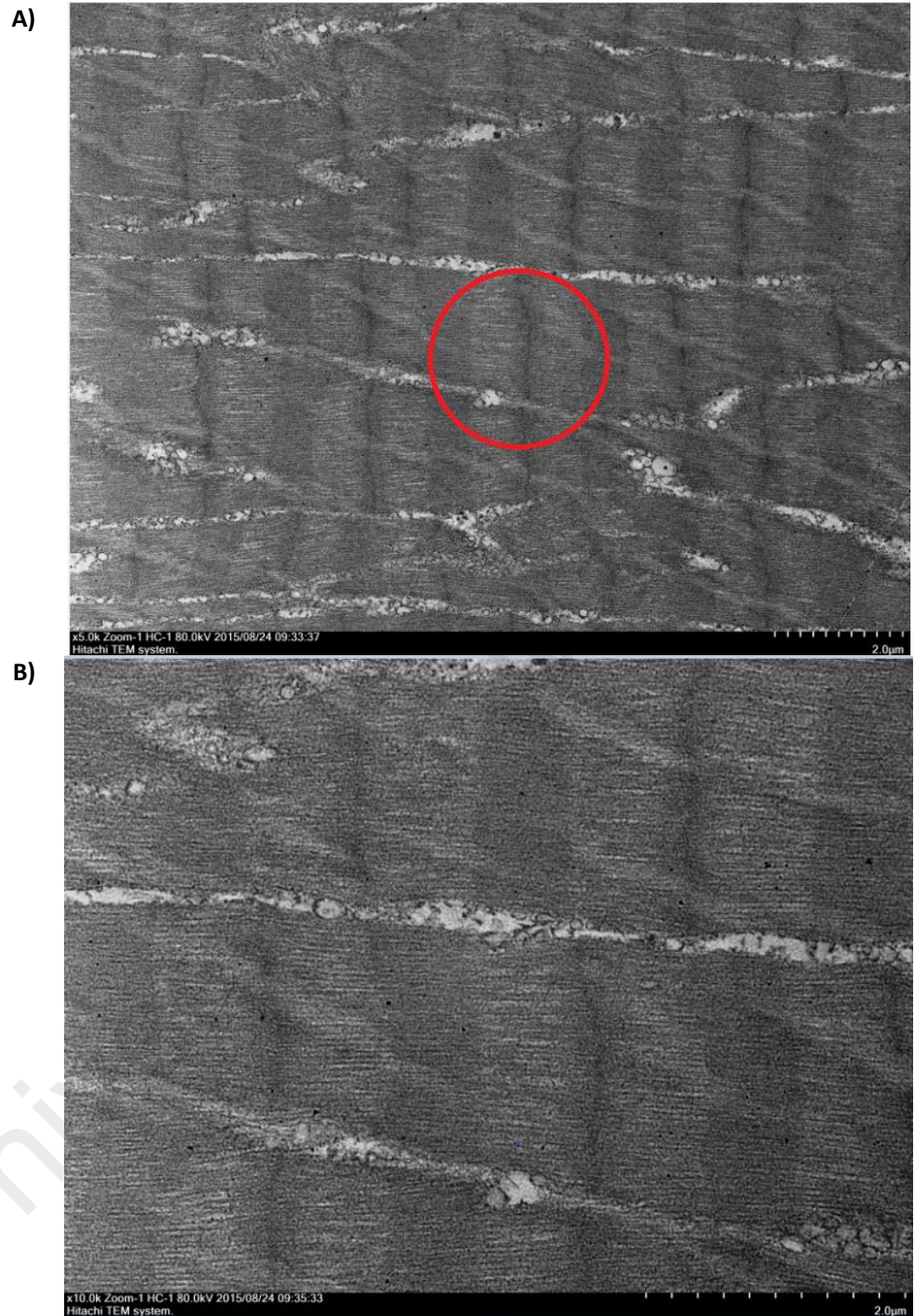
Figure 4.19: Transmission Electron Microscope (TEM) micrographs showing a longitudinal cross section of the muscle morphology of healthy, uninfected *M. rosenbergii* muscle tissues. The first image shows healthy *M. rosenbergii* muscle fibers at a magnification of 5000 X and the second image shows the fibers at a magnification of 10000 X.

.
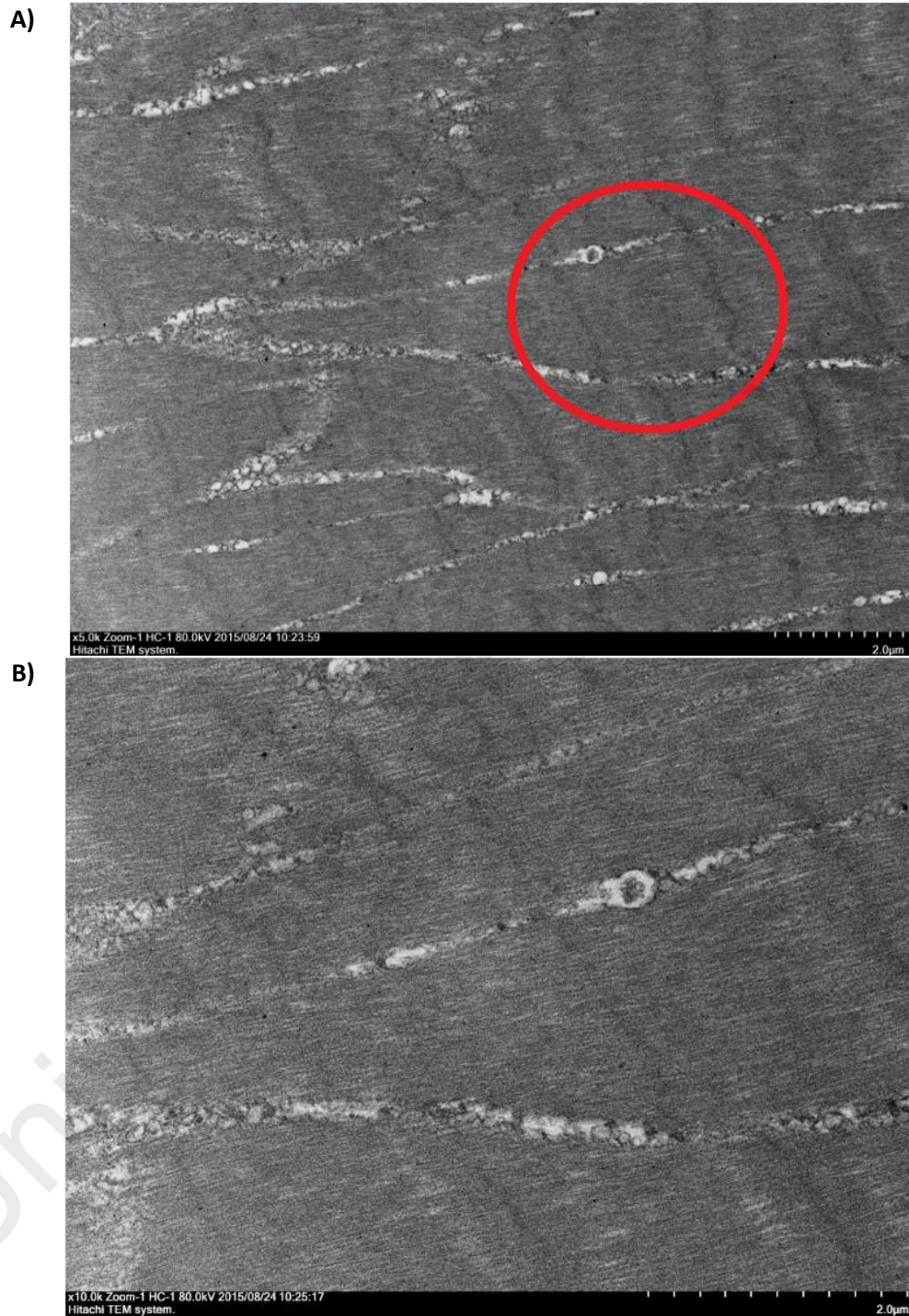
Figure 4.20: Transmission Electron Microscope (TEM) micrographs showing a longitudinal cross section of the muscle morphology of WSSV infected *M. rosenbergii* muscle tissues 24 hours post infection. The first image shows *M. rosenbergii* muscle fibers at a magnification of 5000 X and the second image shows the fibers at a magnification of 10000 X.
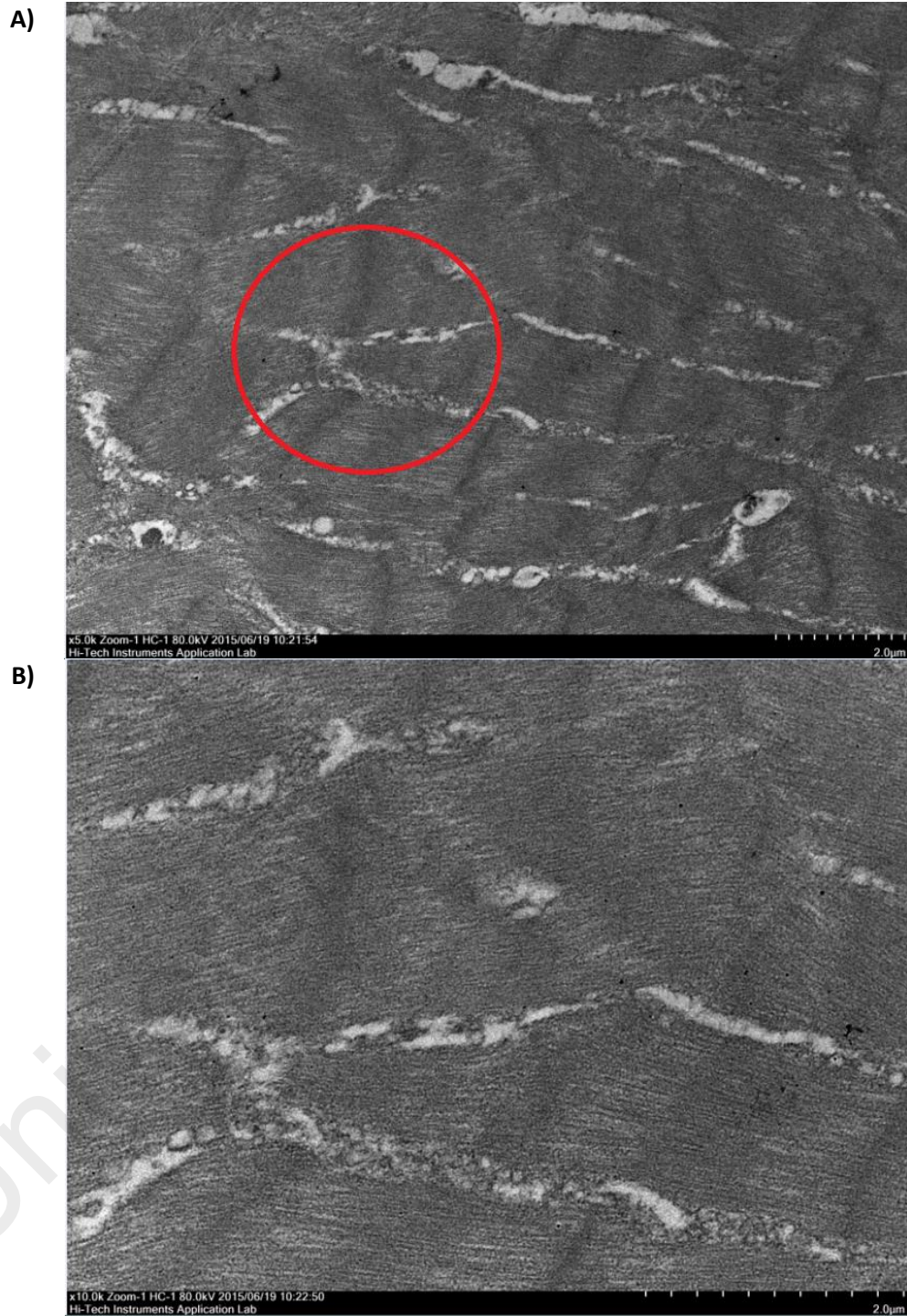
Figure 4.21: Transmission Electron Microscope (TEM) micrographs showing a longitudinal cross section of the muscle morphology of WSSV infected *M. rosenbergii* muscle tissues. The first image shows *M. rosenbergii* muscle fibers at a magnification of 5000 X and the second image shows the fibers at a magnification of 10000 X.
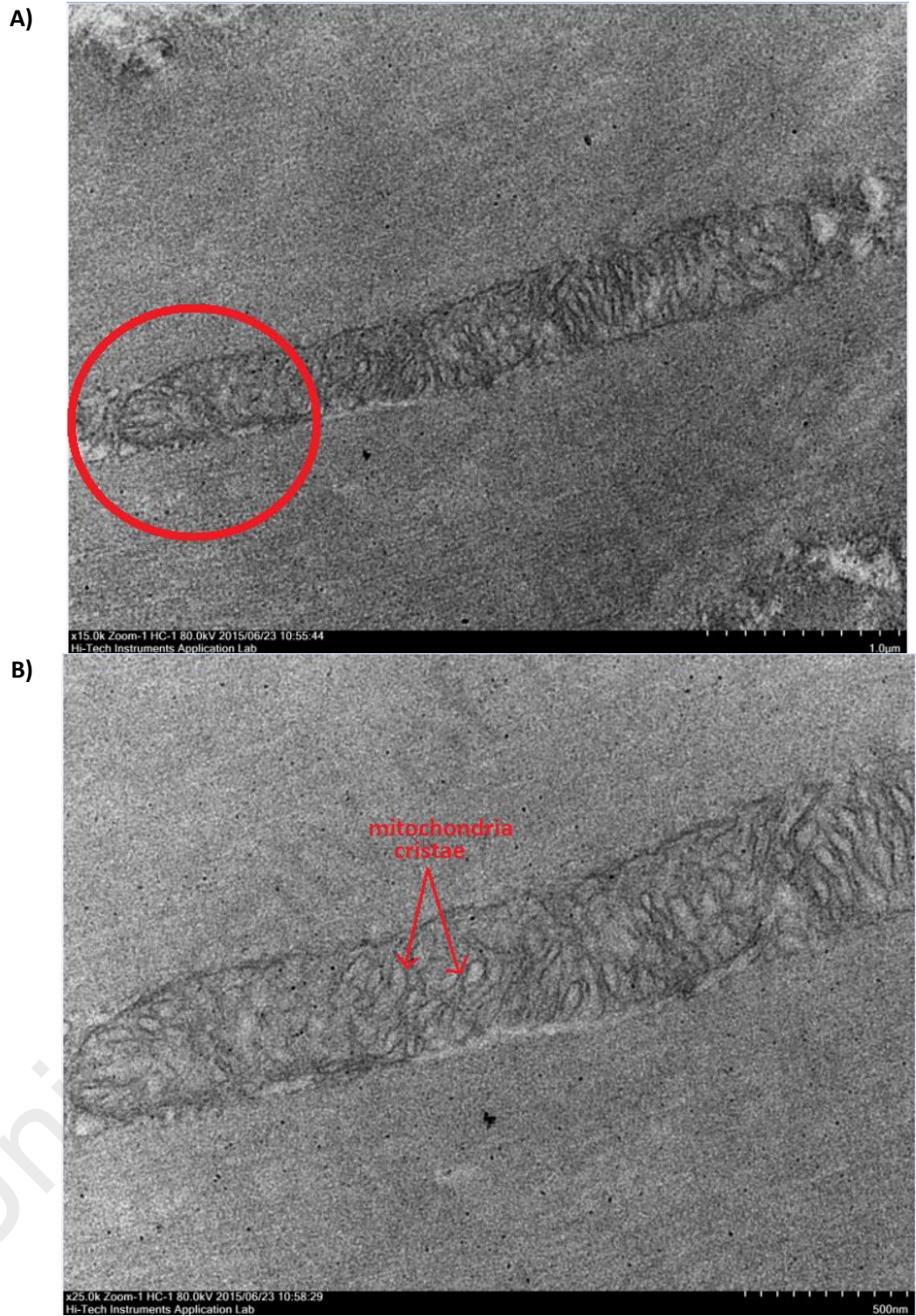
Figure 4.22: **Transmission Electron Microscope (TEM) micrographs showing a mitochondrion organelle in a healthy *M. rosenbergii* muscle tissue. The first image shows the M. rosenbergii ultra-structures at a magnification of 15000 X and the second image fibers at a magnification of 25000 X.**
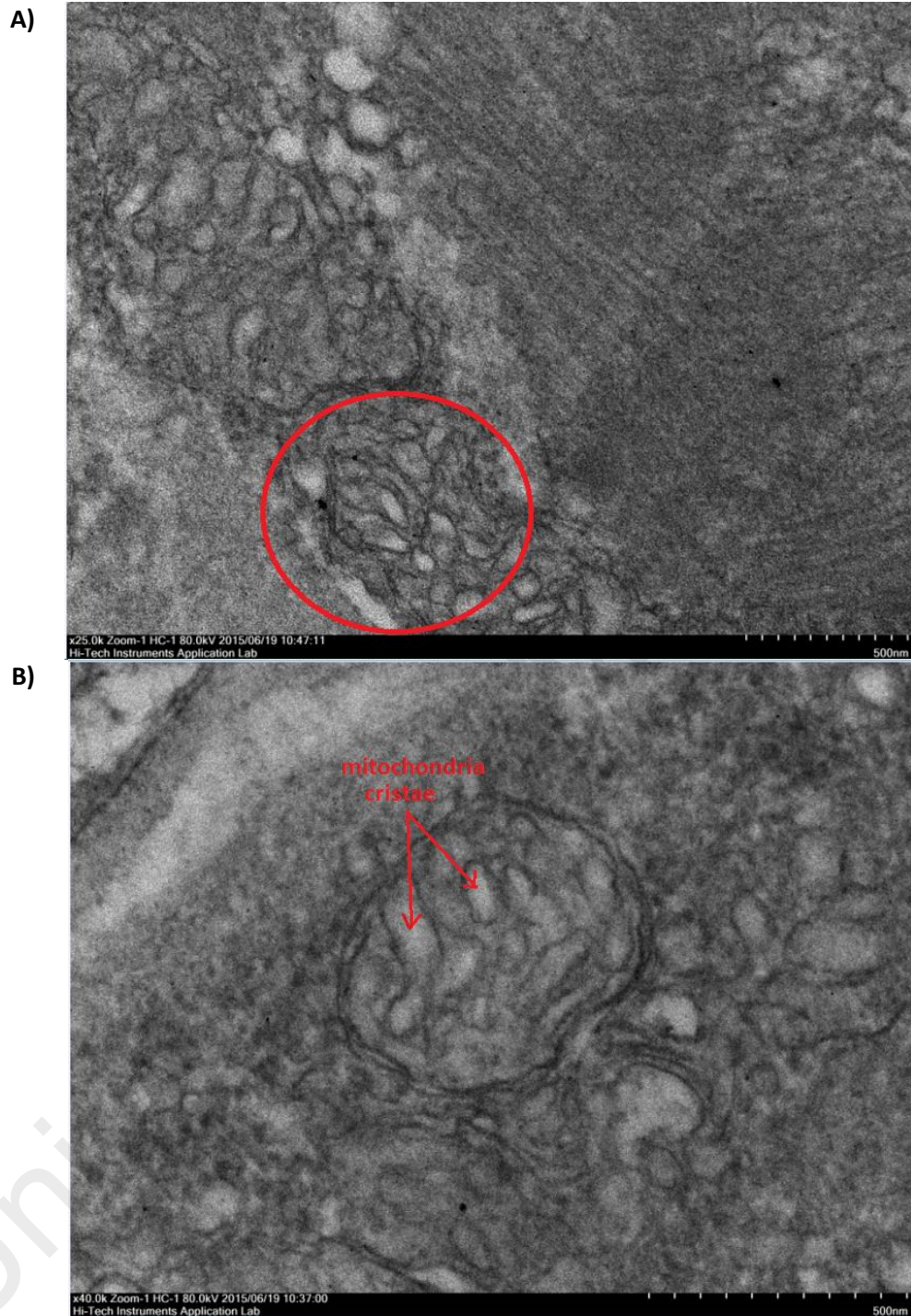
Figure 4.23: Transmission Electron Microscope (TEM) micrographs showing a mitochondrion organelle in WSSV infected *M. rosenbergii* muscle tissues at 24 hours post infection. The first image shows the *M. rosenbergii* ultra-structures at a magnification of 25000X and the second image at a magnification of 40000X.
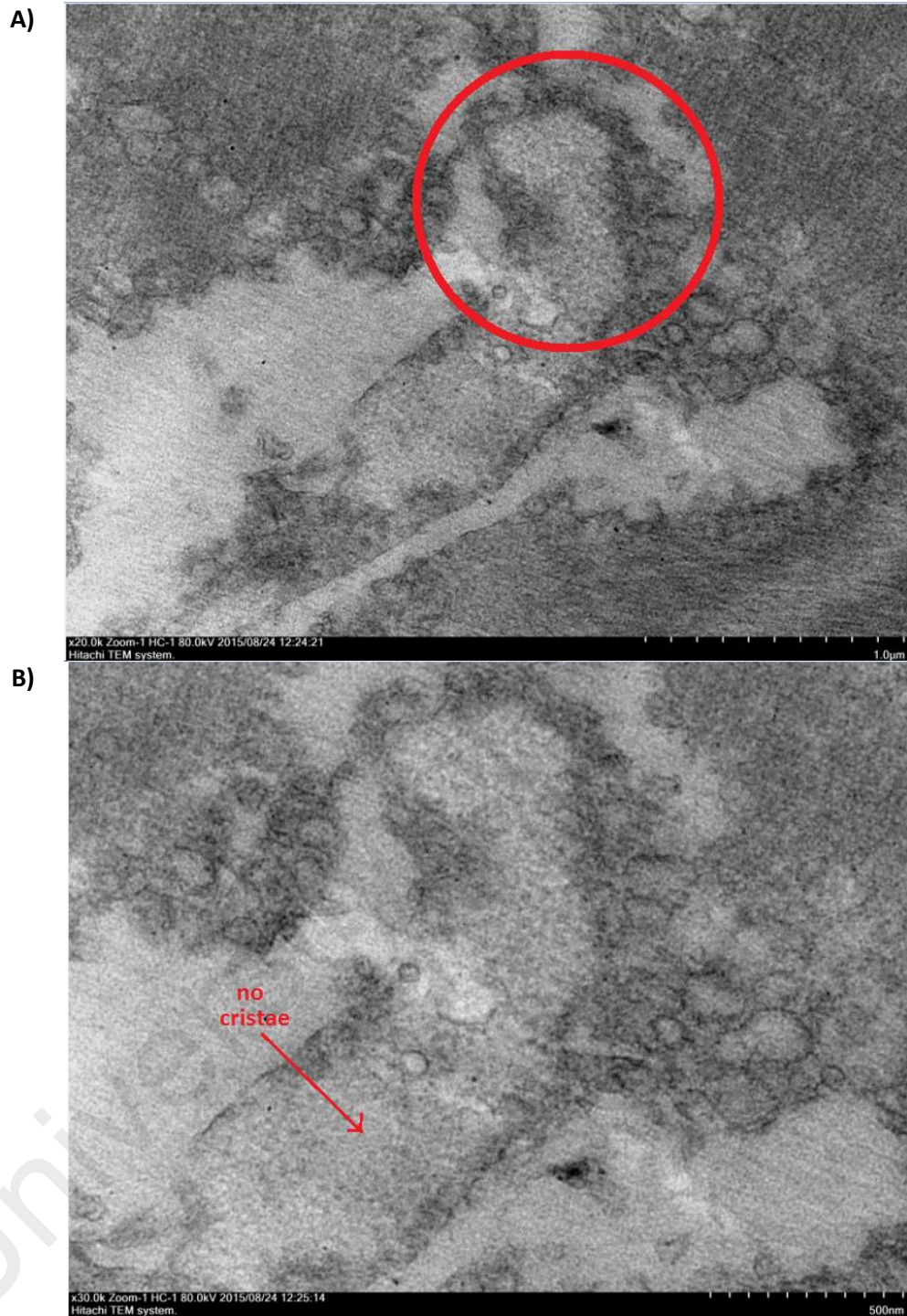
Figure 4.24: Transmission Electron Microscope (TEM) micrographs showing a mitochondrion organelle in WSSV infected *M. rosenbergii* muscle tissues at 48 hours post infection. The first image shows the *M. rosenbergii* muscle ultra-structures at a magnification of 20000 X and the second image fibers at a magnification of 30000 X.

CHAPTER 5 : DISCUSSION

5.1 *Macrobrachium rosenbergii* Dystrophin

A nucleotide sequence encoding the *Macrobrachium rosenbergii* dystrophin was retrieved from the NCBI Short Read Assembly data of *M. rosenbergii* and verified through PCR amplification of *M. rosenbergii* cDNA. The identity of the sequence was further confirmed through a BLAST homology analysis and phylogenetic tree, constructed with dystrophin sequences from various other organisms. Several of the dystrophin sequences with the highest similarities to *M. rosenbergii* dystrophin, in terms of sequence lengths and genetic closeness, such as *Drosophila melanogaster* and *Danio rerio* were further compared by protein alignment and study of predicted physicochemical properties.

Amino acid translation of the *M. rosenbergii* dystrophin sequence produced a 414 amino acids long sequence. This sequence was analyzed to study its physicochemical properties and similar analyses were carried out on homologous *Drosophila melanogaster* and *Danio rerio* as well as the *Homo sapien* dystrophin sequence as comparison. Comparison of physicochemical properties of these dystrophin sequences (Table 4.1), show the longest sequence, the *H. sapien* dystrophin as also the most unstable. However, *M. rosenbergii* dystrophin has a far larger Theoretical pI compared to both of the homolog dystrophin sequences, indicating it is less likely to self-aggregate compared to the other two sequences. It is also the most stable of the dystrophin sequences compared.

The predicted secondary protein structure of *M. rosenbergii* dystrophin (Figure 4.4) was produced with a high confidence of prediction throughout the amino acid sequence. At least

15 different alpha helixes are predicted in this secondary structure, connected by several beta strands and coil loops. This highly accurate prediction method provides a good understanding of the structure of *M. rosenbergii* dystrophin based on just its amino acid sequence, before any further analysis is conducted.

A 3D model of the monomer *M. rosenbergii* dystrophin protein was constructed through target-template alignment to examine its structure (Figure 4.5). Out of the 2572 possible templates found for the *M. rosenbergii* dystrophin amino acid sequence, the template with the highest quality and best alignment was the 1eg3.1A template from the Protein Data Bank (X. Huang et al., 2000). This human dystrophin template profiles key terminal regions of dystrophin, as well as the complex beta-dystroglycan binding site. It shares a 57.36 % identity with the *M. rosenbergii* dystrophin amino acid sequence and has a protein resolution of 2.00 Å, indicating a very small room for possible errors. The alignment produced (Appendix Figure 3) shows 62% sequence coverage with the *M. rosenbergii* dystrophin amino acid sequence between the ranges 104 – 363.

The predicted *M. rosenbergii* dystrophin model showed a dystrophin monomer molecule, with multiple helixes, strands and coils, without any presence of ligands. The model produced obtained a QMEAN4 score of -2.18 and a GMQE score of 0.56. The QMEAN score indicates the statistically significant improvement of the ability of the scoring function to identify the native structure and to discriminate good from bad models. While the GMQE score obtained reflects high reliability and accuracy of the predicted 3D model. These predictions were further validated by plotting all of the protein residue bond angles in a Ramachandran Plot (Figure 4.6).

The Ramachandran Plot constructed visualizes all the angles of the protein backbone, indicating conformations of the polypeptide backbone to the accepted peptide constrictions (Figure 4.6). In this plot of the Phi, Φ and Psi, Ψ angles of protein bonds in the 3D model of *M. rosenbergii* dystrophin, the majority of the residue angles fall in the accepted ranges indicating an ordered protein structure. Almost all of the Proline residues fall within its narrow preferred region of around $\Phi = -60$, indicating good conformity. Similarly, other *M. rosenbergii* dystrophin residues also fall within the contoured areas with very few anomalies, implying a very refined structure. As a majority of the plots on the diagram produced fall within the expected regions, the potential for occurrence of errors in the determination of the position of atoms by computational modeling algorithms is also eliminated.

To assess the quality of both, the 3D protein model and the Ramachandran Plot constructed of *M. rosenbergii* dystrophin, a numerical evaluation was carried out through the protein checking program "WHAT_IF" (Table 4.2). The protein model obtained a backbone conformation Z-score of 0.697, indicating an accurate protein structure model. Meanwhile the Ramachandran plot obtained a Z-score of -0.170, reiterating the conformation of the backbone of all the residues corresponding to the known allowed areas in the plot.

The WHAT_IF protein check on the 3D model constructed also verified that the model had a normal bond angle variability with the RMS-Z score for bond length on 1.242, indicating a high resolution X-ray structure (Parkinson, Vojtechovsky, Clowney, Brunger, & Berman, 1996). The χ_1/χ_2 Rotamer normality score obtained by the constructed protein model expresses how well the χ_1/χ_2 angles of all the residues of the protein correspond to the Protein Data Bank database, while the Inside/Outside distribution score of 1.015 describes

how well the accessibility of the residues observed fi the expected structure. All of these assessments carried out on the predicted 3D protein model of *M. rosenbergii* dystrophin indicate a well refined protein, and verified it suitability to be used in applications with confidence of its accuracy.

Through a conserved domain database analysis, 5 distinct conserved protein domains were seen in the translated protein sequence of *M. rosenbergii* dystrophin. These conserved domains; the WW, EF-Hand-2, EF-Hand-3 and ZZ Dystrophin domains, as well as the Spectrin region, are all well studied components of the human dystrophin protein, with the first four making up the essential central rod region (Jin et al., 2007). The dystrophin central rod region contains two known binding sites, in the WW and ZZ domains respectively. Both binding sites play important roles in the binding of dystrophin to a dystrophin-associated glycoprotein, dystroglycan. The WW domain is a primary binding site for dystrophin and dystroglycan. This interaction is cradled and stabilized by the EF-hand domains of the cysteine rich region (Rentschler et al., 1999).

The ZZ domain coordinates zinc finger motifs in the dystrophin structure, and is recognized as one of the main stabilizing components in the binding of dystrophin to dystroglycan by anchoring these proteins to cell membrane (Hnia et al., 2007; Vulin et al., 2014). This dystroglycan-dystrophin complex binds to intracellular actin muscle cytoskeleton together with extracellular muscle protein, laminin. Its functional role is to maintain the structure integrity of muscle during relaxation and contraction (Hnia et al., 2007). As dystrophin is one of the most important components in building up muscle structure, the presence of these domains in connecting the intracellular cytoskeleton of actin to extracellular matrix is crucial for signalling and maintaining membrane stability (Goldstein & McNally, 2010).

Quantification of *M. rosenbergii* dystrophin expression in this study was conducted using probes targeting the one of these binding sites, namely the ZZ domain.

The NCBI database was searched for dystrophin sequences from other organisms, and 11 such sequences from 5 different phyla were found (Table 3.4). A comparison of the conserved protein domains of all 11 sequences and *M. rosenbergii* dystrophin revealed surprisingly high similarity between the domains found in each sequence despite their origins from 5 different phyla across the animal kingdom (Figure 4.10). The size of the conserved protein domains in each of the dystrophin sequences was found to be proportionate to the genomic size of the organism. The larger sequences of dystrophin retrieved from more complex organisms such as humans and dogs showing a larger portion of the spectrin region. Meanwhile, the short portion of spectrin seen in *M. rosenbergii* dystrophin is similar in proportion to the length of spectrin found in *C. elegans* and *D. melanogaster* dystrophin.

It is also worth noting that despite the vast differences between sequence lengths in mammals (13 960 and 13 887 amino acids respectively in humans and dogs) and arthropods (414 amino acids in *M. rosenbergii*), the size of the EF-hand and ZZ conserved domains in *M. rosenbergii* dystrophin is almost the same length as the mammalian sequences compared. This implies high structural conservation of the gene between vertebrates (mammals) and invertebrates (arthropods). The sequence conservation is graphical shown in Figure 4.9 illustrating the positions of sequence similarities between the *M. rosenbergii* dystrophin amino acid sequence and its counterparts in humans, *D. melanogaster* and *D. rerio*.

The presence of all four characteristic conserved domains of dystrophin, as well as a significant portion of the spectrin region in the relatively shorter *M. rosenbergii* dystrophin sequence is remarkable as most other sequences obtained from the NCBI database lacked one or more of these domains. This suggests potential for the use of *M. rosenbergii* dystrophin in studying the mechanism and functional role of the gene, as well as in dystrophy related studies.

A phylogenetic tree constructed using the same 11 sequences reveal that *M. rosenbergii* dystrophin is most closely related to *Caenorhabditis elegans* dystrophin (Figure 4.8). Given the highly similar proportions of the conserved domains between *M. rosenbergii* dystrophin and *C. elegans* dystrophin sequences as well as their genetic closeness, high functional and structural resemblances of dystrophin could be expected between both organisms. This postulation, if proven true, is highly beneficial in furthering the study of the underlying mechanism of *M. rosenbergii* dystrophin in muscle tissues as *C. elegans* is one of the most studied invertebrate homologue of the gene (Gieseler et al., 2000; Segalat, 2002, 2007; Sparrow, Hughes, & Segalat, 2008). It has also been used extensively as a model organism in the study of muscular dystrophy.

The other arthropod homologue included in the analysis, *D. melanogaster* was also shown in the same clad (Figure 4.8). Other sequences shown to be closely related to *M. rosenbergii* dystrophin through the phylogenetic study are the *Asteroidea sp* and *Pectinidae sp* dystrophin homologues.

5.2 Quantification of White Spot Syndrome Virus (WSSV)

White Spot Syndrome Virus (WSSV) has been established as a pathogenic cause of infection in aquaculture organisms. In order to better understand the relationship between this virus and changes in gene expression of dystrophin, quantification of its viral load changes was conducted in this study. Similarly, detection of the VP28 gene of the virus through PCR in experimentally infected *M. rosenbergii* was used to verify successful infection. The VP28 protein targeted by the PCR probes is vital for virus penetration and its absence leads to observable inhibition of the infection. This VP28 region of the virus is responsible for virus penetration and is broadly used in WSSV detection (Escobedo-Bonilla et al., 2008). All of the challenged *M. rosenbergii* showed positive results for WSSV infection (Figure 4.10 and Figure 4.11) and exhibited several pathological signs of the disease including discoloration of muscles and exoskeleton, lethargy and a drastically reduced appetite.

The viral content of WSSV infected *M. rosenbergii* samples were measured by quantifying the copy number of the VP28 region of the virus through absolute qPCR. Quantification of WSSV load in the infected *M. rosenbergii* muscle tissue did not initially show any significant changes, however increases were observed 24 hours after injection (Figure 4.13). These viral copy number quantification results indicate that the copy number of the injected WSSV in *M. rosenbergii* muscle tissues was not significantly affected in the first 24 hours of the infection. However, a drastic increase in virus copy number was observed at 36 hours, and maintained until 48 hours post infection. A possible assumption for the

observed increment in WSSV copy numbers is the progressive replication of the virus in the host organism.

This delayed onset of WSSV infection in *M. rosenbergii* has been reported in several studies investigating factors contributing to WSSV progression in crustaceans. (Corteel et al., 2012; R B Pramod Kiran, 2002). The parameters maintained throughout the present study, such as water temperature, were also conducive to viral replication and is a possible contributor to the rapid increase in viral copy numbers (Rahman et al., 2006). The observed delayed increase of WSSV copy numbers also corroborate with studies that have demonstrated that the number of infected cells from gills, stomach epithelium, cuticular epithelium and hematopoietic tissue increased after 24 hour WSSV post-infection in *M. rosenbergii* (Corteel et al., 2012). This indicates that the increment of WSSV copy numbers had resulted in increased number of WSSV infected cells from various organs. On the other hand, the virus copy number from 0 to 24 hours post infection did not significantly change, possibly due to the innate immune system of the prawn still being capable of responding to suppress the viral infection.

5.2 Expression of M. rosenbergii dystrophin in response to WSSV infection

The effects of WSSV infection on the expression of *M. rosenbergii* dystrophin was examined by quantifying transcriptional mRNA expressions of *M. rosenbergii* dystrophin in WSSV infected *M. rosenbergii* at a different time intervals post infection. The mRNA expression of *M. rosenbergii* dystrophin in *M. rosenbergii* muscle significantly decreased from 0 hours to 24 hours post-infection to less than half of its expression in the control samples (Figure 4.17). At the same time, WSSV symptoms such as the discoloration of muscles and exoskeleton were also beginning to show. This occurrence of muscle

deterioration with a lowered dystrophin expression corresponds to the functional role of the dystrophin gene in preserving muscle integrity (Shin et al., 2013). Similar muscle deterioration symptoms have also been reported in numerous studies involving WSSV (Leu et al., 2007; Pradeep, Rai, Mohan, Shekhar, & Karunasagar, 2012; Shekhar & Ponniah, 2015). This further supports the postulation of the relationship between muscle deterioration during WSSV infection and reduced dystrophin expression.

Quantification of *M. rosenbergii* dystrophin was conducted through relative qPCR targeting the ZZ domain of the sequence with the ELF-1 gene as a reference gene (Dhar et al., 2009). Immediately upon WSSV infection, the expression of *M. rosenbergii* dystrophin in muscle mRNA showed a sharp drop to almost one-third of its expression in control samples. Although the expression increases slightly, it does not reach the basal expression levels of the control samples until 36 hours post infection. The drop in *M. rosenbergii* dystrophin expression until 24 hours post WSSV infection points towards immune response associated suppression of the gene. The restoration and subsequent rise in *M. rosenbergii* dystrophin expressions after 24 hours could be attributed to the giant freshwater prawn's tendency to exhibit symptoms of WSSV infection only after a period of time. In-depth immunohistochemical studies *of M. rosenbergii* samples similarly injected with a high dosage of WSSV only begin to show WSSV-infected cells after 24 hours post infection (Corteel et al., 2012). This leads to the postulation that the virus pathway of WSSV infection plays a role in the up-regulation of *M. rosenbergii* dystrophin expression in muscle tissues.

Possible linkages between pathogenic infections and the mechanisms of muscular dystrophy have been explored in-depth in the quest for new approaches to muscular

dystrophy treatments. (Emery, 2008). Viral infections have been demonstrated to affect to induce the cleaving of various muscle cytoskeletal protein, accounting for some of the observable clinical symptoms of the infection (Emery, 2008). One such viral infection is WSSV, and the observed decrease in dystrophin expression in the present study corroborates with this theory.

5.3 Protein cleavage analysis of M. rosenbergii dystrophin

The observed decrease in dystrophin expression upon WSSV infection is possibly due to protein cleavage of the sequence by increased proteinase expression as a result of the infection. Further study on this postulation was carried by looking deeper into the potential cleavage sites by proteinase on the dystrophin sequence. Particular attention was given to Chymotrypsin which has previously been reported to have an exponentially increased expression in WSSV infected shrimp (Xue et al., 2013), with a 14 times up-regulation compared to the un-infected control group. It is further reported that in-vivo knockdown of the proteinase expression delays shrimp mortality and reduces WSSV copy number, highlighting its importance in the mechanism of the infection.

In the present study, 52 out of 123 possible Chymotrypsin cleavage sites showed a cleavage occurrence probability of 51% or more. The concurrent occurrences of reduction in dystrophin expression and the reported increase in Chymotrypsin indicate that more frequent cleaving of *M. rosenbergii* dystrophin occurs due to WSSV infection, as reflected by the quantification of *M. rosenbergii* dystrophin shown in Figure 4.17. This also explains the observed increase in mRNA expression of *M. rosenbergii* dystrophin despite increases in copy number of WSSV and worsening muscle condition of infected prawn samples at 36

and 48 hours post-infection. It is likely that the dystrophin mRNA expressed is not able to be fully translated into protein, or is digested by the WSSV induced increase in proteinase expression.

5.4 Intracellular Calcium Concentration in response to WSSV infection

Significantly increased intracellular calcium concentration was observed in the muscle tissues of *M. rosenbergii* from 4.5 mg/dL in the control sample to 8 mg/dL at 6 hours post WSSV infection, through florescent based quantification by calibration to a standard calcium solution (Figure 4.18). This elevated level of intracellular calcium concentration was maintained with very slight fluctions until the end of the study at 48 hours post WSSV infection. This sharp rise in intracellular calcium concentration occurs concurrently with the decrease in *M. rosenbergii* dystrophin expression in WSSV infected muslce tissues, supporting existing throries that both these changes are related. The highest intracellular calcium concentration is observed at the same time as the lowest *M. rosenbergii* dystrophin expressions. Similarly, an increase in *M. rosenbergii* dystrophin expression concurs with the stabilizaton of intracellular calcium concentration in *M. rosenbergii* muscle tissue. These observations suggest that the lowered expression of dystrophin in *M. rosenbergii* has a similar pathophysiological response as those observed in human muscles and other dystrophin deficient model organisms.

Several explanations have been offered to explain intracellular calcium concentration increase in dystrophin deficient muscles, such as that seen in the present study. As it is generally accepted that dystrophin and its associated protein complexes play a role in stabilizing the cell membrane, it is predicted that in the absence of dystrophin, membrane

damage is more frequent (Miyake & McNeil, 2003; Petrof et al., 1993). This leads to leakage or influx of calcium into the cells through damaged membrane. (Mallouk et al., 2000; Yeung et al., 2005). Additionally, identification of a surface membrane calcium channel with increased activity in dystrophin deficient muscles led to the suggestion that the increased intracellular calcium concentration is due to calcium ion entry through this channel (Allen et al., 2010). Furthermore, the relationship between increased calcium presence and muscle weakness has often been explained by the activation of calpains, a calcium-activated enzyme whose role in skeletal muscle protein degradation is well documented (Ma et al., 2011; Sorimachi et al., 2011; Verburg et al., 2006).

5.4 Transmission Electron Microscope observation of *M. rosenbergii* muscle tissues

Transmission Electron Microscope (TEM) observations were carried out on healthy and WSSV infected *M. rosenbergii* muscle tissues at 24 and 48 hours post WSSV infection. Comparisons of the muscle morphology demonstrated that tissues from the control samples were the least intact, while samples at 48 hours post infection were affected the most.

Two different aspects of the muscle tissues were observed, the longitudinal cross section layout of muscle tissues, and the mitochondria found in between the muscle fiber bundles. The longitudinal cross section view was selected for observation as it gives a very profound perspective on the lay and uniformity of the muscle bundles. The muscle bundles were compared based on the alignment of the bundles, inter-bundle space and general appearance of organelles. On the other hand, the comparison criteria used for muscle fibers are the appearances and uniformity of bands, as well as the alignment of muscle fibers within its bundles. Meanwhile, observations of the mitochondria ultra-structure focused on the structure of cristae and membrane within the organelle.

In the longitudinal cross section comparison, two different magnifications, 5000 X and 10000 X, were used to compare the muscle fibers and bundles. The micrograph of longitudinal cross section of healthy *M. rosenbergii* muscle tissues are shown in Figure 4.19, showcasing uniformed and aligned muscle fibers, as well as highly distinct well preserved bands. Micrographs obtained of WSSV infected *M. rosenbergii* muscle tissues 24 hours post infection (Figure 4.20) shows less distinct muscle bands and visible disturbance in its alignments. Upon increasing the magnification (Figure 4.20 B), the appearance of muscle fibers within each bundle also appeared to be less orderly compared to the healthy tissues. At 48 hours post WSSV infection, disturbances in the arrangements of muscle bundles, bands and fibers were very clearly observable even at a low magnification (Figure 4.2). At a magnification of 10000 X, this observation was further iterated by the high level of disorder observed among the muscle fibers.

This progressive deformation of muscle tissue fibers concur with earlier discussed theories on the relationship between muscular changes observed during WSSV infection and the role of the dystrophin gene in *M. rosenbergii.* The worsening muscle tissue conditions seen post infections correspond to the WSSV induced decrease in dystrophin expression. For instance, reduced dystrophin expression at 24 hours post infection (Figure 4.17) is reflected by the observance of disturbances to band alignments seen in the muscle tissues. Further degradations in the state of the muscle tissues are observed at 48 hours post WSSV infection, corresponding to the elevation of protein cleavage in dystrophin, in response to WSSV infection.

Meanwhile, progressive damages were also observed in the mitochondria structure of *M. rosenbergii.* The mitochondrial cristae in healthy *M. rosenbergii* tissues appear well

97

defined and organized within the organelle membrane (Figure 4.22). Swelling of the cristae was observed in muscle tissues 24 hours post WSSV infection, with the cristae structures appearing visible larger and rounded (Figure 4.23). On the other hand, mitochondria organelles seen in muscle tissues 48 hours post WSSV infection completely lacked cristae (Figure 4.44). The micrographs obtained also indicated signs of mitochondria bursting. This, as well as the excessive swellings observed at 24 hours post infection, suggests that WSSV infection leads to swelling and eventually bursting of the cristae structure within the organelle. This destruction of the mitochondria would severely impair the muscle tissue's ability to generate and utilize energy efficiently, as indicated by the blatant muscle weaknesses and lethargy observed in infected *M. rosenbergii* prawn samples.

CHAPTER 6 : CONCLUSION

The dystrophin sequence in *Macrobrachium rosenbergii* contains all four characterized domains of the human dystrophin gene, as well as the coiled spectrin region. The sequence codes for a well-defined protein structure containing multiple beta sheets and coils. It also shares significant similarities with several dystrophin sequences from other organisms including *Drosophila melanogaster* and *Danio rerio* (zebrafish) in the protein comparisons conducted.

Upon WSSV infection, the expression of dystrophin in muscle tissues of *M. rosenbergii* shows a significant drop before increasing to a significantly higher level that the pre-infection expression level. Meanwhile, its intracellular calcium concentration increases to almost twice of the control group's concentrations and this calcium concentration is maintained until the later stages of the infection. TEM images of the *M. rosenbergii* muscle morphology show distinct differences in muscle fibre and band arrangement between the healthy and infected muscle tissue samples. Damages in the mitochondria of infected *M. rosenbergii* were also observed with worsening severity over the course of WSSV infection.

This study is the first report of the dystrophin gene in *M. rosenbergii* as well as its differential expression in response to WSSV immune challenge. The discussions on the relationships between *M. rosenbergii* dystrophin expression, WSSV copy number, intracellular calcium concentration and ultrastructural changes in muscle morphology explored in this study establishes valuable connections between the observed changes.

Further investigations could be undertaken by in-depth biochemical and structural genomic studies on the mechanism of the dystrophin gene, WSSV infection and intracellular calcium concentration. This would verify the effect of WSSV infection on dystrophin and further explore the potentiality of the similarities between *M. rosenbergii* and human dystrophin highlighted here. It is also recommended to extend the immune challenge duration of WSSV infection for a prolonged observation on its pathophysological changes when replicating this study in *M. rosenbergii* or in other WSSV prone crustaceans

LIST OF PUBLICATIONS AND PAPERS PRESENTED

1) Identification of a dystrophin-like gene in *Machrobrachium rosenbergii.* Poster presented at the 19th Biological Sciences Graduate Congress (BSGC), 12-14th December 2014, National University of Singapore, Singapore.

2) Dystrophin gene expression and intracellular calcium changes in the giant freshwater prawn, *Macrobrachium rosenbergii*, in response to White Spot Syndrome Virus (WSSV). Publication under review by Skeletal Muscle Journal.

Table 4: The nucleotide sequence f amplicons obtained through PCR to validate the *Macrobrachium rosenbergii* dystrophin sequence.

Primer name	Amplicon sequence
D1 F	AAAAGACTGCCGATGGGTGGGGCCCCATCTCCCAAGACCCGAACCTCGCAGATCAAC ACGCCCACAATGTTAGAAAATTCCGCGAAGGTTTGGCGGAATTGCAGCGCGGAGTGG ACGACGTCAACGACCAAGCCGCTCGGTTTTCGGCTCATAATGTGCCACTCACCCCTGC TAACTCGGCCAAGCTCCATGACCTCAATAATAGATGGAAAAATTTAGAAACAGCAGTC AATGACAGATGGCGACAAGTGGCCAGCAGATCTCGAGAGGTCACACCACTCACCCCT GCCCAGTTGGCATCTTCAGTGTCACCACCATGGGAGAGGGCTACAACTAGTAATAAA GTGCCTTATTACATCAATCATGATCAGGAGAGTACCC
D1 R	TACTCTCCTGATCATGATTGATGTAATAAGGCACTTTATTACTAGTTGTAGCCCTCTCCC ATGGTGGTGACACTGAAGATGCCAACTGGGCAGGGGTGAGTGGTGTGACCTCTCGA GATCTGCTGGCCACTTGTCGCCATCTGTCATTGACTGCTGTTTCTAAATTTTTCCATCTA TTATTGAGGTCATGGAGCTTAGCCGAGTTAGCAGGGGTGAGTGGCACATTATGAGCC GAAAACCGAGCGGCTTGGTCGTTGACGTCGTCCACTCCGCGCTGCAATTCCGCCAAAC CTTCGCGGAATTTTCTAACATTGTGGGCGTGTTGATCTGCGAGGTTCGGGTCTTGGGA GATGGGGCCCCACCCATCGGCAGTCTTTT
D2 F	TTTGGCGGAATTGCAGCGCGGAGTGGACGACGTCAACGACCAAGCCGCTCGGTTTTC GGCTCATAATGTGCCACTCACCCCTGCTAACTCGGCCAAGCTCCATGACCTCAATAATA GATGGAAAAATTTAGAAACAGCAGTCAATGACAGATGGCGACAAGTGGCCAGCAGA TCTCGAGAGGTCACACCACTCACCCCTGCCCAGTTGGCATCTTCAGTGTCACCACCATG GGAGAGGGCTACAACTAGTAATAAAGTGCCTTATTACATCAATCATGATCAGGAGAG TACCCACTGGGATCATCCAATCATGATGGACTTACTAGACTCTATGAATGAATTCAACC ATATCAAGTTCTCTGCCTACAGAACAGCTACAAAACTTAGAATGTTACAAAAGAAGCT GTCATTGGATCTTGCCAAACTGCAAATGGCAACAGATGTATTTGATGAGCATGGTCTC AGGGGACAAAATGACCGGCTAATTGATGTAGGTGATATGGTCGTGGTTTTGTCAGCT TTATATGCAAATATATCTGCTGATCATCCAGAAGTTAATACCACCCTTGCCATAGACCT CTGTCTCAATTGGCTTTTAAATGTATATGACAGCCAAAGGACAGGGCAGATGAGAGTT CTTTCTTTCAAGATCGGTCTCGTCTGCCTTTGCTGCGGTCATTTAGAAGAAAAATACAG ATACATGTTCCGGCTAATTTTT
D2 R	TTTTTAGCCGGAACATGTATCTGTATTTTTCTTCTAAATGACCGCAGCAAAGGCAGACG AGACCGATCTTGAAAGAAAGAACTCTCATCTGCCCTGTCCTTTGGCTGTCATATACATT TAAAAGCCAATTGAGACAGAGGTCTATGGCAAGGGTGGTATTAACTTCTGGATGATC AGCAGATATATTTGCATATAAAGCTGACAAAACCACGACCATATCACCTACATCAATT AGCCGGTCATTTTGTCCCCTGAGACCATGCTCATCAAATACATCTGTTGCCATTTGCAG TTTGGCAAGATCCAATGACAGCTTCTTTTGTAACATTCTAAGTTTTGTAGCTGTTCTGT AGGCAGAGAACTTGATATGGTTGAATTCATTCATAGAGTCTAGTAAGTCCATCATGAT TGGATGATCCCAGTGGGTACTCTCCTGATCATGATTGATGTAATAAGGCACTTTATTAC TAGTTGTAGCCCTCTCCCATGGTGGTGACACTGAAGATGCCAACTGGGCAGGGGTGA GTGGTGTGACCTCTCGAGATCTGCTGGCCACTTGTCGCCATCTGTCATTGACTGCTGTT TCTAAATTTTTCCATCTATTATTGAGGTCATGGAGCTTGGCCGAGTTAGCAGGGGTGA GTGGCACATTATGAGCCGAAAACCGAGCGGCTTGGTCGTTGACGTCGTCCACTCCGC GCTGCAATTCCGCCAAAAA

D3 F	GGTGAATCGCGCGTCGTGGAGACGTCACGACCAAGCCGCTCGGTTTTCGGCTCATAA TGTGCCACTCACCCCTGCTAACTCGGCCAAGCTCCATGACCTCAATAATAGATGGAAA AATTTAGAAACAGCAGTCAATGACAGATGGCGACAAGTGGCCAGCAGATCTCGAGAG GTCACACCACTCACCCCTGCCCAGTTGGCATCTTCAGTGTCACCACCATGGGAGAGGG CTACAACTAGTAATAAAGTGCCTTATTACATCAATCATGATCAGGAGAGTACCCACTG GGATCATCCAATCATGATGGACTTACTAGACTCTATGAATGAATTCAACCATATCAAGT TCTCTGCCTACAGAACAGCTACAAAACTTAGAATGTTACAAAAGAAGCTGTCATTGGA TCTTGCCAAACTGCAAATGGCAACAGATGTATTTGATGAGCATGGTCTCAGGGGACA AAATGACCGGCTAATTGATGTAGGTGATATGGTCGTGGTTTTGTCAGCTTTATATGCA AATATATCTGCTGATCATCCAGAAGTTAATACCACCCTTGCCATAGACCTCTGTCTCAA TTGGCTTTTAAATGTATATGACAGCCAAAGGACAGGGCAGATGAGAGTTCTTTCTTTC AAGATCGGTCTCGTCTGCCTTTGCTGCGGTCATTTAGAAGAAAAATACAGATACATGT TCCGGCTAATTGCCGACCCAAATCGCCTAGTTGATCAAAGAAAATTAGGGTTATTGTT ACACGACTGTGTACAAGTTCCACGCCAGTTGGGTGAGGTTGCAGCATTTGGTGGATC AAACATTGAACCATCAGTTCGTAGCTGTTTCACAAAAGCTGGAAAGGACAGAGAAAC TATTGAGGCTGTGCATTTCTTGGCATGGGTGCAACAAGAACCCCAGTCACTAGTGTGG TTAGCTGTTTTGCATCGTGTTGCAGCCTCAGAAAAATATTCAGCATCAGGTCAAATGCA ATATCTGCAAGGGATATCCTATTTGTAGGCTTACGGTACCGTTGCCTCAAGTGCCTCA GCTTTGATATGTGTCAGCGGTGCTTCTTTGATGGAAGAACCGGCAAAGTCCAAACTCC CCCTCCCCCCCACGAGAAAA
D3 R	TAGAGGCGTCGTCATTTGAATGTAGCTCAGGTTGTGGTTATGACACCGATTAAGCATT TTGGTCTCCCTCCTGGATTTGTTATCTTTTCTGTTTAATGCCTGAATTGTGCCCAAAAGG ATACCTAATTCGTATTTTTCGTAGGTGAACGTTGAAAAACACTACAGTTTGGGGAAGT CAAAATTTTAGAGCCGACAGGCGTACATTTTGCACCGTGCCACTGTTCTTTCATTCCTA TCAATGATACTGACATGACTTTGGTCCGGTGGGAGTTCCACCAATTGGAAACTTCCTA AATAACTGCCTTGTTGCTTTGACATTTTCCTGACCAAATAGCCTTCCTTCCATTTGATCA ACCAGTCGATTTGGCCCCCCCGCTGGATTTAAACATGACTTAGAATTTTCCCCCATGATC ACAACAGGAAATGTGAAGGATTCCAATATTGAAAGAGTGAACCTTGGACAACCCTGC TCATTGTATGAAATATACAGGGAATTTCCCATTGACACAGAGGTCTATGGCAAGGGTG GTAATAACTCCTGGATGATCAGCAGATATATTTGCATATAAAGCTGACAAAACCACGA CCATATCACCTACATTTATTAGCCGGTTTTTTTGTCCCCTGAGACCATGCTCATCAAATA CATCTGTTGCCATTTGCAGTTTGGCTTGATCCAATGACAACCCCCTTTT

```
D1F    AAAAGACTGCCGATGGGTGGGGCCCCATCTCCCAAGACCCGAACCTCGCAGATCAACACG
D2F    ------------------------------------------------------------
D3F    ------------------------------------------------------------
D3R    --------------------------------------------------TAGAGGCG
D1R    ------------------------------------------------------------
D2R    ------------------------------------------------------------


D1F    CCCACAATGTTAGAAAATTCCGCGAAGGTTTGGCGGAATTGCAGC-GCGGAGTGGACGAC
D2F    ------------------------------TTTGGCGGAATTGCAGC-GCGGAGTGGACGAC
D3F    -----------------------------GGTGAATC-GCGCGTCGTGGAGA
D3R    TCGTC-----------ATTTGAATGTAGCTCAGGTTGTGGTTATGACACCGATTAAGCATT
D1R    ------------------------------------------------------------
D2R    ------------------------------------------------------------


D1F    GTCAACGACC-------------AAGCCGCTCGGTTTTCGGCTCATAATGTGCCACTCACC
D2F    GTCAACGACC-------------AAGCCGCTCGGTTTTCGGCTCATAATGTGCCACTCACC
D3F    CGTCACGACC-----------AAGCCGCTCGGTTTTCGGCTCATAATGTGCCACTCACC
D3R    TTGGTCTCCCTCCTGGATTTGTTATCTTTTCTGTTTAATGCCTGAATTGTGCCCAAAAGG
D1R    ------------------------------------------------------------
D2R    ------------------------------------------------------------


D1F    CCTGCTAACTCGGCCAAGCTCCATGACCTCAATAATAGATGGAAAAAATTTAGAAACAGCA
D2F    CCTGCTAACTCGGCCAAGCTCCATGACCTCAATAATAGATGGAAAAAATTTAGAAACAGCA
D3F    CCTGCTAACTCGGCCAAGCTCCATGACCTCAATAATAGATGGAAAAAATTTAGAAACAGCA
D3R    ATACCTAATTCG------------------TATTTTTCGTAGGTGAACGTTGAAAAAACAC
D1R    ------------------------------------------------------------
D2R    ------------------------------------------------------------


D1F    GTCAATGACAGATGGCGACAAGTGGCC-----------AGCAGATCTCGAGAGGTCACACC
D2F    GTCAATGACAGATGGCGACAAGTGGCC-----------AGCAGATCTCGAGAGGTCACACC
D3F    GTCAATGACAGATGGCGACAAGTGGCC-----------AGCAGATCTCGAGAGGTCACACC
D3R    TACAGTTTGGGGAAGTCAAAATTTTAGAGCCGACAGGCGTACATTTTGCACCGTGCCACT
D1R    ------------------------------------------------------------
D2R    ------------------------------------------------------------


D1F    ACTCACCCCTGCCCAGTTGGCATCTTCAGTGTCACCACCATGG---GAGAGGGCTACAAC
D2F    ACTCACCCCTGCCCAGTTGGCATCTTCAGTGTCACCACCATGG---GAGAGGGCTACAAC
D3F    ACTCACCCCTGCCCAGTTGGCATCTTCAGTGTCACCACCATGG---GAGAGGGCTACAAC
D3R    GTTCTTTCATTCCTATCAATGAT----ACTGACATGACTTTGGTCCGGTGGGAGTTCCAC
D1R    ------------------------------------------------------------
D2R    ------------------------------------------------------------
```

```
D1F     TAGT------------AATAAAGTGCCTTATTACATCAATCATGATCAGGAGAGTAC---
D2F     TAGT------------AATAAAGTGCCTTATTACATCAATCATGATCAGGAGAGTAC---
D3F     TAGT------------AATAAAGTGCCTTATTACATCAATCATGATCAGGAGAGTAC---
D3R     CAATTGGAAACTTCCTAAATAACTGCCTTGTTGCTTTGACATTTTCCTGACCAAATAGCC
D1R     ------------------------------------------------------------
D2R     ------------------------------------------------------------


D1F     ------CC----------------------------------------------------
D2F     -----CCACTGGGATCATCCAATCATGATGGACTTACTAGACTCTATGAATGAATTCAAC
D3F     -----CCACTGGGATCATCCAATCATGATGGACTTACTAGACTCTATGAATGAATTCAAC
D3R     TTCCTTCCATTTGATCAACCAGTCGATTTGGCCCCC-------CGCTGGATTTAAACATG
D1R     ------------------------------------------------------------
D2R     ----------------------------------------TTTTTAGCCGGAACATG


D1F     ------------------------------------------------------------
D2F     CATATCAAGTTCTCTGC-------CTACAGAACAGCTACAAAACTTAGAATGTTACAAAA
D3F     CATATCAAGTTCTCTGC-------CTACAGAACAGCTACAAAACTTAGAATGTTACAAAA
D3R     ACTTAGAATTTTCCCCCATGATCACAACAGGAAATGTGAAGGATTCCAATATTGAAAGAG
D1R     ------------------------------------------------------------
D2R     TATCTGTATTTTTCTTCTAAATGACCGCAGCAAAGGCAGACGAGACCGATCTTGAAAGAA


D1F     ------------------------------------------------------------
D2F     GAAGCTGTCATT-------------GGATCTTGCCA------AACTGCAAATGGCAACAG
D3F     GAAGCTGTCATT-------------GGATCTTGCCA------AACTGCAAATGGCAACAG
D3R     TGAACCTTGGACAACCCTGCTCATTGTATGAAATATACAGGGAATTTCCCATTGACACAG
D1R     ------------------------------------------------------------
D2R     AGAACTCTCATCTGCCCTGTCCTTTGGCTGTCATATACATTTAAAAGCCAATTGAGACAG


D1F     ------------------------------------------------------------
D2F     ATGTATTTGATGAGCATGGTCTCAGGGGACAAAATGACCGGCTAATTGATGTAGGTGATA
D3F     ATGTATTTGATGAGCATGGTCTCAGGGGACAAAATGACCGGCTAATTGATGTAGGTGATA
D3R     AGGTCTATGGCAAGG---------------------------------------------
D1R     ------------------------------------------------------------
D2R     AGGTCTATGGCAAGG---------------------------------------------


D1F     ------------------------------------------------------------
D2F     TGGTCGTGGTTTTGTCAGCTTTATATGCAAATATATCTGCTGATCATCCAGAAGTTAATA
D3F     TGGTCGTGGTTTTGTCAGCTTTATATGCAAATATATCTGCTGATCATCCAGAAGTTAATA
D3R     --GTGGTAATAACTCCTGGATGATCAGCAGATATATTTGCATATAAAGCTGACAAAACCA
D1R     ------------------------------------------------------------
D2R     --GTGGTATTAACTTCTGGATGATCAGCAGATATATTTGCATATAAAGCTGACAAAACCA
```

```
D1F     ------------------------------------------------------------
D2F     CCAC----------------------------------------------CCTTGCCAT
D3F     CCAC----------------------------------------------CCTTGCCAT
D3R     CGACCATATCACCTACATTTATTAGCCGGTTTTTTTTGTCCCCTGAGACCATGCTCATCAA
D1R     ------------------------------------------------------------
D2R     CGACCATATCACCTACATCAATTAGCCGGTCATTTTGTCCCCTGAGACCATGCTCATCAA


D1F     ------------------------------------------------------------
D2F     AGACCTCTGTCTCAATTGGCTTTTAAATGTATATGACAGCCAAAGGACAGGGCAGATGAG
D3F     AGACCTCTGTCTCAATTGGCTTTTAAATGTATATGACAGCCAAAGGACAGGGCAGATGAG
D3R     ATACATCTGTTGCCATTTGCAGTTTGG-------CT------------TGATCCAATGAC
D1R     ------------------------------------------------------------
D2R     ATACATCTGTTGCCATTTGCAGTTTGG-------CA------------AGATCCAATGAC


D1F     ------------------------------------------------------------
D2F     AGTTCTTTCTTTCAAGATCGGTCTCGTCTGCCTTTGCTGCGGTCATTTAGAAGAAAAATA
D3F     AGTTCTTTCTTTCAAGATCGGTCTCGTCTGCCTTTGCTGCGGTCATTTAGAAGAAAAATA
D3R     AACCCCCTTTT-------------------------------------------------
D1R     ------------------------------------------------------------
D2R     AGCTTCTTTTGTAACATTCTAAGTTTTGTAGCTGTTCTGTAG-------GCAGAGAACTT


D1F     ------------------------------------------------------------
D2F     CAGATACATGTTCCGGCTAATTTTT-----------------------------------
D3F     CAGATACATGTTCCGGCTAATTGCCGA-------CCCAAATCGCCTAGTTGATCAAAGA-
D3R     ------------------------------------------------------------
D1R     ------------------------------------------------------------
D2R     GATATGGTTGAATTCATTCATAGAGTCTAGTAAGTCCATCATGATTGGATGATCCCAGTG


D1F     ------------------------------------------------------------
D2F     ------------------------------------------------------------
D3F     ----------AAATTAGGGTTATTGTTA---CACGACTGTGTACAAGTTCCACGCCAGT-
D3R     ------------------------------------------------------------
D1R     --TACTCTCCTGATCATGATTGATGTAATAAGGCACTTTATTACTAGTTGTAGCCCTCTC
D2R     GGTACTCTCCTGATCATGATTGATGTAATAAGGCACTTTATTACTAGTTGTAGCCCTCTC


D1F     ------------------------------------------------------------
D2F     ------------------------------------------------------------
D3F     ------------------------TGGGTG---------------AGGTTGCAGCA
D3R     ------------------------------------------------------------
D1R     CCATGGTGGTGACACTGAAGATGCCAACTGGGCAGGGGTGAGTGGTGTGACCTCTCGAGA
D2R     CCATGGTGGTGACACTGAAGATGCCAACTGGGCAGGGGTGAGTGGTGTGACCTCTCGAGA
```

106

```
D1F     ------------------------------------------------------------
D2F     ------------------------------------------------------------
D3F     TTTGGTGGATCAAACATTGAACCATCAGTTCGTAGCTGTTTCACAAAAGCTGGAAAGGAC
D3R     ------------------------------------------------------------
D1R     TCTGCTGGCCACTTGTCGGCCATCTGTCATTGACTGCTGTTTCTAAATTTTTCCATCTATT
D2R     TCTGCTGGCCACTTGTCGGCCATCTGTCATTGACTGCTGTTTCTAAATTTTTCCATCTATT


D1F     ------------------------------------------------------------
D2F     ------------------------------------------------------------
D3F     AGAGAAACTATTGAGGCTG--TGCATTTCTTGGCATGGGTGCAACAAGAACCCCAGTCAC
D3R     ------------------------------------------------------------
D1R     ATTGAGGTCATGGAGCTTAGCCGAGTTAGCAGGGGTGAGTGGCACATTATGAGCCGAAAA
D2R     ATTGAGGTCATGGAGCTTGGCCGAGTTAGCAGGGGTGAGTGGCACATTATGAGCCGAAAA


D1F     ------------------------------------------------------------
D2F     ------------------------------------------------------------
D3F     TAGTGTGGTTAGCTGTTTTGCATCGTGTTGCAGCCTCAGAAAATATTCAGCATCAGGTCA
D3R     ------------------------------------------------------------
D1R     CCGAGCGGCTTGGTCGTTGACGTCGTCCACTCC---GCGCTGCAATTCCGCCAAACCTTC
D2R     CCGAGCGGCTTGGTCGTTGACGTCGTCCACTCC---GCGCTGCAATTCCGCCAAAAA---


D1F     ------------------------------------------------------------
D2F     ------------------------------------------------------------
D3F     AATGCAATATCTGCAAGGGATATCCTATTTGTAGGCTTACGGTACCGTTGCCTCAAGTGC
D3R     ------------------------------------------------------------
D1R     GCGGAATTTTCTAA----------------------------------------CATTGT
D2R     ------------------------------------------------------------


D1F     ------------------------------------------------------------
D2F     ------------------------------------------------------------
D3F     CTCAGCTTTGATATGTGTCAGCGGTGCTTCTTTGATGGAAGAACCGGCAAAGTCCAAACT
D3R     ------------------------------------------------------------
D1R     GGGCGTGTTGAT------CTGCGAGGTTCGGGTCTTGGGAGATGGGGCCCCACCCATCGG
D2R     ------------------------------------------------------------


D1F     --------------------
D2F     --------------------
D3F     CCCCCTCCCCCCACGAGAAAA
D3R     --------------------
D1R     CAGTCTTTT------------
D2R     --------------------
```

Figure 25: Multiple sequence alignment of PCR products obtained from all 3 primers used to validate the *M. rosenbergii* dystrophin.

```
D.rerio           ------------------------------------------------------------     0
M.rosenbergii     ------------------------------------------------------------     0
D.melanogaster    MDQFLAFLSETETLCENAESDIERNPIMFKDLQSEIETHRVVYDRLDGTGRKLLGSLTSQ    60


D.rerio           ------------------------------------------------------------     0
M.rosenbergii     ------------------------------------------------------------     0
D.melanogaster    EDAVMLQBRLDEMNQRWNDIRSKSIAMRNRIESNSEHWNALLLSLRELTEWVIXKDSELS   120


D.rerio           ------------------------------------------------------------     0
M.rosenbergii     ------------------------------------------------------------     0
D.melanogaster    TLGLGPVLTDAASLQFQLDDHKAFRRQLEDKRPIVESNLTSGRQYIANEAAVSDTSDTEA   180


D.rerio           ------------------------------------------------------------     0
M.rosenbergii     ------------------------------------------------------------     0
D.melanogaster    NHDSDSRYMSAEEQSRELRRSIRREVGKLSEQWNNLLDRSDNWKHRLDEYLTKMRQFQEI   240


D.rerio           ------------------------------------------------------------     0
M.rosenbergii     VEEVSGRLDEAQKTADGWGPISQDPNLADQHAHNVRKFREGLAELQRGVDDVNDQAARFS    60
D.melanogaster    LEDLSSRVALAEQTKTSWLPPSSVGEAI-SRCNSCSRLRDKMTTASALLDDCNEQQSFFT   299


D.rerio           -----------------------DA-----FSTVEREHRGGDSPPKQH--RAIG         24
M.rosenbergii     AHNVPLTPANSAKLHDLNNRWKNLETAVNDRWRQVA----------------SRSREVT   103
D.melanogaster    ANEVLVPTPCLSKLEDLNTRMKLLQXAMDERQKVLRQXEPSRRTRGGDXGRTTSNSGTIG   359
                                                  :                       :


D.rerio           IF-------KKGTLWIRRLLGGEMRENLRNHQTQTTCWDHPKMAELYQSIADLNNVRFSA    77
M.rosenbergii     PLTPAQLASSVSPPWERATTSNK-VPYYINHDQESTHWDHPIMMDLLDSMNEFNHIKFSA   162
D.melanogaster    P--LPNLGQSXKPPWERATTAAN-VPYYIDHERQTTHWDHPEMIEIMKGIADLNEIRFSA   416
                      .   *  *    .  :        :*: ::* **** * :*  .: ::*.::***


D.rerio           YRTAMKLRRMQKALCLDLLSMPAACEAFEQHNLK-QNEQFMDIVQVINCLTSIYDRLEQQ   136
M.rosenbergii     YRTATKLRMLQKKLSLDLAKLQMATDVFDEHGLRGQNDRLIDVGDMVVVLSALYANISAD   222
D.melanogaster    YRTAMKLRSVQKRLALDRISMSTACESFDRTWPACQNDELIDIPDMTTVLHSLYVTID--   474
                  ****.*** .** *.**  .:  *.:*:.     **.::*:*. ::  *.:.* .:.


D.rerio           HSSLVNVPLCVDMCLNWLLNVYDTGRAGKIRTLSFKTGIISLCKAHLEDKYRFLFREVAS   196
M.rosenbergii     HP-EVNTTLAIDLCLNWLLNVYDSQRTGQMRVLSFKIGLVCLCCGHLEEKYRYMFRLIAD   281
D.melanogaster    ---EIDLTLMLDLAXNWILNVYDSQRTGQIRVLSFKVGLVLLCKGHLEEKYRYLFRLVAD   531
                     :: *.:*:. **:*****. *.*::*.**** *:: ** .***:***.:** :*.


D.rerio           ATGFCDQRRLGLLLHDAIQIPRQLGEVASFGGSNIEPSVRSCFQFANNKP---------E   247
M.rosenbergii     PNRLVDQRKLGILLHDCVQVPRQLGEVAAFGGSNIEPSVRSCFTKAGKDRE-------T   333
D.melanogaster    TDRRADQRRLGILLHDCIQVPRQLGEVAAFGGSNIEPSVRSCLEQAGISQEAIDGNQDIS   591
                      ***.*******.:*:.********.*************:.    *  .
```

```
D.rerio         LEASVFLDWMRLEPQSMVWLPVLHRVAAAETAKHQAKCNICKECPIIGFRYRSLKHFNYD    307
M.rosenbergii   IEAVHFLAWVQQEPQSLVWLAVLHRVAASENIQHQVKCNICKAYPIVGLRYRCLKCLSFD    393
D.melanogaster  IELQHFLGWLQHEPQSLVWLPVLHRLAAAEAAKHQAKCNICKEYPIVLFRYRCLKCFNFD    651
                :*    ** *:: **** :*** ****:**:* .:**.****** **.:***.** ::.:*

D.rerio         ICQSCFFSGRVAKGHRMQYRMVEYCTPTTSGEDVRDFAKVLKNKFRTERYFAKHPRMGYL    367
M.rosenbergii   MCQRCFFDGRSGKSHKITHRM---------------------------------------    414
D.melanogaster  MCQRCFFFGRNAKNHKLTHRMHEYCTTTTSTEDVRDFTRALKNKFKSRKYFKKHPRVGYL    711
                :** *** **  .* **:  :**

D.rerio         FVQTILEGDNMETPASSPQLS----HDDTHSRIEHYASRLAEMENRNGSYVNDNVSPNES    423
M.rosenbergii   ------------------------------------------------------------    414
D.melanogaster  FVQSVLEGDALESPAPSPQHTTHQLQNDMHSRLEMYASRLAQVEYGG-----TGSNSTPD    766

D.rerio         MDDEHLLIQHYCQSLNQGSPLSQPQSPAQILISMETEEKGELERVLYDLEQENRKLQAEY    483
M.rosenbergii   ------------------------------------------------------------    414
D.melanogaster  SDDEHQLIAQYCQALPGTSNGSAPKSFVQVMAAMDAEQREELEAIIRDLEEENANLQAEY    826

D.rerio         DRLKKAHDHKG----LSPLPSPPQMLFVSPQSPRDAELIAEAKLLRQHKGRLEARMQILE    539
M.rosenbergii   ------------------------------------------------------------    414
D.melanogaster  QQLCSKEQSGMPEDSNGMQHSSSSMTGLSGQGEQGQDMMAEAKLLRQHKGRLEARMQILE    886

D.rerio         DHNRQLESQLTRLRQLLEQTESKVNGTALS-SPSTAS---PRSDTSLASLRVAASQTTET    595
M.rosenbergii   ------------------------------------------------------------    414
D.melanogaster  DHNRQLEAQLQRLRQLLDEPNGGGSSATSSGLPSAPGSALNSKPNTLQTRSVTALR----    942

D.rerio         MGDDELSSPTQDASTGLEDVIEQLNNSFPHSQGGGRLNP          634
M.rosenbergii   --------------------------------------          414
D.melanogaster  -------NXTR---------------THRQR-----          951
```

Figure 26: Multiple sequence alignment of the amino acid sequence of *M. rosenbergii* with the two other arthropod dystrophin protein sequences *D. melanogaster* and *D.rerio.*. Regions of high similarity between these three sequences are indicated with asterisks.

```
Target    VEEVSGRLDEAQKTADGWGPISQDPNLADQHAHNVRKFREGLAELQRGVDDVNDQAARFSAHNVPLTPANSAKLHDLNNR
1eg3.1.A  -------------------------------------------------------------------------------

Target    WKNLETAVNDRWRQVASRSREVTPLTPAQLASSVSPPWERATTSNKVPYYINHDQESTHWDHPIMMDLLDSMNEFNHIKF
1eg3.1.A  ---------------------PASQHFLSTSVQGPWERAISPNKVPYYINHETQTTCWDHPRMTELYQSIADLNNVRF

Target    SAYRTATKLRMLQKKLSLDLAKLQMATDVFDEHGLRGQNDRLIDVGDMVVVLSALYANISADHPE-VNTTLAIDLCLNWL
1eg3.1.A  SAYRTAMKLRRLQKALCIDLLSLSAACDALDQHNLK-QNDQPMDILQIINCLTTIYDRLEQEHNNLVNVPLCVDMCLNWL

Target    LNVYDSQRTGQMRVLSFKIGLVCLCCGHLEEKYRYMFRLIADPNRLVDQRKLGLLLHDCVQVPRQLGEVAAFGGSNIEPS
1eg3.1.A  LNVYDTGRTGRIRVLSFKTGIISLCKAHLEDKYRYLFKQVASSTGFCDQRRLGLLLHDSIQIPRQLGEVASFGGSNIEPS

Target    VRSCFTKAGKDRETIEAVHFLAWVQQEPQSLVWLAVLHRVAASENIQHQVKCNICKAYPIVGLRYRCLKCLSFDMCQRCF
1eg3.1.A  VRSCFQFANNKPE-IEAALFLDWMRLEPQSMVWLPVLHRVAAAE--------------------------------

Target    FDGRSGKSHKITHRM
1eg3.1.A  ---------------
```

Figure 27: The target-template alignment of the *M. rosenbergii* dystrophin amino acid sequence and the best template identified for its structural protein modeling; the dystrophin template 1eg2.1.A.

Table 5: The cycle threshold values of the serially diluted viral extract obtained in the preparation of standard curve.

Concentration (ng/µl)	Cycle threshold (ct) values			Average ct value
	Prawn 1	Prawn 2	Prawn 3	
0.001	36.21	36.12	36.08	36.14
0.01	35.36	35.08	35.39	35.28
0.1	33.46	34.55	33.52	33.84
1	33.36	33.39	33.29	33.35
10	32.98	32.46	32.75	32.73

Table 6: Table of qPCR results used to construct the standard curve for the quantification of the viral content in tissues samples. The concentrations shown are the concentrations of the serially diluted viral extract of known copy numbers.

Concentration (ng/µl)	In Concentration	Copy number	Average of ct value
0.001	-6.90776	6.62E+06	36.14
0.01	-4.60517	6.62E+07	35.28
0.1	-2.30259	6.62E+08	33.84
1	0	6.62E+09	33.35
10	2.302585	6.62E+10	32.73

Table 7: Cycle threshold values obtained through qPCR of WSSV viral load in experimentally infected *M. rosenbergii.* Its corresponding copy number was obtained using the equation of the standard curve.

Time post infection (hours)	Cycle threshold (ct) values			Average ct value	Corresponding copy number
	Prawn 1	Prawn 2	Prawn 3		
3	37.45	38.2	37.88	37.84	5.47×10^4
6	39.48	39.22	38.16	38.95	2.95×10^3
12	37.4	37.2	36.82	37.14	3.45×10^5
24	36.85	37.4	36.75	37.00	4.99×10^5
36	34.74	35.56	35.35	35.22	5.40×10^7
48	33.94	33.8	33.76	33.83	2.10×10^9

Table 8: The spectrometry readings obtained for each prawn in the quantification of intracellular calcium concentrations.

Sample	Absorbance units (OD)			Average OD	Intracellular calcium concentration (mg/dl)
	Prawn 1	Prawn 2	Prawn 3		
Control	0.401	0.483	0.469	0.451	4.51
3	0.682	0.713	0.681	0.692	6.92
6	0.780	0.799	0.800	0.793	7.93
12	0.732	0.746	0.547	0.675	6.75
24	0.738	0.759	0.753	0.750	7.50
36	0.752	0.742	0.671	0.722	7.22
48	0.718	0.739	0.730	0.729	7.29

Table 9: The Cycle threshold value obtained in the qPCR quantification of the dystrophin gene in WSSV infected *M. rosenbergii.* It relative expression was calculated using the internal control gene EF1-alpha as a reference gene according to the dd(ct) method.

Sample	Ct (Dys)	Mean Ct Calibrator Dys	Ct (Ef1a)	Mean Ct Calibrator (Ef1a)	d(CT) Dys - Ef1a	d Ct (Mean Calibrator)	dd (CT)	2^dd(Ct)	Mean Fold Change in Gene Expression	S.E.M
Control	22.815	22.692	17.32	17.246	5.495	5.446	0.049	0.9666	1.002908983	0.073522
	22.702	22.692	17.402	17.246	5.300	5.446	-0.146	1.1065		
	22.56	22.692	17.018	17.246	5.542	5.446	0.096	0.9356		
T3	28.075	22.692	21.697	17.246	-6.378	5.446	-0.932	0.524	0.331333333	0.080139
	28.147	22.692	20.736	17.246	-7.411	5.446	-1.965	0.256		
	28.234	22.692	21.022	17.246	-7.670	5.446	-2.224	0.214		
T6	25.395	22.692	18.071	17.246	-7.324	5.446	-1.878	0.272	0.363333333	0.113456
	25.203	22.692	18.923	17.246	-6.280	5.446	-0.834	0.561		
	24.599	22.692	17.193	17.246	-7.406	5.446	-1.960	0.257		
T12	28.181	22.692	23.448	17.246	-4.733	5.446	0.713	0.61	0.726666667	0.127628
	28.401	22.692	23.876	17.246	-4.525	5.446	0.921	0.528		
	27.902	22.692	22.515	17.246	-5.387	5.446	0.059	1.042		
T24	27.633	22.692	20.939	17.246	-6.694	5.446	-1.248	0.421	0.706	0.166283
	27.583	22.692	21.542	17.246	-6.041	5.446	-0.595	0.662		
	27.146	22.692	21.75	17.246	-5.396	5.446	0.050	1.035		
T36	24.539	22.692	19.653	17.246	-4.886	5.446	0.560	1.474	1.692333333	0.406192
	24.717	22.692	19.799	17.246	-4.918	5.446	0.528	1.442		
	24.247	22.692	19.91	17.246	-4.334	5.446	1.112	2.161		
T48	26.961	22.692	21.86	17.246	-5.101	5.446	0.345	1.27	1.643333333	0.171966
	26.454	22.692	22.081	17.246	-4.373	5.446	1.073	2.104		
	26.899	22.692	22.091	17.246	-4.808	5.446	0.638	1.556		

112

Table 10: A list of the possible protease cleavage sites in the protein sequence of *M. rosenbergii* dystrophin, shown with the position of each cleavage site and the probability of cleavage occurrence.

Position of cleavage site	Resulting peptide sequence	Peptide length (amino acid)	Peptide mass [Da]	Cleavage probability
414	HPM	3	383.466	-
92	NDRW	4	589.608	96.60%
342	AW	2	275.307	94.60%
163	AY	2	252.27	94.30%
272	EEKY	4	567.596	94.30%
384	Y	1	181.191	94.30%
81	RW	2	360.416	92.30%
216	Y	1	181.191	91.80%
18	ADGW	4	447.448	87.10%
238	NW	2	318.332	87.10%
352	VW	2	303.361	87.10%
130	Y	1	181.191	85.40%
256	SF	2	252.27	85%
243	VY	2	280.324	83.80%
324	F	1	165.192	83.80%
274	Y	1	181.191	82.20%
38	F	1	165.192	82.10%
276	F	1	165.192	82.10%
311	AF	2	236.271	81.70%
169	L	1	131.175	75.90%
195	GL	2	188.227	75.90%
382	PIVGL	5	497.635	75.90%
140	HW	2	341.37	75.60%
190	DVF	3	379.413	74%
392	SF	2	252.27	74%
400	F	1	165.192	74%
59	F	1	165.192	72.50%
160	F	1	165.192	72.50%
155	EF	2	294.307	71.30%
339	F	1	165.192	67.80%
118	W	1	204.228	67.60%
387	RCL	3	390.501	66.90%

234	L	1	131.175	65.20%
263	VCL	3	333.446	65.20%
176	L	1	131.175	62.50%
212	L	1	131.175	62.50%
254	L	1	131.175	62.50%
390	CL	2	234.313	62.50%
399	F	1	165.192	59.40%
260	IGL	3	301.386	55.20%
285	L	1	131.175	55.20%
350	SL	2	218.253	55.20%
27	L	1	131.175	54.50%
42	EGL	3	317.342	54.50%
110	L	1	131.175	54.50%
180	L	1	131.175	54.50%
230	L	1	131.175	54.50%
340	L	1	131.175	54.50%
353	L	1	131.175	54.50%
66	VPL	3	327.424	54%
105	TPL	3	329.396	54%
202	L	1	131.175	51%
278	L	1	131.175	51%
251	M	1	149.208	49.80%
74	L	1	131.175	48.50%
295	L	1	131.175	48.50%
356	L	1	131.175	48.50%
129	VPY	3	377.441	44.30%
291	L	1	131.175	43.40%
305	L	1	131.175	43.40%
45	EL	2	260.29	39.60%
172	L	1	131.175	39.60%
183	L	1	131.175	39.60%
185	M	1	149.208	36.40%
408	SH	2	242.235	34.40%
357	H	1	155.156	34.30%
77	L	1	131.175	33.60%
236	CL	2	234.313	33.60%
240	L	1	131.175	33.60%
148	L	1	131.175	31.60%
239	L	1	131.175	31.60%
293	GL	2	188.227	31.60%
294	L	1	131.175	31.60%
208	IDVGDM	6	648.729	31.10%

157	H	1	155.156	28.70%
84	L	1	131.175	28.40%
268	L	1	131.175	28.40%
8	L	1	131.175	26%
149	L	1	131.175	26%
178	SL	2	218.253	26%
125	N	1	132.119	25.10%
145	DHPIM	5	611.714	24.40%
33	AH	2	226.235	24%
62	AH	2	226.235	24%
12	DEAQ	4	461.429	22.50%
173	Q	1	146.146	22.50%
146	M	1	149.208	21.50%
394	DM	2	264.296	21.10%
31	ADQH	4	469.454	20.90%
79	NN	2	246.223	19.20%
283	PN	2	229.236	19.20%
46	Q	1	146.146	17.50%
246	Q	1	146.146	17.50%
288	DQ	2	261.235	17.50%
396	CQ	2	249.285	17.50%
70	TPAN	4	401.42	17.30%
348	QEPQ	4	500.509	16.70%
367	IQH	3	396.446	15%
171	RM	2	305.395	13.90%
167	T	1	119.12	11.90%
275	M	1	149.208	11.90%
325	T	1	119.12	11.90%
193	DEH	3	399.36	11.80%
55	NDQ	3	375.338	11.60%
227	N	1	132.119	11.60%
34	N	1	132.119	10.90%
63	N	1	132.119	10.90%
241	N	1	132.119	10.90%
120	ER	2	303.318	10.50%
94	Q	1	146.146	9.90%
300	VQ	2	245.279	9.90%
368	Q	1	146.146	9.90%
344	VQ	2	245.279	9.70%
333	T	1	119.12	9.30%
215	AL	2	202.253	9.10%
174	K	1	146.189	8.60%

267	CGH	3	315.347	8.60%
374	C	1	121.154	8.40%
36	R	1	174.203	8.30%
289	R	1	174.203	8.30%
93	R	1	174.203	8%
303	R	1	174.203	8%
218	AN	2	203.198	7.90%
315	N	1	132.119	7.90%
364	SEN	3	348.313	7.90%
372	CN	2	235.258	7.90%
123	T	1	119.12	7.70%
132	IN	2	245.278	7.70%
156	N	1	132.119	7.70%
326	K	1	146.189	7.50%
375	K	1	146.189	7.50%
109	TPAQ	4	415.447	6.90%
158	I	1	131.175	6.90%
304	Q	1	146.146	6.90%
14	T	1	119.12	6.60%
26	PN	2	229.236	6.60%
83	N	1	132.119	6.60%
86	T	1	119.12	6.60%
165	T	1	119.12	6.60%
47	R	1	174.203	6.50%
196	R	1	174.203	6.50%
338	VH	2	254.289	6.30%
98	R	1	174.203	6.20%
321	R	1	174.203	6.20%
403	DGR	3	346.343	6.20%
117	SPP	3	299.327	6.10%
184	Q	1	146.146	5.90%
250	Q	1	146.146	5.90%
198	Q	1	146.146	5.80%
273	R	1	174.203	5.80%
383	R	1	174.203	5.80%
73	K	1	146.189	5.40%
168	K	1	146.189	5.40%
175	K	1	146.189	5.40%
182	K	1	146.189	5.40%
290	K	1	146.189	5.40%
52	DDV	3	347.325	4.90%
88	V	1	117.148	4.90%

226	ADHPEV	6	666.689	4.90%
7	SGR	3	318.333	4.80%
201	NDR	3	403.395	4.80%
248	T	1	119.12	4.80%
277	R	1	174.203	4.80%
284	R	1	174.203	4.80%
322	S	1	105.093	4.60%
97	S	1	105.093	4.50%
99	S	1	105.093	4.50%
138	T	1	119.12	4.40%
411	T	1	119.12	4.40%
151	DS	2	220.182	4.30%
153	MN	2	263.312	4.30%
298	HDC	3	373.384	4.30%
56	A	1	89.094	4.20%
309	GEVA	4	374.394	4.20%
360	VA	2	188.227	4.20%
58	R	1	174.203	4.10%
377	AY	2	252.27	4.10%
21	GPI	3	285.343	3.90%
35	V	1	117.148	3.90%
219	I	1	131.175	3.90%
320	EPSV	4	430.458	3.90%
229	T	1	119.12	3.70%
231	A	1	89.094	3.70%
323	C	1	121.154	3.70%
398	RC	2	277.342	3.70%
24	QD	2	261.235	3.60%
72	A	1	89.094	3.60%
181	A	1	89.094	3.60%
257	K	1	146.189	3.60%
281	IAD	3	317.342	3.60%
409	K	1	146.189	3.60%
60	S	1	105.093	3.30%
71	S	1	105.093	3.30%
122	T	1	119.12	3.30%
161	S	1	105.093	3.30%
213	S	1	105.093	3.30%
220	S	1	105.093	3.30%
228	T	1	119.12	3.30%
39	R	1	174.203	3.20%
100	R	1	174.203	3.20%

126	K	1	146.189	3.20%
331	DR	2	289.291	3.20%
370	K	1	146.189	3.20%
388	K	1	146.189	3.20%
43	A	1	89.094	3%
121	A	1	89.094	3%
164	R	1	174.203	3%
166	A	1	89.094	3%
186	A	1	89.094	3%
247	R	1	174.203	3%
373	I	1	131.175	3%
87	A	1	89.094	2.90%
336	EA	2	218.21	2.90%
354	A	1	89.094	2.90%
4	VEEV	4	474.511	2.80%
114	SV	2	204.226	2.80%
302	VP	2	214.265	2.80%
406	SGK	3	290.319	2.80%
197	G	1	75.067	2.70%
249	G	1	75.067	2.70%
329	GK	2	203.241	2.70%
82	K	1	146.189	2.50%
312	G	1	75.067	2.50%
327	A	1	89.094	2.50%
57	A	1	89.094	2.40%
410	I	1	131.175	2.40%
137	HDQES	5	614.569	2.30%
13	K	1	146.189	2.20%
22	S	1	105.093	2.20%
76	HD	2	270.245	2.20%
147	D	1	133.104	2.20%
179	D	1	133.104	2.20%
233	ID	2	246.263	2.20%
245	DS	2	220.182	2.20%
252	R	1	174.203	2.20%
358	R	1	174.203	2.20%
96	VA	2	188.227	2.10%
102	EV	2	246.263	2.10%
111	A	1	89.094	2.10%
361	A	1	89.094	2.10%
37	K	1	146.189	2%
49	GV	2	174.2	2%

85	E	1	147.131	2%
159	K	1	146.189	2%
264	C	1	121.154	2%
286	V	1	117.148	2%
316	I	1	131.175	2%
332	E	1	147.131	2%
334	I	1	131.175	2%
124	S	1	105.093	1.90%
314	GS	2	162.145	1.90%
187	T	1	119.12	1.80%
369	V	1	117.148	1.70%
211	VVV	3	315.413	1.60%
253	V	1	117.148	1.60%
355	V	1	117.148	1.60%
112	S	1	105.093	1.40%

Table 11: The 11 dystrophin sequences retrieved from the NCBI database. The NCBI accession number for each sequence is shown in parenthesis below each species name.

Species name	Nucleotide sequence	Length (bp)
Asteroidea sp *(X99737.1)*	CACCCTAAGATGACTGAGCTCTTCCAGTCGTTAGGAGAYCTAAACGATGTCAAGTTCTCAGCCTACAGAACTGCCATGAAGCTCAGAAAACTCCAAAAAG CTCTCTGCTTCGACCTCCTGAGCATGAACAGCGCAGACAGGATCTTCCAGCAGCATGAGGTTCTCTCCTCNTGCACAGACCGGACGTTAGACGTCAACGA GATCATCACCTTACTCACCTCCGTCTACGAGAACATCACCTCGGATCATCCCAACTTGGTGGTCGTGCCTCAGTGTGTAGACATGTGCCTCAACTGGCTTC TCAACGTCTACGACACGGTGCGAAGCGGTCGCATGAGGGCGCTCTCCTTCAAACTCGCCATCATCATTCTCAGCAAGGCGCACTTGGAGGACAAATACA GATATATTGTACGATGTGTTGCCAACAACAACGGCTTCATTGATCAGCGAGGTCTCGGCCTCCTGCTTCATGACTGCATTCAGATACCACGGCAACTCGG GGAGGTAGCGTCGTTTGGAGGCAGCAACATTGAGCCTAGTGTCCGTAGCTGCTTCAACATGAGCGGAGGCAAGTCTCACATCGAGCCTCCCCAGTTCTT GGCTTGGATGAAACTTGAGCCTCAATCCATGGTCTGGATGCCCGTCTTACACCGCATGGCGGCTGGGGAGACCGCCAAGCACCAGGCTAAATGCAACGT CTGTAAGGAGTTCCCCATCGTTGGGTTGAGGTATCGCTGCCTCAAGTGCTTCAACTTTTACATGTGCCAGAACTGCTTCTTCTCGGGCCGGGTGTCCAAAA ACCACAAACACTCCCACCCCATGCAGGAGTATTGTACGACGACTACCTCTGGAGAAGATGTGAGGGACTTCGCCAAAGTGGTGCGCAACAAGTTCAAGT CTAAGCGCTACTACCAGAAACATCCGCGGGTGGGCTACCTGCCTGTCCAGTCAGTACTAGAGGGCGATGACCTGGAATCACCTGTGACATCACCACAGC AGATGGCCAATGAGGACATGCATACCAGATTAGAGCTGTATGCCAGCAGGTTAGCCGAGGTTGAGGAGCACGGGTCATTCCTTACNCCTTCCCCCGACT TGGATGATGAGCACCAGTTGATTGCACAGTACTGCCAGAGCTTAGGAGGAGATGTGTCTTCTGTGCCACGAAGCCCTGCACAGATCGTGGTGGCTATCG AGTCTGATCAGAGGGGTGAGCTAGAAAACCAAATTCTCATCTTGGAAGAAGAGAACAGAACTCTCCTTGCCGAGCTGGACCGCCTGCGTGCCCTCCGCT ACGACGACCACACGCGTGCTGAGGTCAGCTCCGAGGAAGACTTCCGCCATCCCCGGCCGACACGGAGCTGGTTGCAGAGGCCAAGCTTCTCAGACAA CATAAGGGTCGTCTGGAGGCA	1400
Branchiostoma *lanceolatum* *(X99736.1)*	GACCCCAAGATGACAGAACTCTTCCAATCATTAGCTGACTTGAATGACGTNAGATTCTCAGCCTACCGCACAGCCATGAAACTCAGGAGGTTACAGAAG GCTCTATGTTTGGACTTGCTAAGCATGAACAACGCCATTGATGCCGAGATCATCAACTGTCTGACCACGATCTACGACAATCTGGAGCAGGACCACGGTA ACCTGGTCAACGTGCCGCTCTGCGTAGACATGTGCCTCAATTGGTTACTCAACGTGTATGATACGGGTAGGAGTGGGAAGATCAGGGTGCTATCGTTCA AGGTGGGGATCATCTCTCTCTGTAGAGCGCATCTCGAGGACAAGTACAGATTCAACTTCAGGTTGATAGCAGAGGCCACAGGATTTGCCGACCAGAGAA AGCTGGGGTTACTGCTGCATGATCTAATACAGGTTCCTCGTCAGCTAGGTGAGATTGCGTCATTCGGCGGGAGCAATATCATCGAGCCCAGCGTGCGCA GCTGCTTCGAGCGGGCAGGCGGGAAGCCTGAGATCGAGGCGGCCCATTTCCTGGACTGGATGAAGCAGGAGCCCCAGTCCATGGTGTGGTTACCTGTC CTGCACAGGCTAGCCGCAGCTGAGACGGCCAAACACCAGGCCAAGTGCAATATCTGTAAAGAATACCCAATTGTAGGATTTAGGTATCGCTGTTTGAGG TGCTTCAACTTTGACATGTGCCAGAGCTGTTTTCTGTCGGGCAGGAAAGCCAAGGGACACAAGCTGTCCCACCCAATGCAGGAGTACTGCACAGCTACG ACCTCAGGGGAGGATGTGTGCGGGACTTCGCCAAGGTCGTCCGCAACAAGTTCAGGTCCAAGAAATCCTTGAAAAAACACCCCCGACTGGGGTACCTTCCG GTGCAGACCGTGTTAGAGGGGAGACAACCTTGAAACGCCGTCGTCCTCGCCGCAGCAGGCCGTTTCTCAGGACATGCATAGCCGGCTGGAGCTGTACGCC AGCAGGTTGGCTGAAGTAGAACAGAGTATGAACAGCATGAGCAGTTCTTCAGACATGGAGGACGAGCACACACTGATCCAGCAGTATTGCCAGAGTCT GGGAGGAGACTCAGGAGTGCCAAAGAGCCCAGCTCAGATCGTGGTGGCCATCGACCAGGAACAGCGAGCAGAGCTGGAAGCAGTGATTAAGGACCTG GAGGAGGAAAAATAGRAATTTACGAATGGAGTATGAGCGGTTGAAAATCCAGAAGCAGCAACAGCCGTCGCCGCACGAGCCGAAGTCGGAGGTGAACC GCGACAACGAGCTGATCGCGGAGGCGAAGCTGCTGCGGCAACACAAGGGCCGCCTA	1338

120

| *Caenorhabditis elegans* (NM_0013063 15.1) | ATGGATGATAATTCTCGAGAAAAGCTCTCAAAATTAGCAAAAGCCAAAGATCAAATCACAGCGAGAGCTAATGAAGCGCTGGCTGCTCTTACACGGACTGTGTCAGAATGTGAAGATTTCGAGAAGCAGATAATGCTCTTCCAGAATTGGAGCGCTCGAATCGGATTTCTATTACAAGCTAGAAAATCTGCAGACATTTCCGCATTTGACATTCCCCATGAGTACCACAAACTCTCCCGCGAGTTTGAGGAATGGACTGTGAAGCTCAATGAGATGAACTCGACGGCCACTGAGAAAGACGACTCGGCGAGAATGCGGGAGCAGCTTAATCATGCTAATGAAACAATGGCAGAGCTCAAGCGGAAATTCAATGAATTTAAGAGGCCTAAAGGTTTTGAGGAGAAGCTTGAAAAAGTTATCACAACCCTGTCAAATGTTGAAATGGGACTAGATGACACCACTGGAATCGATGGATCTGAGTGTGGTGGGGCTCTGATGGAAGTTCGAGCACTTGTTAGAATGCTCGATGGGGCACAGGAGAAATGGAAGGATTTGGCTGAAAACCGAGAGCAACTGGTGAAGGATCGGGTTCTAGATGAGGAGACTTCTAAGGAGACCCTCCAAAAGCTGCAGTATGCAAAAACCAAGTCAAAAGAGCTCTACGAGAGAAGTTCTACGTGCATCGAACGTCTGGAAGACTGTGTTGAAATGTATCAGAGGCTGAAAATGGAATCAGACGAAATTGAGCGATTTTTGGAGGAAATGGAAGGGAAATTGGACCAATACGCGGCAAGTGATCGGCCGGAAGAAGCTGAAATTGTCAATGAGCTCATAAGTGAATGGAATCGGAATGAGGCTGCGATGAAGAATGCCGAGCACTTGCAGAGACAGTTGAACGAACGCGCCATAAAAATCCCTGACGACGTGCTCTCCCTGAAGCGGCTCCGTGCAGATGCTCTCAAAAACCGCCTGAACTCGTGGTGCCGTACCATTCAGGAAATGTCCGAGGACGATGAGTCTGCCCTGTTGGAAATCGACGAGCTCCACCAGAACCTTGAGAAAGAGCTTAAGCTAGTTTCTGATAAGGAACCATCGAAAAATTGCGGAGAAGCTGCGATTCTTGAGAGCCGACAGAGATCGATTGAGTTCTCGAACCAGGAAGCTAGCGGCCAAGAATCCCCGGCTAGCTGCCACGTCATCTGATGTTCTAGCTGGATTGAATCAGAAATGGAAGGAGTTGGAAGTAAAGGCTTCTGCTGAAAAAGCTCCTGCTCCAGAGCTCCGTGACGCTAGATTAAGCTCCCCGTCCGAGCAGCCATTCGACAAGAGAGTTCAAGAACTGTGCGACCTATTTGAAAACTTGGAAGCTCAGCTGGATTTCAATGGATCTCCCGTATCAATGGTGACTGAATATCAGAAGCGAGTTGAGAATTTGGATGAGTATTTGGATGAGTACCGACCAGCTCTTGATGATACTATTGAGGAAGGCCGGAAAATTGCAGAGACTGGAAGACTGGAGCTTCAGCACATTCGGCTATTGAGAAGCTGGATGAGCTGACGAATCGGATTGAGCAGGTGGAGGTGGAGTTGGATAAGCATCGGGATAAGGTGCCGTCGCTTGTGGAGCAGCATGAGCAGCTGAAGAAGGATATTGACTCATTCCTCCTGGTACTCGACGTCTTCACCGACCGTAATCTTGATGACGTGGACATTGCTAAATCGACACGTAAAGAGCTTGCCGAGCGAGATTCCCATATTGTCTCCCTAACTTCCAGAGCAACTGCCATACATTGTGCTCTTCCGGGAAAAAGGCCCCCAACTCCATGACGTCACCCTGGACAAGCTCCGTGACAGGATAGAGAAGCTCGAGGCTCGACTTTCGGCGACTGAGAAGAAGCCAGTGGAGACTGTCAAATCTACAATACCCGATAGACCAGAAGTTCCGGAGGAGCCGGAGAAATCCTCGCCAGATCGCACGAGCCGCTCGAGTTTGCAGCTGGCAATGGAGGCATATGGAACAGCTACAGAGGATGATAGTGTGATCAGTGAGGCGGTTACTGTAGGCCAAAAATCTGTGGATCAGGTAGATCCTGTGGAGCAATTGGAGCCAGTGGAGCCTGTTGAGCCAAAGTTAGAGGGTTAAGCAGCTCAAGGATGAGGCGACCGAGGAGGAGGAGAAACGAACAATTATTCTACCGGATGAGACAGAGAAAGTGATAGAAACTATCCCGGCTGCGAGGCCCTCAGCGGGACCCTCAGAGGGTACTGTAGCAGAAGTATCCACGTCAGAGATCCTTAAAGCCCGACCTGCACAAGAGTCAATTGAACGGACTGTTCGAGAAGTTCCGGTTGATGAATATGAAGAGACTGCGAATATTTCCAGTGGAGATGAGCTTCAGGATCACAAAATTTCCTCGGCGGTGCCGGATTCCGAGTCTGAAATCGCTTCAATGTTCGAAGTGCTCGACTCAATCGAGGACTCGCACACCAATTTCGAAGAGTTTCCGTTTGACTACCTAGATAGTGCTGATGACGATTTGAAGAAAACCCTACTCAAACTGGAGTCCTGTGAGAAGACACTGGCCAAAAATGAAATGACTATTAACATTGCGCAAGCCGAGAATGCGAGGGAGAGGATTACGATGCTGCGTCAAATGGCACTTCAACGGAAGGACAAGCTGCCAAAGTTTAACGAGGAATGGAATGGCAATGCAATGCAAGAGTTAATCCAACTCGCCGACGCCCTTGTCGACGAGGCCGAACGATACGAATCCGACCAGATCCCCCAAATGGATCGGAAATCTGCTCCAAATGTGCTCGGCGAGCTGCGAAAACGTGTCGCCAACGCAGAAGGACCGGTGATAGATCTCGTGAAGAAGCTCTCCCAGTTGGTACCCCGAATGCAAGAAGATTCCCCAAAATCTCAGGACATTCGGCAGAAAGTCTACGGGATTGAGGATAGATTCAGAAGAGTCGGTCAAGCCGAGGGAGCAGCGATCTCTAAGGCTTTATCCAGTGCTCTCACGGAGCCGGAGCTCAAGTTGGAGCTGGATGAAGTTGTGAGATGGTGTGAGATGGCAGAGAAGGAGGCAGCTCAGAATGTGAATTCGCTGGATGGAGATGGATTGGAGAAGCTTGACGGACGGCTTGCACAGTTTACAAAAGAATTGCAGGAGAGAAAAGACGATATGGTGCAACTGGAAATGGCAAAGAACATGATCATCCCCTCCCTAAAAAGGAGACGCTCATCACGATCTCCGCCGTAACTTTTCCGATACCGCGAAACGTGTTGCAATGGTTCGTGATGAGCTCTCAGATGCTCACAAATGGGTTGCCACGTCACGTGACACATGTGACACTTTCTGGGCAGACATTGACTCCCTGGAGCAGTTGGCACGGGATGTGGTGAGAAGAGCCAACGGGATCCGGATGGCTGTGATTTATACTCCGTCGAGGGAGAATGTGGAAGGTGTGCTGCGGGATGTTCAGAGATTGAAGATGTCGATTGGGGATGTGAAGAAGAGGGTTCAGACGGCGAATCTACCGCCGGCGATCAAGTTGGCCGGGAAGAATGCGAAGAGAGTGGTCCAAGTGCTCACCGAGACAGCCACCACGATAGCCGACTGTCATGACATACCTACCTACCTGATCGATGAGATGAATGATTCAGGCGGTGACACTACGGAATCCAGGAGTACTGTTGTCGAGATGACTTCAGTGCACACTAAGCAAAGCTCCTCAAGCTCTTCTAACAAAACCCCATCCGCTGGTGGTGAATCTGACGATGCTCACACCTTAAACGGAGATGATGAGCAATCGGAGGAAGATCAGAAGATCTACTCACGAGAGTCTTCGTCAACATTGCCACGTGGAGTTTCCAGCCTAGGATCAACAGGTTCTAGTGGGGTCCTTGACCCAGTAGCGGTTCAGCTGACACACACAAGGCACTGGCTTCATGATGTAGAACGGGATGCCTCGATTACTGTAGATCTGGCACAATGGCAGCCGGCTCGAGAGCTGTGGCAGAGTATTCAGGGGATTATCGATGAGATTCGGTTGAGAAGTGTTCATGTCACAGGTGCCCATGATGCGAGCCCCAATCGACAAGTTAGGCAGCAAGCCGCGCAGTTGCTGACAGAAATGCGCCGAACCATCGAAAATTGCGAGAAGCGTTGTCTGATCCTCAATCAAATATCAGATATCGCGCGACAAAACGAGGCATCTCGAAACGAAATGGAGCTCTGGCTTAAAAGTGCAAGTGACGTCATCGGAGAGCGAAGAGTTGAAGAGCTCAGCGAAGAGGTTGTTCGGCAGGAGTTGCAGGTTTTGGAACGAGTTGTTGAGCAGTTAACGGAGAGAAAGGACAAAATGGCGGAGATTAATTCCCAGGCGAACAAGATTGTGGACACCTACACAAAGGATGAAGCTCATAATTTGAGTCATTTGCTCAGCAGGCTTAATATGAGCTGGACGAAGTTTAATGATAACATTCGTATCCGCCGAGCCGTACTCGAAGCTTCACTCCGCTCCCGCCGGGACTTCCACTCTGCACTTTCAGAATTTGAAAAGTGGCTTTCACGGCAAGAAGACAATTGCTCAAAACTCTCAGCGGACACATCTAATCATCAAGCTATTAAAGACACT | 7515 |

	TCGAAGCGGAAAAATTGGACGCAAAGCTTTAAAACTTTGAATGCGGAGCTGAATGCTCATGAGGATGTGATGAAAAGTGTCGAGAAAATGGGAAAAAT GTTGGCTGAAAGTTTGGAATCTGGAAATGAAAAAGTTGAGCTTTTGAAAAGAGTTGGTGAGACCACAAGAAGATGGACAGCGCTGAGGAAGACTACTA ATGAGATTGGAGAACGTCTCGAAAAAGCCGAACAAGAATGGGAGAAGCTTTCTGACGGTCTAGCTGACCTCTTATCGTGGGTTGAAGCCAAAAAACAA GCAATAATGGATGAGCAACCGACTGGTGGAAGCCTTTCTGCAGTTATGCAGCAAGCCTCATTTGTGAAAGGACTGCAAAGGGAAATTGAATCAAAAACT GCAAATTATAAGAGTACTGTAGAAGAAGCTCATAGTTTCCTGATGCAACATGACTTGAGACCGAAACTTCATAGCCCACATGTCCTGGATGATGATTATG AGAAGGAGGAGCTCGCCAACTTGGAGCAACGTCGTCGTGGCCTCGAGATCAATGCAAATTGCGAGAGGCTCAAGAAAAACTGGGCAGAGCTCGGCATC GAAGTAGAATCATGGGATAAGCTCGTTCAGCATGCAATGCAGCGGCTCCAGGAGCTCGAGAGAAACTTGGCAGAGTGCCAACTTCATCTGACGTCATCG GAAAATGAGATTGAAACGATGAAGGCTGTCGAGAAGATTCATTTGGAAGATTTGAAAATTGCGCGCGAAGAAACTGATCAAATTTCTAAGAGAATCGAT GAAGTTCGGCTTTTTGTCGATGATGTAAATGATGCGGCGGCAAGGTTGTTGGCTGAGGATTTGAAGCTTGATGAGCACGCAAAGGGACAAATTGAGCA TGTTAATAAGAGATATTCAACCCTAAAACGTGCGATTCGAATCCGGCAGGCGGCGGTCAGAAGACGCCGCATCGGATTTTGGACCAACTTCTGAACATTTC CTAAACCAATCAGTCACCTTGCCATGGCAACGAGCTATCAGTAAATCCAACCTGTTACCCTATTATATTGAACAAACCAGTGAAAAAACACAATGGGAGC ATCCAGTATGGGTGGAAATCGTAAAAGAGCTCTCACAATTCAACCGCGTCAAATTCCTCGCCTACCGTACCGCCATGAAGCTCCGAGCCCTTCAAAAACG GCTCTGTCTCGATCTAGTCGACCTAACACTTTTGGAAAAGGCGTTTGTTCGACTGAAAGGACTTTCAGCTGAAGAATGTCCGGGACTGGAGGGTATGGTT TGTGCTCTGCTGCCGATGTATGAAGCTCTTCATGCAAAATATCCAAATCAAGTGCAAAGTGTCTCGTTGGCTGTCGATATTTGCATTAATTTCTTGTTGAAC CTTTTTGATCAATCCCGCGACGGAATAATGCGTGTTCTCTCCTTTAAAATTGCCATGATCGTATTCTCGAACATTCCATTAGAGGAAAAATATAGATATCTT TTCAAATTAGTTTCCCAGGACGGTCATGCGACCCAAAAACAAATTGCATTGCTTTTATATGATCTTATCCACATTCCTCGTTTGGTCGGTGAATCCGCCGCA TTTGGAGGTACCAATGTGGAGCCATCGGTCCGATCGTGCTTCGAAACTGTACGATTGGCGCCAACAATATCGGAAGGCGCTTTTATCGATTGGGTCAAG AAGGAACCTCAATCAATAGTCTGGCTAGCTGTCATGCATCGATTAGTCATCTCCGAAAGCACAAACATGCGTCCAAATGCAACGTGTGCAAAATGTTCC CAATCATCGGCATCCGTTACCGATGCCTCACCTGCTTCAACTGTGACCTCTGCCAGAATTGTTCTTCAGTCAGAGAACTGCAAAAAGTCATCGAACAAAT CATCCAATGCAGGAGTACTGTGAAAAAAACTACGTCATCCGATGACGCAAGGGATTTCGCGAAAATGATTCGGAATAAGTTCCGGGCGTCGAAACGGCA GAAGGGGTATTTGCCGATTGATGTGGCCGAAGAAGGCATCCCTCTTACATGTCCCCCGGCAAAGGTCACCAACCAGGCCACGGAGCAAATGAACGCAG ACACCTCTCAAATGACTGCACATCTTGCAAAGCTCTCGGCAGAGCACGGTGGAGGCGCCGAGCACATGGAGCCCGTGCAATCACCCCTTCAGATTATTAA TCAGGTAGAACAACTGCAACGGGATGAAATGGACCAGATGCTTCACCGTCTGCAATTCGAGAATAAGCAGCTGAGAAAGGAGCTCGAGTGGAAACGAG GTGCAGCTAGTACTATGGAAATTGATAGAAGCTCAAAACGTCATCAGGAACGACATCAGAGTGAGAGTCGGGGAGGAACGTTACCATTGAGAAATGGT AGAAGTGTAGTGTCATTGAAGAGCACACAGAGCCAGAATGACGTAATGGACGAGGCGAAAGCCCTACGCCTCCACAAGCAACGTTTGGAGCATCGCTC TCGAATTCTAGAGCAGCAAAATGAGCAGCTCGAGATGCAGCTGCAAAGGTTGAAGAAAGTTATAGATGCGCAAAAGCAACAAGCTCCCCTCTCAACCAA CTCGTTGCTCCGCGGAAGTCATCATCAACCATGGAGCCCCGAACGTGCTCGTTCGGGCTCCGCGAGCACACTGGATCGTGGTCTTATTGTGAGTAGTCGT CATCAGGAGCAGGCAGAAGCGGCCGGCGGCGGAGCTGAAGACTCATCGGATGAAGCTGGCGGAGCCGGTGGTGGGCCACGTGGCAGCAGTGTCGGA CAAATGCAAAATCTAATGACAGCTTGTGATGATTTAGGAAAAGCAATGGAATCCCTGGTGGTGTCTGTTGTATATGATTCCGATGATGAGGAGAATGA TTGA	
Canis lupus (NM_0010033 43.1)	GGTCCTGGAATTTGAAATATCCAGGGGCCTCTACAGAATCCTGGCATCAGTTACTATGTTGACTCACTCAGTGTTGGGATCACTCACTTTCCCCTTACAGG ACTCAGATTTTGGAGGCAATTAGCTTCATCAGAGAAAAACGAATAGGAAAAACTGAAGTGTTATTTTTTAAAAGGCTGCTGAAGTTGGTTGATTTCTCAT CATTCCTAAACCTATTGGAGCACTTAAGTTGGGAGAAAATTTACCAAGATTTTTACTGCTGCCTTGATACATACTTTTTTTTTCCCCAAATGCTTTGGTGGGA AGAAGTAGAGGACTGTTATGAAAGAGAAGATGTTCAAAAGAAAACATTCACAAAATGGGTAAATGCACAGTTTTCTAAGTTTGGGAAGCAGCACATAG AGAACCTCTTCAGTGACCTACAGGATGGGAGACGCCTCCTAGACCTTTTGGAAGGCCTGACAGGGCAAAAACTGCCAAAAGAAAAAGGATCCACAAGA GTTCATGCCCTGAACAATGTCAACAAGGCACTGCGCGTCTTGCAGAAAAATAATGTTGATTTAGTGAACATTGGAAGTACTGACATAGTAGATGGAAATC ACAAACTGACTCTTGGTTTGATTTGGAATATAATCCTCCACTGGCAGGTCAAAAATGTAATGAAAAATATCATGGCTGGATTGCAACAAACCAACAGTGA AAAGATTCTCCTGAGCTGGGTCCGACAATCAACTCGTAATTATCCACAGGTTAATGTCATTAACTTCACCACCAGCTGGTCTGATGGCCTGGCTTTGAACG CTCTCATCCACAGTCATAGGCCAGACCTGTTTGATTGGAATAGTGTGGTTTGCCAGCAGTCAGCCACACAACGCCTGGAACATGCATTCAACATTGCCAA ATATCAATTAGGCATAGAGAAACTGCTTGATCCTGAAGATGTTGCCACCACTTATCCAGATAAGAAGTCCATCTTAATGTATATCACATCACTCTTCCAAG TTTTGCCTCAACAAGTGAGCATTGAAGCCATCCAGGAAGTGGAAATGTTGCCAAGGCCATCTCAAGTTACTAGAGAAGAACATTTTCAGATACATCATCA AATGCACTATTCTCAACAGATCACAGTCAGTCTAGCACAGGGATATGAACGAGCCCCTTCCTTTCCTAAGCCTCGGTTCAAGAGCTATGCCTACACACAG GCTGCTTATGTCACCCACTTCTGACCCCACACGGAGCCCACTTCCTTCACAGCATTTGGAAACTCCTGAAGACAAGTCATTTGGCCGGTCATTGACAGAGAC CGAAGCAAACCTGGACAGTTATCAAACAGCTTTGGAAGAAGTACTCTCGTGGCTTCTTCAGCTGAGGATGCACTGCAAGCCCAAGGAGAGATTTCTAAT GATGTCGAAGAAGTGAAAGAACAATTTCATACTCATGAGGGATATATGATGGACTTGACATCCCATCAGGGACGGGTCGGTAATGTTCTCCAACTGGGA AGTCAACTGATTGGAACAGGGAAATTATCAGAAGATGAAGAAACCGAAGTGCAGGAACAAATGAATCTCCTCAATTCAAGATGGGAATGCCTCAGGGT	13887

AGCTAGCATGGAAAAACAAAGCAATTTACATAAAGTTCTAATGGATCTCCAGAATCAGCAACTGAAAGAGTTAAATGACTGGCTAACCAAAACAGAAGA
GAGAACAAGGAAAATGGAGAAGGAGCCCCTTGGACCTGATATTGAAGACCTAAAACGCCAAGTACAACAACATAAGGTGCTTCAAGAAGACTTAGAAC
AGGAACAAGTCAGGGTCAATTCCCTCACTCATATGGTGGTGGTAGTCGATGAATCTAGTGGAGACCATGCAACTGCTGCTTTGGAAGAACAACTTAAGG
TACTGGGAGATCGATGGGCAAACATCTGTAGGTGGACAGAAGATCGCTGGGTTCTTTTACAAGACATCCTCCTAAAATGGCAGCGTTTTACTGAAGAAC
AGTGCCTTTTTAGTGCATGGCTTTCGGAGAAGGAAGATGCAGTGAACAAGATTCACACAACTGGCTTTAAGGATCAAAGTGAAGTGTTATCAAATCTTCA
GAAACTGGCTGTCTTAAAAACAGATCTGGAAAAGAAGAAGCAAACCATGGACAAACTCTGCTCACTCAACCAAGACCTTCTTTCAGCGCTGAAAAACAC
AGTGGTAGCCCACAAGATGGAAGCATGGCTGGACAACTTTGCCCAGCGCTGGGATAATTTAGTCCAGAAACTTGAAAAAAGTTCAGCACAGATTTCACA
GGCTGTCACCACCACTCAGCCATCACTAACACAGACAACTGTAATGGAAACAGTAACTATGGTGACCACGAGGGAACACATCTTGGTAAAGCATGCCCA
AGAGGAACTGCCACCACCACCCCCTCAGAAGAAGAGGCAGATTATCGTGGATTCTGAAATTAGGAAAAGGTTGGATGTCGATATAACTGAACTTCACAG
TTGGATTACTCGTTCAGAAGCTGTGTTGCAGAGTCCTGAATTTGCAATCTATCGGAAGGAAGGCAACTTCTCAGACCTTAAAGAAAAAGTCAATGCCATA
GAGCGAGAAAAAGCCGAGAAGTTCAGAAAACTGCAAGATGCCAGCAGATCAGCTCAGGCCCTGGTGGAACAGATGGTGAATGAGGGTGTTAATGCTG
ACAGCATCAAACAAGCCTCCGAACAACTGAACAGCCGGTGGATAGAGTTCTGCCAATTGCTAAGCGAGAGACTTAACTGGCTGGAGTATCAGAACAACA
TCATCACTTTCTATAATCAGCTACAACAATTGGAGCAGATGACAACTACTGCTGAAAACTGGTTGAAAACCCAGCCTACCACCACATCAGAGCCAACAGC
AATTAAAAGCCAGTTAAAAATTTGTAAGGATGAAATCAACCGACTGTCAGCTCTTCAGCCTCAAATCGAGCGATTAAAAATTCAAAGCATAGCCCTGAAA
GAGAAAGGACAAGGGCCAATGTTCCTGGATGCAGACTTTGTGGCCTTTACAAATCATTTTAACCAAGTCTTTGCTGATGTGCAGGCAAGAGAAAAAGAG
CTACAAACAATTTTTTGACAGTTTGCCACCCATGCGCTATCAGGAGACTATGAGTACCATCCTGACATGGATCCAGCAGTCAGAAACCAAACTCTCTATACC
TCAGGTTACTGTCACTGAATATGACATCATGGAACAGAGACTCGGAGAGCTACAGGCTTTACAAAGCTCTCTGCAAGAGCAACAAAATGGCCTAAACTA
TCTCAGCACCACTGTGAAAGAGATGTCAAAGAAAGCACCACTGTCTGATATTAGTCGGAAATATCAATCAGAATTTGAAGAGATTGAGGGACGTTGGAA
GAAGCTGTCTTCCCAGCTGGTTGAACATTGTCAAAAGTTGGAGGAGCAAATGGCTAAACTTCGAAAAATTCAGAATCACATAAAAACTCTGAAGAAATG
GATCACTGAAGTCGATGTTTTCCTGAAGGAGGAATGGCCTGCCCTTGGGGATTCAGAAATTCTGAAAAGACAGCTGAAACAGTGCAGGCTTTTAGTCAA
TGACATTCAGACCATCCAGCCTAGTCTCAACAGTGTCAATGAAGGGGGCTCAGAAGATGAAGAATGAAGCAGAACCAGAGTTTGCTGGCAGACTTGAGA
CAGAGCTCCGAGAACTTAACACCCAGTGGGATTACATGTGCCGCCAGGTCTATGCCAGGAAGGAAGCCTTAAAAGGAGGTTTGGATAAAACTGTAAGTC
TTCAGAAAGATCTGTCAGAGATGCATGAGTGGATGACACAAGCTGAAGAAGAATACCTAGAGAGAGATTTCGAATACAAGACCCCTGATGAATTACAG
ACAGCAGTTGAAGAGATGAAGAGAGCTAAAGAAGAGGCCCAGCAAAAAGAAGCAAAAGTGAAATCCTAACCGAGTCCGTCAATAGTGTCATAGCTCA
GGCTCCACCTGCAGCACAAGAGGCCTTAAAAAAGGAACTTGACACTCTCACCACCAACTACCAGTGGCTCTGCACCAGGCTCAATGGCAAATGCAAGAC
CTTGGAAGAAGTTTGGGCGTGCTGGCATGAGTTATTGTCCTACTTGGAGAAGGCAAACAAGTGGCTAAGTGAAGTAGAAGTCAAGCTTAAAACCACTG
AAAATATTTCTGGGGGAGCTGAGGAAATCGCCGAGGTGCTTGATTCGCTTGAAAATTTGATGCAACATTCAGAGGATAACCCGAATCAGATTCGCATAT
TGGCACAGACCTTGACAGATGGTGGAGTCATGGATGAACTGATCAATGAGGAGCTTGAGACATTTAATTCTCGTTGGAGAGAACTCCATGAAGAGGCT
GTGAGGAGGCAAAAGTTGCTTGAGCAGAGTATCCAGTCGGCCCAGGAGATAGAAAAATCCTTGCACTTAATTCAGGAGTCCCTCTCTTCCATTGACAAG
CAGTTGGCAGCTTATATTGCTGACAAAGTGGATGCAGCTCAGATGCCTCAGGAAGCCCAGAAAATCCAATCAGATTTGACAAGTCATGAGATCAGTTTA
GAAGAAATGAAGAAACATAACCAGGGAAAGGAGACTGCCCAAAGGGTACTATCCCAAATTGATGTGGCACAGAAAAAATTGCAGGATGTTTCCATGAA
GTTTCGATTATTCCAGAAGCCAGCCAATTTTGAGCAGCGCCTACAAGAAAGTAAAATGATTTTAGATGAAGTGAAGATGCATTTACCTGCGTTGGAAACA
AAGAGTGTGGAACAGGAAGTAGTACAGTCACAGTTAAATCATTGTGTGAACTTGTATAAAAGTCTGAGTGAAGTGAAGTCTGAAGTGGAAATGGTAAT
AAAAACTGGACGTCAGATTGTACAGAAGAAGCAGACGGAAAACCCGAAAGAGCTTGATGAAAGAGTTACAGCTTTGAAATTGCATTATAATGAGCTGG
GAGCAAAGGTGACAGAAAGAAAGCAACAGTTGGAAAAATGCTTGAAATTGTCCCGTAAGATGCGAAAGGAAATGAATGCCCTGACAGAATGGCTGGC
AGCTACAGATATGGAACTGACAAAGAGATCGGCAGTTGAAGGAATGCCTAGTAATTTGGATTCTGAAGTTGCCTGGGGAAAGGCTACTCAGAAAGAGA
TTGAGAAACAGAAGGTTCACCTAAAGAGTGTCACAGAGGTAGGAGAGGCCTTGAAAACGGTTTTGGGCAAGAAGGAAATGTTGGTGGAAGATAAACT
GAGTCTTCTGAATAGTAACTGGATAGCCGTCACTTCCCGAGCAGAAGAGTGGTTAAACCTTTTATTGGAATACCAGAAACACATGGAAACTTTTGACCAG
AATGTGGATTACATCACAAACTGGATCATTCAGGCTGATGCACTTTTGGATGAATCTGAGAAAAAGAAACCTCAGCAAAAAGAAGACATACTTAAGCGT
TTAAAGGCTGAAATGAATGACATACGTCCAAAGGTGGATTCTACACGTGACCAAGCAGCAAACCTGATGGCAAACCGCGGCGACCACTGCAGGAAAGT
AGTAGAGCCCAAAATCTCAGAGCTCAACCATCGATTTGCAGCCATTTCTCACAGAATTAAGACTGGAAAGGCCTCCATTCCTTTGAAGGAATTGGAGCAG
TTTAACTCAGATATACAAAAATTGCTTGAACCACTGGAGGCTGAAATTCAGCAGGGGGTGAATCTGAAAGAGGAAGACTTCAATAAAGATATGAGTGAA
GACAATGAGGGTACTGTAAAAGAATTGTTGCAAAGAGGAGACAACTTACAACAAAGAATCACAGATGAGAGAAAGCGAGAGGAAATAAAGATAAAAC
AACAGCTGTTACAGACAAAACATAATGCTCTCAAGGATTTGAGGTCTCAAAGAAGAAAAAAGGCTCTAGAAATTTCTCACCAGTGGTATCAGTACAAGA
GGCAGGCTGATGATCTCCTGAAATGCTTGGATGACATTGAAAAAAAATTAGCCAGCCTACCTGAACCCAGAGATGAAAGGAAAATAAAGGAAATTGATC
GTGAATTGCAGAAGAAGAAAGAGGAGCTGAATGCAGTGCGTAGGCAAGCTGAGGGCTTGTCTGAGGATGGGGCCGCAATGGCAGTGGAGCCAACTCA
GATCCAGCTCAGCAAGCGCTGGCGGGAAATTGAGAGCAAATTTGCTCAGTTTCGAAGACTCAACTTTGCACAAATTCACACTGTCCATGAAGAGTCAGT
GGTGGCGATGACTGAAGACATGCCTTTGGAAATTTCTTATGTGCCTTCTACTTACCTGACTGAGATCACTCATGTCTCACAAGCCCTATCAGAAGTGGAA
GAGCTTCTTAATGCTCCCGACCTTTGTGCTCAAGATTTTGAAGATCTCTTTAAACAAGAGGAATCCTTGAAGAACATAAAAGACAGCCTGCAACAAATCTC

123

AGGTCGGATTGACATCATTCACAATAAAAAGACAGCAGCATTGCACAGTGCCACTCCTGCAGAAAGGGCAAAGCTCCAGGAAGCTCTCTCACGGCTTGA
TTTCCAATGGGAAAGAGTTAACAATATGTACAAGGACCGACAAGGGAGATTTGACAGATCTGTGGAAAAATGGCGGCGGTTTCATTATGATATGAAGAT
ACTTAATCAATGGCTAACAGAAGCTGAACAGTTTCTCAAAAAGACACAAATTCCTGAGAATTGGGAACATGCCAAATACAAATGGTATCTTAAGGAACTC
CAGGATGGCATTGGACAGCGGCAAAGTGTTGTCAGGGTATTGAATGCAACTGGGGAAGAAATAATTCAACAGTCCTCAAAAACAGATGCCAGTATTCTC
CAAGAAAAACTGGGAAGCCTGAATCTGCGGTGGCAGGAGGTCTGCAAACAGCTGGCAGAAAGAAAAAAGAGGCTAGAGGAACAGAAGAATATCTTGT
CAGAATTTCAAAGAGATGTAAATGAATTTGTTTTATGGTTGGAAGAAGCGGATAACGTTGCTAATATTCCACTTGAACCTGGAAATGAGCAGCAGCTAAA
AGAAAAACTTGAACAAGTCAAGTTACTGGCAGAAGAGTTGCCCCTGCGCCAGGGAATTCTAAAACAATTAAATGAAACTGGAGGAACAGTGCTTGTAA
GTGCTCCCCTAAGCCCAGAAGAGCAAGATAAACTTGAAAATAAGCTCAAGCAGACAAATCTTCAGTGGATAAAGGTTTCTAGAAATCTGCCTGAGAAGC
AAGAAGAAATTGAGGCACACGTAAAAGACCTTGGACAGCTGGAAGAGCAGTTAAATCATCTGCTTCTATGGCTGTGTCTCCTATTAGGAATCAGTTGGAAA
TTTACAATCAGCCAAATCAAACAGGACCATTTGACATCAAGGAAATTGAAGTAGCAGTTCAAGCTAAACAGCCGGATGTGGAAGGGATTTTGTCTAAAG
GGCAGCATTTGTACAAGGAAAAACCAGCCACTCAGCCAGCGAAGAGAAAAGCTGGAAGATCTCAGCTCTGATTGGAAGGTGGTAACTCAGTTGCTTCAA
GAGCTGCGGGCAAAGCAACCTGGCCCAGCTCCTGGACTGACCACTGTCAGAGCCCCTCCCAGTCAGACTGTTACTCTGGTGACACAACCCGCGGTTACC
AAGGAAACTGCCATCTCCAAACTAGAAATGCCATCTTCATTGCTGTTGGAGGTACCTGCACTGGCAGATTTCAACCGAGCTTGGACAGAACTTACCGACT
GGCTGTCTCTGCTTGATCGAGTTATAAAATCACAGAGGGTGATGGTGGGTGATCTTGAAGACATTAACGAGATGATCATCAAGCAGAAGGCAACGCTGC
AGGATTTGGAACAGAGGCGCCCCCAGTTGGAAGAACTCATTACCGCTGCCCAGAATTTGAAAAACAAGACCAGCAATCAAGAGGCTAGAACAATCATTA
CTGATCGAATTGAAAGAATTCAGAGTCAGTGGGATGAAGTACAGGAACATCTTCAGAACCGGAGGCTACAGTTGACTGAAATGTTAAAGGATTCCACAC
AATGGCTGGAAGCTAAAGAGGAGGCTGAGCAGGTGTTGGGGCAGGCCAGAGCCAAGCTTGAGTCATGGAAGGAGGCTCCCTACACAGTAGATGCAAT
CCAAAAGAAAATCACAGAAACCAAGCAGTTGGCCAAAGACCTCCGCCAGTGGCAGATAAATGTAGATGTTGCAAATGATTTGGCACTGAAACTTCTCCG
AGATTATTCTGCAGATGATACCAGAAAAGTACACATGATAACAGAGAACATCAATGCCTCTTGGGCAAGCATCCATAAAAGATTGAGTGAGCGAGAGGC
TGCTCTGGAAGAAACCCACAGATTACTGCAACAGTTCCCCTTGGACCTGGAGAAGTTCCTTGCCTGGCTTACAGAAGCCGAAACAACTGCCAACGTCCTG
CAGGATGCCACCCATAAGGAAAGGCTTCTAGAAGATTCCAAGGGAGTAAGAGAGCTGATGAAACAATGGCAAGACCTCCAAGGAGAAATCGAAGCTCA
CACAGATATCTATCACAACCTGGACGAAAATGGCCAAAAAGTCCTGAGATCCCTGGAAGGTTCTGACGATGCAGCCTTGTTGCAAAGACGTTTGGATAA
CATGAACTTCAAGTGGAGCGAACTTCGGAAAAAGTCTCTCAACATTAGGTCTCACTTGGAAGCCAGTTCTGACCAGTGGAAGCGTCTGCACCTTTCTCTT
CAGGAACTTCTGGTATGGCTCCAGCTGAAAGATGATGAGTTAAGCCGGCAGGCACCCATTGGAGGGAGACTTTCCAGCGGTGCAGAAGCAGAATGATGT
ACACAGGGCCTTCAAGAGGGAATTGAAAACGAAAGAACCTGTAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCAGCCTTTAGAAGG
ACTAGAGAAACTCTACCAGGAGCCCAGAGAGCTGCCTCCTGAAGAGAGAGCCCAGAATGTCACACAGCTCCTACGAAAGCAAGCTGAGGAGGTCAACA
CTCAGTGGGAAAAACTGAACGTGCACTCTGCAGACTGGCAGAGAAAAATAGACGAGGCCCTCGAAAGACTCCAGGAGCTTCAGGAAGCAACAGATGA
GCTGGATCTCAAACTACGTCAGGCAGAGGTGATCAAAGGATCCTGGCAGCCTGTGGGTGACCTCCTCATTGACTCTCTCCAAGATCACCTCGAAAAAGTC
AAGGCGCTTCGAGGAGAAATTACACCTCTGAAAGAGAATGTCAGCTACGTCAATGACCTTGCTCGCCAACTCACTACGTTGGGCATTCAGCTGTCACCAT
ATAACCTCAACACTCTGGAAGACCTGAACACCAGATGGAAGCTTCTGCAGGTGGCCATTGAGGACCGCATCAGGCAGCTGCATGAAGCGCACAGGGAC
TTTGGACCAGCCTCCCAGCACTTCCTTTCCACTTCTGTCCAGGGTCCCTGGGAGAGAGCCATCTCACCAAACAAAGTGCCCTACTATATCAACCACGAGAC
CCAAACAACTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGCTGACCTGAATAATGTCAGGTTCTCAGCTTACAGGACTGCCATGAAA
CTCCGAAGACTGCAGAAGGCCCTTTGCTTGGATCTCTTGAGCCTATCGGCTGCATGCGATGCCTTGGACCAGCACAACCTCAAGCAAAATGACCAGCCCA
TGGATATCCTGCAGGTCATTAACTGTCTGACCACTATTTATGATCGCCTAGAGCAAGAGCACAACAATCTGGTCAACGTCCCTCTCTGCGTGGATATGTGT
CTCAATTGGCTGCTGAATGTTTATGACACGGGACGAACGGGGAGGATCCGGGTCCTGTCTTTTAAAACTGGCATCATTTCTCTGTGTAAAGCCCATTTGG
AAGACAAGTACAGATACCTCTTCAAGCAAGTGGCAAGTTCGACAGGATTTTGTGACCAGCGCAGGCTGGGCCTCCTCCTGCATGACTCTATCCAGATCCC
AAGACAGTTGGGTGAAGTCGCATCGTTCGGGGGCAGTAACATTGAGCCGAGTGTCAGGAGCTGCTTCCAGTTTGCTAATAATAAGCCTGAGATCGAAG
CGGCCCTCTTCCTAGACTGGATGCGCCTGGAGCCCCAGTCCATGGTGTGGCTGCCTGTCCTGCACCGAGTGGCTGCCGCGGAAACTGCCAAGCACCAGG
CCAAGTGCAACATCTGCAAGGAGTGTCCCATCATCGGATTCAGGTACAGGAGTCTAAAGCACTTTAATTATGACATCTGCCAAAGTTGCTTTTTTTCTGGT
CGAGTTGCAAAAGGCCATAAAATGCACTATCCCATGGTGGAATACTGCACTCCGACTACATCGGGAGAAGATGTCCGTGACTTTGCCAAGGTACTAAAA
AACAAATTTCGAACCAAAAGGTATTTTGCGAAGCATCCCCGAATGGGCTACCTGCCAGTGCAGACTGTCTTAGAGGGGGGACAACATGGAAACTCCTGTC
ACTCTGATCAACTTCTGGCCGGTAGATTCTGCGCCTGCCTCGTCCCCTCAGCTTTCACACGATGATACTCATTCACGCATTGAGCATTATGCTAGCAGGCT
AAAAAAAAATGGAAAACAGCAATGGATCTTATCTAAATGATAGCATCTCTCCTAATGAGAGCATAGATGATGAACATTTGTTAATCCAGCATTACTGGCGA
AGTTTGAACCAGGAATCCCCCCTGAGCCAGCCTCGTAGTCCTGCCCAGATCTTGATTTCCTTAGAGAGTGAGGAAAGAGGGGAGCTAGAGAGAATCCTA
GCAGATCTTGAGGGGAGAAACAGAAATCTGCAAGCAGAGTATGATCGTCTAAAGCAGCAGCATGAACACAAAGGCCTGTCCCCACTGCCATCCCCTCCT
GAAATGATGCCTACTTCTCCCCAAAGTCCCCGGGATGCTGAGCTCATCGCTGAGGCCAAGCTGCTGCGTCAACACAAAGGCCGCCTGGAAGCCAGGATG
CAAATCCTTGAAGACCATAACAAACAACTGGAATCCCAGTTACACAGGCTCAGGCAGCTGCTGGAGCAACCCCAGGCAGAGGCCAAGGTGAATGGTAC
AACGGTGTCTTCTCCTTCTACCTCTCTTCAGAGGTCAGATAGCAGTCAGCCTATGCTGCTCCGGGTGGTCGGCAGTCAGACTTCAGAATCCATGGGCGAG
GAAGACCTGCTCAGCCCTCCCCAGGACACAAGCACAGGGTTAGAGGAAGTGATGGAGCAGCTCAACCACTCCTTCCCTAGTTCCAGAGGAAGAAATACC

	CCTGGGAAGCCAATGAGAGAGGACACAATGTAGGAAGCCTTTTCCACTTGGCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTTTCAGTCATGACAG ATGAAGAAGGGGCAGAATAAATGTTTTACAACTCCTGATTCCCGCATGGTTTTTATAATATTCATACAACAAAGAGGATTAGACAGTAAGAGTTTACAAG AAAAAAAAAAACAAATCTATATTTTTGTGAAGCGTAGTGGTACTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAGGT TTCACACGTCTATGCAATTGTACAAAAAAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCTTTTGG GTTGTTTAAAAATTTATAACAGTTATAAAGAAAGATTGTAAACTAAAGTGTGCTTTATAAAAAAAAAAGTTGTTTATAAAAACCCCTAAACACACACACACA CACACACACACACACACACACACACTTTGGGGCAGCGCATTGTTTTGCATCGTTTTAGTGTGATATCCACATGAAATTCATGGTTTTTTTTTTTCTTTTCT GGCATATTAAAGATAGGACTTCCTCTAACACCACCAAATGACTACTTCATATGGCTCTTTTAGGAATTGGTGACTGATTGGAACTGGCTGTGGGTTCCGC TTTATGCATCTATATGTCTAAAATTATGTAAATACTGTAGGTATATAGAGAAATAGATCTGAATTTCTAGAGACCTTTGAAATTTCTTGGTCAAATGATCTG GCCACTTGCTAATGGCAAGAGATTGCTTCTGGTCGTCCAGGCTTACCTGCTTGGTCTAGAATGAATTTTCCCTGGAGCCTGAAACCAGGAGGCAACTACC ACACTAAAACATTGTCTGGGGCTCCAGATGTTTCTCATTTGAAACTTTCCACTGACAGCTACGCTATTGAATCTTTTAAAAGGGACATACGAATGAATACA CAGGACTTACTATGTCAGAGTAATCCATGGGTTGATACATTCAGCTTCCCGCTGCTGGCAATGCCACGATCTACATTTAATGATGCTTCAGTGGAAGTCAC GACATCAGAAGACATTTTTAACTCCCAGACAGTGAGGAGGACTTTCTAGTAGGGCTGGAGGGCTGTGGACTCCCAGCCAATCCCTGGGAGGGAGGGGG GTCACTCTTCAAGTTAAAGGATTGGATGAATGTTCATAACCATACATGGAGATCTTTGTAATTACAACAAAATCACTTCCCTCTCCTCACAGTGAAATAGC AGTGGGCATTTGAGGGGGTTTTGCTTTCTTGCTTTATTGCTGCTTTTTTTTTTTTTTTTAGAATTTTTCCCCTATTGCAAGGAGTTTTTAAATGCCACAAGACT TCATTTAAAATAACCCTTGGGAAGGTGTGTAAGACAGCCTCATCATGGTGTGCTCCTGGTAAATATCTACCCAGAAGCCCATGAACTGTGTTTCCCTACTT CACATTTCTCTACAACTAATTCCACACAGGTTTGTAAGGGGGGAAGGTAGATGGATTCACATGTTCCAAAGGGTACTCCAAACCCTTGCTGGTGGGTGAG GAAGGTTAGAGATACAAACATTGTTCAAAAAGAGAATAACCCAAGAGGACTAACTGGTAGGAACAAGCTTTACTCTATCATCTTCTTGGCCTTTCATTTG TTTTCATTCATTCATTCACTAGAAATCACCATGTGACCTATTATTTTGCAAATCTGTTACCTCTTAGATCAAGTATCATTAACTTTTGCAGAGTGGGTTTTGT CTATTATTTTATTAATGACAATTAACATCCAACATGGCTTCTCATGCTATTTCTACCTCACTTTGGTTTTGGGGTGTTCCTAATAATCGTGCACACCTGATTT CACAGCTTCACTACTTGTCCATCATGTGAACATTTTTCTCCATTTTTTCCCCCACTGTCATTTCCAAAGGAAAACCCAAAGCTCTAAGGTAACAGAATGACA TGAAGATTTGGTTTTGGTCTTGCATTTTTTCTCCTTTATGTGACACTGGACTTTTTCTTTCCCCGCAGATTTTTTTTTTAACCATGATTTAAAACAAGGGGTT GTGTTACTTCTACTAAGCAGTTTTTAAGTTTCATTCTAAAAGCAGAGGTAAATAGCGTGCATAATTTTATTTTAATCTTTTCATTTTATCTTTTAGACATTAGT TCTGGAGTGAGTCTGTCAAAATATTTGAATAAAAATTGAGAGCTTTATTGCTACGTTTTAAGCATAATTTGGACATTATTCCGTGTTCTTTATAACCACCAA GTATTAAACTGTAAATCATAATGTAACTGAAGCATAAACATCACGTGGCATGTTCTGTCATTGTTTTCAGGTACTGGGTTCTTGAGTATCATAATATATTGT GTTTTAACACCAACACTGTAACATTTACTAATTATTTTTTTAAACTTCAGTTTTACTGCATTTTCACAACATATCAGACTTCACCAAATATATGCCTTACTATT GTATTATATTACTGCTTTACTGTGTATCTCAATAAAGCACGCAGTTATGTTAC	
Danio Rerio (NM_131785.1)	GGATGCTTTTTCCACTGTAGAGAGGGAACACCGCGGAGGTGATTCGCCCCCAAAACAACACCGAGCCATCGGGATTTTTAAAAAAGGAACACTTTGGAT TCGGCGTTTGCTGGGAGGCGAGATGAGGGAGACTTGAGAAATCACCAGACCCAAACAACATGTTGGGACCCACCCAAAGATGGCAGAACTCTACCAGT CATTAGCGGATCTCAACAACGTGCGGTTCTGGCATACAGGACGGCAATGAAGCTCAGACGAATGCAGAAAGCCCTCTGTTTGGATCTTCTGAGCATGC CTGCAGCCTGTGAAGCCTTTGAGCAGCACAATCTCAAACAGAACGAGCAGTTCATGGACATCGTGCAGGTGATCAACTGTCTGACCAGCATCTACGACC GTCTGGAGCAGCAGCACAGCAGCCTGGTCAACGTGCCTCTCTGTGTGGACATGTGTCTCAACTGGCTGCTCAACGTTTACGATACAGGACGAGCTGGGA AGATTCGTACCCTATCCTTCAAAACAGGAATAATCTCTTTGTGCAAAGCTCACCTTGAAGATAAGTACAGATTTTTATTTCGAGAGGTGGCCAGTGCCACA GGCTTCTGTGACCAGCGGCGCCTCGGCCTCCTCCTGCATGATGCCATTCAGATCCCCAGGCAGCTGGGTGAAGTGGCGTCTTTCGGAGGGAGCAATATT GAGCCCAGTGTGCGCAGCTGCTTTCAGTTCGCCAATAACAAACCGGAGTTAGAGGCTTCAGTCTTCCTGGACTGGATGCGTTTAGAACCTCAGTCGATGG TTTGGCTTCCTGTCCTTCACCGTGTAGCGGCCGCTGAGACAGCAAAGCACCAGGCTAAATGCAACATTTGTAAGGAATGTCCTATTATTGGCTTCAGGTA CCGAAGTTTAAAGCACTTTAACTATGATATCTGCCAAAGCTGCTTCTTTTCTGGCAGAGTGGCCAAAGGTCACAAGATGCAGTACCCTATGGTTGAATATT GCACACCGACGACGTCAGGAGAGGATGTGAGAGACTTTGCCAAGGTGTTAAAGAACAAGTTCAGGACAGAGCGCTATTTTGCCAAGCACCCTCGCATG GGTTACCTTCCCGTCCAGACCATCCTTGAAGGAGACAACATGGAGACGCCTGCATCTTCCCCTCAACTCTCCCATGACGACACCCACTCTCGCATCGAGCA TTACGCTAGCAGGCTAGCTGAAATGGAGAACCGAAATGGCTCGTATGTGAATGACAATGTATCGCCCAATGAAAGCATGGATGATGAGCATCTGCTGAT CCAGCTACTGCCAGAGTCTGAATCAAGGCTCTCCTCTCAGCCAGCCGCAGAGCCCCGCTCAGATCCTCATCTCAATGGAGACTGAAGAGAAGGGAGA GCTGGAGAGAGTGCTCTATGATCTGGAGCAGGAAAACAGGAAGCTGCAAGCGGAGTATGATCGTCTGAAAAAGGCACACGATCACAAGGGTCTGTCAC CGTTGCCTTCACCTCCACAGATGCTTCCGGTGTCGCCTCAAAGTCCACGCGACGCAGAACTGATCGCAGAAGCCAAACTACTGCGGCAACACAAGGGAC GATTGGAGGCAAGAATGCAAATCCTGGAGGATCACAACAAGCAGCTGGAATCGCAGCTTACACGCCTAAGACAGTTACTCGAGCAGACTGAGTCCAAG GTGAATGGCACTGCCCTGTCCTCACCCTCCACTGCCTCTCCGAGATCTGACACCAGCCTGGCCTCACTGCGTGTGGCCGCAAGCCAAACCACAGAGACCA TGGGTGATGATGAGCTGTCCAGTCCCACCCAGGATGCAAGCACTGGATTAGAAGACGTCATCGAGCAGCTCAACAACTCCTTCCCTCACAGCCAAGGAG GCGGACGGCTAAATCCATGAGAGGTCCCAATGTGGGAAGTCTTTTCCACATGGTTGATAACATTGGACATGCCATGGAGTCGCTGGTTCATGTAATAAC	2173

	AGAGGAGCAGGATTTAGACTGAAGACGTCTTCTCCTCGTTGCATGCTTTTGTAGTGTCAACAACTGGACCGGATATGTTTACAATGGGGAATATCAATAA AAATCTATTTTTCTGAAGGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	
Drosophila melanogaster (X99757.1)	CCCCGAGCGGTGCACAGCCTTCAATCTTTTGATCGAGCCATGGATCAGTTCCTTGCTTTTCTTTCGGAAACGGAAACACTTTGTGAAAACGCCGAGTCTGA TATTGAACGCAATCCATTAATGTTCAAGGACCTACAATCGGAAATTGAAACCCATCGCGTGGTTTACGATCGTCTCGATGGAACAGGCCGCAAACTTTTG GGTTCGCTCACCTCACAAGAGGATGCGGTTATGCTGCAGCGTCGCCTCGATGAAATGAATCAACGTTGGAATGATTTAAGATCGAAGAGTATTGCAATG AGGAATCGCCTGGAGTCCAATTCGGAGCACTGGAACGCCCTGCTCTTGTCGCTCCGCGAACTGACCGAGTGGGTTATCCGYAAGGACTCCGAACTGAGC ACTTTGGGCTTGGGTCCCGTTCTGACGGATGCCGCCAGTCTACAGAAACAGTTGGACGATCATAAAGCCTTTCGCCGCCAATTGGAGGATAAGCGGCCA ATTGTCGAGAGCAATTTAACGAGCGGTCGTCAGTACATCGCCAACGAAGCAGCCGTCTCGGATACCAGCGATACGGAGGCCAACCATGATTCCGACTCG CGTTACATGTCCGCCGAGGAGCAGAGCCGTGAGCTAAGACGCAGCATCCGGCGGGAGGTTGGAAAACTTTCCGAGCAGTGGAACAATCTCCTCGATAG ATCTGACAATTGGAAGCATCGCCTGGATGAGTACTTGACGAAGATGCGTCAGTTCCAGGAGATCCTGGAGGATCTTAGCTCCCGCGTTGCCCTTGCCGA GCAGACAAAGACCTCCTGGCTGCCACCCTCATCCGTGGGCGAGGCAATTAGCAGATGCAACAGTTGCAGCCGCCTGAGGGACAAGATGACGACGGCCA GTGCATTGCTGGATGATTGCAACGAGCAGCAGTCCTTCTTCACCGCCAATGAGGTCCTGGTGCCCACACCCTGCCTCTCCAAGCTGGAGGACTTAAACAC GCGCATGAAACTCCTGCAAATHGCGATGGACGAGCGCCAGAAGGTCCTTCGTCAGGTNGAGCCCAGCAGACGCACGAGAGGCGGCGACGAYGGACGC ACCACATCGAACAGCGGCACCATTGGACCACTCCCCAATCTCGGCCAGAGCGTNAAGCCGCCATGGGAGCGGGCTACCACAGCCGCCAATGTGCCTTAC TATATCGACCACGAACGCCAGACCACGCACTGGGACCATCCGGAGATGATTGAGCTGATGAAGGGCCTGGCAGATCTCAACGAGATACGCTTTTCGGCC TACAGAACGGCCATGAAACTGCGATCAGTTCAGAAGCGATTGGCTCTCGACAGGATTAGCATGTCCACCGCCTGCGAGTCCTTCGATCGAACATGGCCT GCGTGCCAAAACGACGAGCTGATCGACATACCCGACATGACCACGGTGCTGCATTCGCTCTACGTGACAATCGATGAAATTGATTTGACGCTAATGCTG GACCTGGCCATYAACTGGATCCTGAATGTCTACGACTCACAGAGGACCGGCCAAATCCGCGTCCTCAGCTTCAAGGTCGGACTGGTCCTCCTCTGCAAGG GTCACCTGGAGGAGAAGTATCGCTATCTGTTCCGCCTGGTTGCAGACACTGACCGACGAGCGGATCAGAGGCGACTGGGACTCCTGCTGCACGATTGCA TTCAGGTACCCCGCCAACTAGGTGAAGTAGCCGCTTTCGGTGGCTCCAACATAGAGCCCTCAGTGAGATCCTGCCTGGAGCAGGCTGGCATCTCGCAGG AGGCCATCGATGGCAACCAGGACATCTCCATCGAGCTGCAGCACTTCCTCGGCTGGCTGCAGCATGAACCGCAGAGTCTCGTGTGGCTTCCGGTGCTCC ATCGCCTGGCAGCGGCGGAGGCGGCCAAGCACCAGGCCAAATGCAACATCTGCAAGGAGTACCCGATCGTTCTCTTCCGCTACAGATGCCTCAAGTGCT TCAACTTCGACATGTGCCAAAAGTGTTTCTTCTTCGGCCGCAACGCCAAGAACCACAAGCTAACTCATCCCATGCACGAGTACTGCACCACTACCACATCC ACCGAGGACGTTCGTGACTTTACACGCGCCCTGAAGAACAAATTCAAGAGTCGCAAATACTTCAAAAAGCATCCAAGAGTGGGATACCTGCCCGTTCAG AGTGTTTTGGAGGGTGATGCTCTGGAGAGTCCGGCGCCAAGTCCGCAGCACACTACCCACCAGCTGCAGAACGACATGCACTCCCGCCTGGAGATGTAC GCCTCACGACTGGCTCAGGTGGAGTACGGCGGAACGGGCTCGAATTCCACGCCGGACAGTGATGATGAGCACCAACTGATCGCTCAGTATTGCCAGGC TCTGCCCGGAACTTCAAATGGAAGTGCGCCCAAGTCGCCTGTCCAAGTAATGGCCGCCATGGATGCCGAGCAGCGGGAAGAACTAGAGGCGATCATCC GCGATCTTGAGGAGGAAAACGCCAACCTGCAGGCAGAGTACCAGCAATTGTGCAGCAAGGAGCAGAGCGGCATGCCCGAGGACTCCAATGGCATGCA ACACTCCAGTTCCAGCATGACGGGTCTTTCTGGCCAGGGCGAGCAAGGCCAGGACATGATGGCGGAAGCGAAGCTGCTGCGCCAGCACAAGGGACGCC TTGAGGCGCGCATGCAAATCCTGGAGGATCACAATCGGCAACTGGAAGCTCAGTTGCAGCGACTACGCCAACTTCTCGATGAACCCAATGGCGGCGGCA GTAGCGCCACCAGCAGCGGTCTACCCAGCGCTCCCGGTTCCGCATTGAACTCCAAACCCAATACCCTTCAGACCCGGTCGGTGACCGCCCTCCGCAATTR AACACGGACTCACCGGCAAAGATGAACCAACAGAACGGCCACTACTAGCAYAACTCAAAGAACTCAAGTGGCCTAGTGACCGTAATCACGGA	2962
Gallus gallus (X13369.1)	GCCGGGGTTTAGGTGGGAACACACACGGCTGTCAATCTGCAGGGACGAGCCTCATTCGAAAGCGACTGAAGCAGCAAAGTGGTGGGAATATACTTTGC CAAGAAGCGTTTTGCTGTTGTTTTCTTTTCCTCTGTTTTGCGGGGAGATTTCATGTCTGCTCACGTGCTTTGGTATGAGGAAGTGGAAGATGATTATGAAC GGGAAGATGTTCAAAAGAAAACATTCACGAAATGGATAAATGCACAGTTTGCTAAGTGTGGAAGACGTTGCATTGAAGATCTTTTTAATGATTTTCGAG ATGGACGAAAACTTCTGGAGCTCCTGGAATGCCTTACAGGCCAAAAAATTGCAAAAGAAAAGGGCTCCACAAGAGTTCATGCTCTGAACAACGTCAACA AAGCACTGCAGATTTTGCAAAGAAATAATGTTGATTTGGTAAATATTGGGAGCTCAGACATTGTGGACGGCAATCATAAACTGACCCTTGGTTTGATCTG GAATATAATCCTCCACTGGCAGGTCAAAGATGTAATGAAAAACATTATGGCTGGACTGCAGCAGACAAACAGTGAGAAGATTCTGCTGAGCTGGGTCCG TCAATCAACTCGTAATTACCCACAGGTCAATGTTATCAATTTCACCAGTAGCTGGTCTGATGGATTGGCTTTCAATGCACTCCTTCACAGTCACAGACCAG	13575

```
ACCTGTTTGATTGGAATGCTGTTGCTTCTCAGCAGTCACCTGTGCAACGATTAGACCATGCATTTAACATAGCCAGGCAACACCTGGGCATAGAGAAGCT
CCTTGATCCTGAAGACGTTGCAACTGCCTGTCCAGATAAGAAGTCCATCTTAATGTATGTGACTTCCCTCTTCCAAGTTCTGCCACAGCAAGTCACTATGG
AGGCCATCAGGGAGGTGGAAATGCTGCCACGGCACTCAAGGGTCACTACAGAGGAGCACATACAAGTACATCATCAACAGCATTTTTCACAAGAGATCA
CAGTCAATATACCCCAGAGACCTTCACCTTCTCCTAAACCACGGTTCAAAAGTTATGCATATGCACAAACTGCGTATGTCATACCCCCTGACCAAAAAAGG
AGGCAGGTTCCTCCACAGTTTTTAGAAACTGTTGAAAAAAAGAACGTATACCACCACAGTGATGAGGTCCGAAATGGATCTTGACAGCTACCAAACAGCTT
TAGAAGAAGTACTCACGTGGCTTCTCTCTGCTGAAGATGCATTACAAGCACAAGGGGATATATCCAGTGATGTAGAAGTTGTTAAAGAGCAGTTTCATAC
TCATGAGGGTTTCATGATGGAATTAACAGCTCACCAAGGCCGTGTTGGTAATGTTTTGCAAGTTGGAAGTCAACTTCTAGCAATGGGAAAGCTTTCTGAT
GATGAAGAAAATGAAATACAAGAACAAATGAATCTACTCAATTCTCGATGGGAAAGTCTCAGGGTTGCAAGTATGGAAAAACAAAGCAATTTACATAAA
ATCCTAATGGATCTGCAGAATCAGCAACTAGCCCAGTTAGCTGACTGGCTAACAAAAACAGAGGAAAGGACAAAGAAGATAGATTCAGAGCCCTTAGG
TCCGGATCTAGAGGACCTAAAGCGTCAAGTAGAAGAACATAAGGCATTTCAAGATGATTTGGAACAGGAACAAGTGAAGGTGAACTCTCTAACCCATAT
GGTGGTGGTGGTGGATGAAAACAGCGGAGACAAAGCGACTGCTGCACTGGAAGAGCAGCTTCAGCACTTTGGAAGCCGGTGGGCAGCCATCTGCAGA
TGGACAGAAGATAGATGGGTCCTTCTGCAAGATATTCTTAGAAAATGGCAGCACTTTGCAGAAGAGCAGTGCCTTTTTTGATGCATGGCTTACTGAGAAA
GAAGGATCGCTTAGCAAAATCCAGACAAGTGACTTTAAAGATGAAAATGAGATGCTGACTAGTCTACGCAAATTAGCTATCCTAAAAGGAGATATAGAA
ATGAAAAAGCAAATGATGAGTAAACTGAAGTCGCTCAGCAGAGACCTCCTGGTAGCTGTGAAAAATAAAGCAGTGGCTCAGAAGCTGGAAAGCCGGCT
TGAAAATTTTGCCCAGCGCTGGGATAGCCTTGTGCAGAAGTTGGAGAGCGATTCTAAGCAGGTTTCGCAGGCTGTCACTACCACTCAGACATCACTAACA
CAGACAACTGTAATGGAAACTGTCACTATGGTGACCACAAGAGAACAAATCCTGGTTAAGCATGCTAAAGAGGAGCTCCCACCACCACCTCCCCACAAA
AAGAGGCAACTCCTTGTGGATTCGGAAATTAGGAAGAGGTTTGACTCGGATACAACAGAACTCCATAGCTGGATGACACGTTCAGAAGCAGTACTTCAG
AGTCCTGAGTTTGCAATATATCGAAAAGAAGGAAATCTCTCAGATCTCAGAGAAAGAGTCAATGCCATCCAGCGAGAAAAGCCTGAGAAGTACAGAAA
GCTGCAAGATGCCAGCAGATCAGCTGAGGCCCTGGTGGAACAGATGGTGAATGAGGGTCTGAATGCTGACAACATTAGACAAGCCTCAGAGCAGCTGA
AGAGTCGCTGGATTGAGTTCTGTCAGTTACTGAGTGAAAGACTGGTATGGTTGGAGTACCAAAATAGCATCATTGACTTCTATTCCCAGCTGCAGCGACT
GGAGCAGACAGCAATTACAGCAGAAAACTGGCTGAAAGCACAACCCACACCAGCAACAGATCCTGCTACAGTAAAAATTCAGTTGGAAAAATGCAAGG
ACGAAATCATCCGAATGTCAACTCTTCAGCCTCAAATTGAACGGCTTAAGGCTCAGAGTCAAGCTCTGAAAGAGAAAGAACAATGCCCAGTGTTTCTGGA
TGCTGACCTTGCTGCTTTTACCAGCCACTTCAAACAAATACTTGCTGACATGCACACCCAGAGAAAAGCAACTACAGACCATTTTTGACAGTTTGCCTCCTG
CACGCTATAAAGACACAGTGACTACTATACTTTCATGGATCCAGCAGTCAGAAACTAAAGTCTCCATACCTCCAGTTGCAGTGGCTGAATATGAAATCAT
GGAACAGAGACTCGGGGAGCTCAAGGCTCTACAAAGTTCTCTGCAAGAGCAGCAAAAAGGCCTGAAATATCTCAACACAACTGTTGAAGACTTGTCTAG
GAAAGCCCCTGCAGAAGTCAGCCAGAAATACCGATCAGAGGTTGAGTTGATCGTTGGCCGCTGGGAGAAGCTGTCATCACAGTTGGTGGAACATTGCC
AGAAACTGGAGGATCTTATGACTAAACTCCAACGATTCCAGAATGACACAAAAACATTGAAGAAGTGGATGGCTGAAGTAGATGTCTTTCTGAAGGAGG
AATGGCCTGCTCTTGGTGATTCAGAAGCTCTGGAAAAGCAACTTGAGCAGTGTACAGCTTTAGTAAATGATATCCAGACTATCCAGCCGAGTTTGAACAG
TGTTAATGAGATTGGGAAGAAAATGAAGAGGGAAGCAGAGCCAGAATTTGCTTCCAGAATAGCAACAGAACTAAAGGATCTCAATGCTCAATGGGAAC
ATATTTGCCAACAGGCACATGCTAAGAAGGCAGCATTAAAAGGTGGTTTGGATAAGACTGTGAGCCTCAGAAAGGATTTGTCAGAGATGCATGAATGG
ATAACACAAGCTGAGGAAGAATATCTGGAAAGAGATTTTGAGTATAAAACACCCGAAGAATTACAGAAAGCTGTTGAAGAACTGAAGAGAGCAAAGGA
GGATGCCATGCAGAAAGAAGTGAAAGTGAAACTTATTACTGATTCCGTGAATAATTTTATAGCAAAGGCTCCACCTGCAGCTAATGAGGCTTTGAAAAA
GGAGCTTGATGTTCTAATTACCAGCTACCAGAGACTCTGCAGCAGACTGAATGGAAAGTGCAAAACTTTGGAGGAAGTATGGGCATGTTGGCATGAATT
ATTGTCATATTTGGATGCAGAAACAAATGGTTAAATGAGGTGGAATTGAAACTGAAGGCAACTGAAAATATCCAGGGAGGTGCAGAAGAGATTTCTG
AGTCTTTAGATTCTTTGGAACGTTTAATGAGACATCCAGAAGATAATCGCAATCAGATTCGGGAATTAGCTCAGACTTTAACTGATGGTGGAATCTTGGA
TGAACTGATCAATGAGAAACTTGAGAAGTTCAATACTCGATGGGAAGAACTGCAGCAGGAGGCTGTGAGAAGACAAAAGAGTCTTGAACAGAGTATTC
AATCTGCCCAGGAGACTGACAAAACCCTCCGCTTAATTCAAGAGTCTCTTGCTGCTATAGACAAACAGCTGACAGCCTACACTGCAGACAGAGTTGATGC
AGCACAAGTGCCTCAGGAAGCACAGAAAATACAATCTGAATTAACAAGCCATGAGATTAGTTTGGAAGAAATGAAGAAACGAAACCGAGGCAAGGAAT
CTGCAAAAAGAGTTCTTTCCCAAATTGATGTGGCACAGAAAAAGCTGCAGGATGTCTCCATGAAGTTTCGCTTGTTTCAGAAAACCAGCTAATTTTGAACA
GCGTCTACAAGAATGCAAAAGAATTCTAGATGAAGTGAAGTTGCAAGTGCCCAAGTTGGAGACGAAGAGTGTTGGACAGGAAGTAGTGCAGTCACATT
TGGACCACTGCATGAAATTATATAAAAGCCTGAGTGAGGTGAAGTCTGAAGTGGAAACAGTGATAAAAACTGGGAGGCAGATTGTTCAAAAGCAGCAG
ACAGAGAACCCAAAAGAACTGGATGAAAGGCTTACAGCTTTGAAGTTGCAGTATAATGAATTGGGTGCGAAGGTGACAGAAAAAAAGCAGGAGTTAG
AGAAATGCTTGAAATTGTCCCGGAAGCTACGAAAAGAAATTAATTCGCTGACAGAATGGCTTGCAGCAACAGATGTGGAATTGACAAAGAGATCAGCT
GTGCAAGGGATGCCTAGCAATTTGGATGCTGAAATTGCCTGGGGCAAGGCAACACGGAAAGAGATTGAGAAACGCCAAGTCCAGCTTAAGAATATCTG
TGATTTAGGAGAGAATTTGAAGACAGTACTGAAAGGAAAGGAAAGTCTAGTGGAAGATAAACTCAGTCTCCTGAACAGTAATTGGATAGCAGTAACTTC
ACGTGCCGAGGAATGGTTAAATCTGTTAATGGAATATCAAAAGCACATGGAGGCTTTTGATCAGAAAGTAGCTAATGTCACGACTTGGATATATCGTGC
TGAAATACTGTTGGATGAATCTGATAAGCAAAAGCCCCAGCAAAAAGAGGAAACTCTTAAGCGCTTAAAGGCTGAGCTGAATGATATGCATCCAAAGGT
GGACTCTGTGCGTGACCAAGCAGTAGACTTGATGACAAACCGTGGTGATCACTGCAGGAAAGTAATAGAGCCTAAACTATCTGAGCTCAACCATCGATT
TGCTGCCATATCACAAAGAATTAAGAGTGGAAAGCCCTTCATTCCTTTGAAGGAATTGGAGCAATTTGACTTCGATATACAAAAATTGCTTGAACCACTG
```

127

```
GAGGTTGAAATTCAGCAGGGGGTGAATCTGAAAGAGGAGGACTTCAATAAAGATATGAGTGAAGATGATGAGAGCACAGTGAAAGAATTGCTGCAAA
GGGGCGACACGCTTCAGAAAAGAATCACAGATGAGAGAAAACGGGAGGAAATAAAGATAAAACAACAGCTGTTGCAGACTAAACATAATGCTCTCAAG
GACTTGAGGTCTCAAAGAAGAAAAAAGGCTTTAGAGATTTCTCATCAATGGTATCAGTACAAGAGGCAGGCTGATGACCTAATGACATGGCTGGATGAC
ATTGAGAAAAAGTTAGCCAGTCTACCAGACCACAAAGATGAGCAGAAGCTAAAGGAGATTGGTGGAGAGTTGGAGAAGAAGAAGGAAGATCTGAATG
CGGTGAACAGACAGGCTGAACGCCTGTCTAAGGATGGGGCTGCAAAAGCAGTGGAGCCAACCCTGGTACAGCTCAGCAAGCGCTGGCGAGATTTTGAG
AGCAAATTTGCTCAGTTTCGAAGACTCAACTATGCACAAATTCAAACAGTTCTAGAAGATACAACTTTTGTGATGACTGAAAGTATGACTGTGGAAACCA
CTTACGTGCCTTCTACATACCTGGCAGAGATCCTTCAGCTTCTGCAAGCCTTGTCTGAAGTAGAAGAACGCCTTAATTCTCCTGTTCTGCAGGCCAAGGAC
TGTGAGGATCTCTTGAAACAAGAAGAATGTCTCAAGAACATTAAAGATTGCCTGGGGAGACTCCAGGGTCATATAGACATTATTCACAGCAAGAAGACA
CCGGCTTTGCAAAGTGCTACACCACGGGAAACAGCAAATATACAAGACAAGCTGACTCAGCTTAATTCCCAATGGGAGAAAGTTAACAAGATGTACCGG
GACCGGCAGGCACGCTTTGACAAATCCAAGGAAAAGTGGCGGCTTTTTCATTGCGAAATGAAGAGTTTTAATGAGTGGCTAACTGAAACTGAAGAGAA
ACTGTCAAGAGCACAGATAGAGGCTGGAGACGTGGGTCATGTGAAAACCAAGCAATTTCTTCAGGAGCTTCAGGATGGCATTGGGCGACAGCAAACTG
TTGTCAAAACACTGAATGTAACTGGCGAAGAAATTATTGAGCAGTCATCAGCAGCAGATGCTAACGTGCTGAAGGAGCAACTGGGAAATCTGAATACCC
GGTGGCAGGAGATCTGCAGACAGCTGGTAGAGAAAAGAAAGAGGATAGAGGAAGAAAAGAATATTTTATCAGAATTTCAAGGAAGACTTGAACAAGCT
GATTTTATGGTTAGAGGAAACAGAGAACGTCATTGCTATTCCCCTTGAACCAGGGAATGAAGACCAGCTAAGAGACTGCCTTGGCAAAGTAAAGTTAAG
AGTTGAAGAGCTGCTGCCACACAAGGGAATACTGAAACGATTAAATGAAACTGGAGGAACAACGCTTGGAAGTGCATCACTGAACCCAGAAAGAAAAC
ATAAGCTTGAGAGTACACTGAAGGAGGCTAGCCGTCGCTTGTTAAAGGTGTCCAGAGATCTACCAGAGAAGCAAAAAGAAATAGAGATTCTGCTAAAG
GATTTCATCGAACTTAATCAGCAAATAAATCAACTGACACTCTGGATAACACCTGTCAAAAACCAGCTAGAGCTTTATAACCAAGTGGGTCAACCAGGAG
CTTTTGATATTAAGGAAACCGAAGCAGCAGTGCAGGCTAAACAGCCGAATGTGGAAGAGGTTTTGTCTAAAGGGTGTCATTTATATAAGGAAAAACCAG
CCACTCATCCAGTAAAGAAAAAACTAGAAGACTTGAATGCTGACTGGAAGGCAATAAACCACTTAATTCTACAACTGAAGGAGAAGCCAACATTTGGAG
AGCCTGCCCTTACCTCTCCAGGTGTCTTAACTTCTGGTCAAACTGTTGCTGTGGATACACAAGCCAGGGTAACCAAGGAAACCACCAGCTTCACACCAGA
AATGCCATCTTCTGTGCTTTTGGAGGTTCCAGCCTTAGCTGACTTCAATAAGGCATGGGCAGAACTCACTGACTGGCTTTCTCGACTGGATCGAGAGATA
AAAGCTCAGAGAGTGACAGTAGGTGATCTTGATGATATCAACGACATGATCATCAAACAAAAGGCTAACATGCAAGATCTGGAGCAAAGACGTCCCCA
GCTGGATGAACTAATAACTGCAGCACAAAATCTCAAAAACAAGACGAGCAATCAAGAGGCCAGAACAATAATTACTGACCGCATTGAAAAGATACAGA
GCCAGTGGGATGATGTGCACGGATACCTCCAAAACCGAAGACAACAGCTTCATGAGATGCAAAAGGATTCAACACAGTGGCTAGAAGCTAAACAAGAA
GCTGAACAGGTTCTTGAACAAGCAAAAGCAAAGCTTGAGTCATGGAAAGAAATTTCCTATACTGTGGAAGCTCTGAAAAAGCAGAACTCTGAGCTTAAG
CAATTTTCAAAAGAGATACGACAGTGGCAAATGAATATAGAAGGGGTGAATGACGTGGCACTTAAGCCTGTCCGCGATTATTCAGCAGATGACACCAGA
AAAGTAGAACTGATGACAGATAACATTAATGCGACATGGGCTACAATCAATAAGAGGGTTAGTGAACGTGAAGCCGCACTGGAATCAGCTCTACTGATG
TTGCAGGAATTCTACCTGGATCTTGAAAAGTTCCTTGCTTGGCTTACAGAAGCTGAAACAACTGCTAATGTCCTGCAGGATGCTACACACAAGGAAAAGA
CACTAGAGGATCCCCAGATGGTTCGGGAGCTCATGAAGCAGTGGCAGGATCTACAGGCAGAAATTGATGCACATACTGACATCTTCCACAACCTGGATG
AAAACGGGCAGAAAATCCTGAGATCCCTGGAAGGCTCAGAGGATGCTGTCCTGTTGCAGAGACGTCTGGATAACATGAACTTCAGATGGAGTGAGCTT
AGGAAGAAATCTCTAAACATTAGATCTCATTTGGAAGCCAGCACAGACCAGTGGAAGCGTTTACATCTCTCTCTTCAGGAACTTTTGGCATGGCTGCAAT
TGAAGGAGGATGAATTAAAACAGCAAGCACCCATTGGTGGAGATATTCCCACTGTGCAGAAGCAGAATGATGTTCATAGGACTTTCAAGAGGGGAGCTG
AAAACAAAAGAACCTGTTATCATGAATGCACTTGAGACTGTGCGACTCTTCCTGGCAGATCAACCAGTAGAGGGACTGGAAAAGGTCTATCCAGAACCA
AGAGACCTATCACCTGAGGAGAGGGCCCAGAATGTCACTAAAGTTCTCCGAAGGCAAGCAGATGATGTCAGAACTGAGTGGGATAAGCTAAATCTACG
TTCTGCTGATTGGCAAAAGAAGATAGATGATGCTCTTGAAAGACTGCAGGGTCTTCAGGAGGCAATGGATGAACTAGACCTGAAACTGCGCCAGGCTG
AAGCATTCAAGGGATCCTGGCAGCCAGTGGGGGATCTGCTGATAGACTCTCTGCAGGATCACTTAGAAAAAGTCAAGGTTTATCGAGCAGAAATGGTG
CCCCTTAAAGAGAAGGTGCATCAAGTCAATGAGCTGGCTCACCGGTTCGCTCCCCCTGATATTCAGCTCTCCCCATACACTCTCAGCTGTCTGGAGGACCT
GAACACAAGGTGGAAGGTGCTACAGGTGGCCATTGATGAGCGCATCAGGCAACTGCATGAAGCTCACAGGGATTTGGCCCTACTTCCCAGCATTTTCT
TACCACTTCTGTCCAAGGCCCCTGGGAGAGGGCAATCTCGCCAAACAAAGTGCCCTATTACATCAACCATGAGACGCAGACAACCTGCTGGGATCATCCC
AAAATGACCGAGCTCTACCAGTCTTTAGCGGACCTGAACAATGTCAGATTCTCAGCATACAGAACTGCCATGAAGCTCCGCAGGCTGCAGAAAGCTCTCT
GCTTGGATCTCCTGAATCTGTCTGCTGCATGCGATGCCTTGGACCAGCACACAACCTCAAGCAAAATGACCAGCCGATGGATATTCTGCAGATCATTAACTG
CTTGACCACTATTTATGATCGACTGGAACAGGAGCACAATAATCTGGTCAATGTTCCTCTCTGCGTAGACATGTGCCTCAACTGGCTGCTGAATGTCTATG
ACACGGGTCGAACAGGAAGGATCCGTGTCTTATCTTTCAAAACTGGTGTTGTATCCCTTTGTAAAGCACATCTGGAAGATAAGTATAGATACCTGTTCAA
GCAGGTGGCGAGCTCCACTGGCTTCTGTGACCAGCGCCGGCTGGGACTGCTGCTGCACGACTCCATCCAGATCCCACGGCAGCTGGGGGAGGTCGCTTC
GTTTGGGGGCAGCAACATCGAGCCGAGTGTCAGAAGCTGCTTCCAGTTTGCCAATAACAAGCCTGAGATCGAAGCAGCCTTGTTCCTGGACTGGATGAG
GCTGGAACCACAATCCATGGTGTGGCTGCCCGTGCTGCACAGGGTGGCTGCTGCCGAAACTGCCAAACACCAAGCAAAGTGTAACATCTGCAAGGAGT
GCCCCATTATTGGATTCAGGTACAGAAGCTTAAAGCACTTTAACTATGACATCTGCCAAAGTTGCTTCTTCTCTGGCCGTGTTGCAAAAGGTCACAAAATG
CACTATCCCATGGTGGAGTACTGCACACCGACAACTTCTGGAGAAGATGTCCGTGACTTTGCCAAGGTACTAAAAAACAAATTTCGAACAAAAAGATATT
TTGCAAAGCACCCACGAATGGGCTACCTGCCTGTGCAAACTGTCTTGGAGGGAGACAACATGGAAACTCCTGTTACTCTGATCAACTTCTGGCCAGTAGA
```

| | TTCTGCGCTAGCAGAAATGGAGAACAGCAATGGTTCTTACCTAAATGACAGTATTTCACCTAATGAGAGCATCGATGATGAACACTTGTTAATCCAGCAC
TACTGCCAAAGTCTGAACCAGGAATCCCCCCTGAGCCAGCCCCGAAGCCCTGCCCAGATCTTGATTTCTCTGGAGAGTGAAGAAAGAGGTGAACTGGAG
AGAATTCTTGCAGATCTGGAAGAAGAGAATCGAAACTTGCAAGCGGAGTATGACCGTTTGAAGCAACAGCATGATCACAAAGGATTATCTCCACTGCCA
TCCCCACCAGAGATGATGCCAGTTTCTCCACAGAGTCCTCGCGATGCTGAACTCATTGCAGAAGCCAAACTGCTTCGCCAGCACAAGGGCCGCCTGGAG
GCCAGGATGCAGATTCTGGAGGATCACAACAAACAGCTGGAGTCACAGCTGCACAGGCTGAGGCAGCTGCTGGAGCAGCCACAGGCAGATGCCAAGG
TGAATGGTACAACACTATCATCTCCTTCTACCTCTTTGCAGAGGTCAGACAGCAGTCAGCCAATGCTTCTTCGTGTAGTTGGCAGCCAGACTTCAGAAACC
ATGGGCGAGGACGACCTGCTCAGCCCTCCCCAGGACACAAGCACAGGTTTGGAGGAAGTGATGGAGCAGCTTAACAACTCCTTCCCCAGTTCCAGAGG
AAGAAATGCCCCTGGAAAGCCAGTGAGAGAGGCCACAATGTAGGGAGCCTTTTCCACATGGCAGATGACTTGGGAAGAGCGATGGAAACCTTAGTGAC
GGTGATGACTGACGACAAGGTGTTAGAATAGCAAGATGTGTTTTACAACTTCAGGCGTCCCGCATGGTTTTTATAATATTCATACTACAAAGAGGATTAG
ACTGTAAGAGTTTACAAGAAAACAAATCTATATTTTTGTGAAGGGTAGTGGTACTATACTGTAGATTTCAGTAGTTTCCTAAGTCTGTTTCTGTTTTGTTAA
CAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAAATTATAAGAAAACTACATGTAAAATCTTGATAGCTGAATAACTTGCCATTTCTTTATATGGAA
CGCATTTCGGGTTGTTTAAAAAAAAATTTATAACAGTTATAAAGAAAGATTGTAAACTAAAAAGTGTGCTTTATAAAAATAATTGTTTATAGAAACCCCCTT
TAAAAAAAAATACAAAAAAAAAAAATACTGAGGCAGTGCATTTGTTCAGTATCATTTCAGCTCTGCAGCTGCATCAGAGAGAGATTCATGTTCTGTGTTCTTT
TCCTGGCCTGTCAAAGAGGAGAGGAAGAAAAAAGCATTTCCTGCAGCGACACCAAAATGGCCAACTTAACATTTCTCATTGGTTTTGTCTGAGTAGGGCT
GGCTTTAAAATAACTCCCGGACTTCCAGTGAAGATAGTTATTTCACTTCTTTCCAATAAAATTAACGATTACATGAATCACACTAGTTGGCAATAATTAAAT
GAAAGAGCACATGACATGACAGAGCCTGCAAAAAATTAATGGAAAGGAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAGCAGAAGAAAGAAGGAAA
GAATTTAATGATGTCCAATACGCCAGCATTTCCTCTGCTGACTTAACCCTGAATGTCCTTAGTGAAGGAACGTAGAATTGCCTTACCGGATGAAACCTTTG
GACCTTACTGTCCGGTACTCCGCTGGCTGCAGCCAGCGCCCAGAGCTTCAGCAGAAGCTGCGTATGGCTCTCCCTGCCCGCTGCCACCGTTAGCGACACC
TCGCTGCACACGAGAGAAGAAAGGGAGAAGTTCCCTCCTGACCCCACGCGGATCAGCTTCTGACATGAAGCACAAGTGGGTTATTCACAGAGCACCTTT
GCTTGTATGGCTTCCAATGCTGTTAATGAACGTAGAGCTACTTGATCCTCTGGAAAATAAAACCAGACCTTTCCGCAATGTCTGTCTCAGCCTGCTAGAGC
AGCTGGTCCCAAGGAGCTGTAATGCAGCAAGCAACCTTTGAGGAAGGCTGTAGAAGAGTTACTGCTGTTCTTATATGTGTATGGTGATCAGGTGAAGGT
AGCCACACTGCTCCCCGAGATATTTATCTTCAAAACAAAGGCCCTGATGCTTTCTCCTTCCCTGTTTCTCTCTACAAGTAATACCGCACAGTTAACAATAAA
GTGACAAAGAGAGGAACTGATAGAAAAACAGATTAGCCAGCGTAACTGCAGCTCTTGAGAGGTAGGTTTGGGTATGAGGTGCTGTCAGCAAAAGCTTT
ACCACTTCATTATCTTAACCAGCTTTTAATTTTCTCTTAGATACATAGAAACACACCGTCTGACTTTATATTGCAAAGTAAAAATCTGTTACCTCTTCTTGTA
TGAAGTGTAATTAAACTTGGGAAAGATTTTATTTTGAATGCTAGTTAACATTGAACATGGCTAGTCATGTCATTTCTACCTCACTTTGGTTTTGTGGTGTTC
CTGACAACTACGTACAGATATGTTACATCATCACCACTCGTCCATTGTGTGAACGGGTCTCATCATTTAAGTAAGAATGCTAATAATTCACAACGAGTAAG
TACAATCCCCCCAAACCTCTGCTTTCGTAATGAAGAGCTTTGTTTAACACTTTCTTGCAGTAGGAACCCCATGTAATTCTCGTGGTCTTCTTTACCAGAAGG
TTTGTTCTGAGATCTGGGAAAGAGATGGGTTTCTTCTCGTTTTTCCCCCATGTGCCCCCCCCTTAAGCCTTCTCTTCAGAAAACAGAAGGAGGAAAATGGAT
TGCACGGTTGTTTTCATCAGGGTTTTGTTTTGTTTGTTTTACTTTCTAGGTCTTAGTTTTGGAGTGAGTCTGTCAAAAATATTTAAGAAGTTGAAAGCTTTA
TCACTGCTTTTTGAAGCATAATTTGGGAATTGTTTCATATTCCTTATAACGACAAAGTATTAAACTGTAAAACATAATCAGTGTAATTGAATACATAAATAC
CATATGGCATGTTATTTCATTGTTACTCAGTACTGGTTTATTACTTGGATATCATAATATATTGTGTTTTAACACCAACACTGTAACATTTACTAATTATTTTT
ATAAACTTCTGTTTTACTGCATTTTTCACAACATATCAGATTTCACCAAATATATGCCTTACTAATG | |
| *Homo sapiens*
(M18533.1) | GGGATTCCCTCACTTTCCCCCTACAGGACTCAGATCTGGGAGGCAATTACCTTCGGAGAAAAACGAATAGGAAAAACTGAAGTGTTACTTTTTTTAAAGC
TGCTGAAGTTTGTTGGTTTCTCATTGTTTTTAAGCCTACTGGAGCAATAAAGTTTGAAGAACTTTTACCAGGTTTTTTTTATCGCTGCCTTGATATACACTTT
TCAAAATGCTTTGGTGGGAAGAAGTAGAGGACTGTTATGAAAGAGAAGATGTTCAAAAGAAAACATTCACAAAATGGGTAAATGCACAATTTTCTAAGT
TTGGGAAGCAGCATATTGAGAACCTCTTCAGTGACCTACAGGATGGGAGGCGCCTCCTAGACCTCCTCGAAGGCCTGACAGGGCAAAAACTGCCAAAA
GAAAAAGGATCCACAAGAGTTCATGCCCTGAACAATGTCAACAAGGCACTGCGGGTTTGTCAGAACAATAATGTTGATTTAGTGAATATTGGAAGTACT
GACATCGTAGATGGAAATCATAAACTGACTCTTGGTTTGATTTGGAATATAATCCTCCACTGGCAGGTCAAAAATGTAATGAAAAATATCATGGCTGGAT
TGCAACAAACCAACAGTGAAAAGATTCTCCTGAGCTGGGTCCGACAATCAACTCGTAATTATCCACAGGTTAATGTAATCAACTTCACCACCAGCTGGTCT
GATGGCCTGGCTTTGAATGCTCTCATCCATAGTCATAGGCCAGACCTATTTGACTGGAATAGTGTGGTTTGCCAGCAGTCAGCCACACAACGACTGGAAC
ATGCATTCAACATCGCCAGATATCAATTAGGCATAGAGAAACTACTCGATCCTGAAGATGTTGATACCACCTATCCAGATAAGAAGTCCATCTTAATGTAC
ATCACATCACTCTTCCAAGTTTTGCCTCAACAAGTGAGCATTGAAGCCATCCAGGAAGTGGAAATGTTGCCAAGGCCACCTAAAGTGACTAAAGAAGAAC
ATTTTCAGTTACATCATCAAATGCACTATTCTCAACAGATCACGGTCAGTCTAGCACAGGGATATGAGAGAACTTCTTCCCCCTAAGCCTCGATTCAAGAGC
TATGCCTACACACAGGCTGCTTATGTCACCACCTCTGACCCTACACGGAGCCCATTTCCTTCACAGCATTTGGAAGCTCCTGAAGACAAGTCATTTGGCAG
TTCATTGATGGAGAGTGAAGTAAACCTGGACCGTTATCAAACAGCTTTAGAAGAAGTATTATCGTGGCTTCTTTCTGCTGAGGACACATTGCAAGCACAA
GGAGAGATTTCTAATGATGTGGAAGTGGTGAAAGACCAGTTTCATACTCATGAGGGGTACATGATGGATTTGACAGCCCATCAGGGCCGGGTTGGTAA | 13957 |

TATTCTACAATTGGGAAGTAAGCTGATTGGAACAGGAAAATTATCAGAAGATGAAGAAACTGAAGTACAAGAGCAGATGAATCTCCTAAATTCAAGATG
GGAATGCCTCAGGGTAGCTAGCATGGAAAAACAAAGCAATTTACATAGAGTTTTAATGGATCTCCAGAATCAGAAACTGAAAGAGTTGAATGACTGGCT
AACAAAAACAGAAGAAAGAACAAGGAAATGGAGGAAGAGCCTCTTGGACCTGATCTTGAAGACCTAAAACGCCAAGTACAACAACATAAGGTGCTTC
AAGAAGATCTAGAACAAGAACAAGTCAGGGTCAATTCTCTCACTCACATGGTGGTGGTAGTTGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTGG
AAGAACAACTTAAGGTATTGGGAGATCGATGGGCAAACATCTGTAGATGGACAGAAGACCGCTGGGTTCTTTTACAAGACATCCTTCTCAAATGGCAAC
GTCTTACTGAAGAACAGTGCCTTTTTAGTGCATGGCTTTCAGAAAAAGAAGATGCAGTGAACAAGATTCACACAACTGGCTTTAAAGATCAAAATGAAAT
GTTATCAAGTCTTCAAAAACTGGCCGTTTTAAAAGCGGATCTAGAAAAGAAAAAGCAATCCATGGGCAAACTGTATTCACTCAAACAAGATCTTCTTTCA
ACACTGAAGAATAAGTCAGTGACCCAGAAGACGGAAGCATGGCTGGATAACTTTGCCCGGTGTTGGGATAATTTAGTCCAAAAACTTGAAAAGAGTACA
GCACAGATTTCACAGGCTGTCACCACCACTCAGCCATCACTAACACAGACAACTGTAATGGAAACAGTAACTACGGTGACCACAAGGGAACAGATCCTG
GTAAAGCATGCTCAAGAGGAACTTCCACCACCACCTCCCCAAAAGAAGAGGCAGATTACTGTGGATTCTGAAATTAGGAAAAGGTTGGATGTTGATATA
ACTGAACTTCACAGCTGGATTACTCGCTCAGAAGCTGTGTTGCAGAGTCCTGAATTTGCAATCTTTCGGAAGGAAGGCAACTTCTCAGACTTAAAAGAAA
AAGTCAATGCCATAGAGCGAGAAAAAGCTGAGAAGTTCAGAAAACTGCAAGATGCCAGCAGATCAGCTCAGGCCCTGGTGGAACAGATGGTGAATGA
GGGTGTTAATGCAGATAGCATCAAACAAGCCTCAGAACAACTGAACAGCCGGTGGATCGAATTCTGCCAGTTGCTAAGTGAGAGACTTAACTGGCTGGA
GTATCAGAACAACATCATCGCTTTCTATAATCAGCTACAACAATTGGAGCAGATGACAACTACTGCTGAAAACTGGTTGAAAATCCAACCCACCACCCCAT
CAGAGCCAACAGCAATTAAAAGTCAGTTAAAAATTTGTAAGGATGAAGTCAACCGGCTATCAGGTCTTCAACCTCAAATTGAACGATTAAAAATTCAAAG
CATAGCCCTGAAAGAGAAAGGACAAGGACCCATGTTCCTGGATGCAGACTTTGTGGCCTTTACAAATCATTTTAAGCAAGTCTTTTCTGATGTGCAGGCC
AGAGAGAAAGAGCTACAGACAATTTTTGACACTTTGCCACCAATGCGCTATCAGGAGACCATGAGTGCCATCAGGACATGGGTCCAGCAGTCAGAAACC
AAACTCTCCATACCTCAACTTAGTGTCACCGACTATGAAATCATGGAGCAGAGACTCGGGGAATTGCAGGCTTTACAAAGTTCTCTGCAAGAGCAACAAA
GTGGCCTATACTATCTCAGCACCACTGTGAAAGAGATGTCGAAGAAAGCGCCCTCTGAAATTAGCCGGAAATATCAATCAGAATTTGAAGAAATTGAGG
GACGCTGGAAGAAGCTCTCCTCCCAGCTGGTTGAGCATTGTCAAAAGCTAGAGGAGCAAATGAATAAACTCCGAAAAATTCAGAATCACATACAAACCC
TGAAGAAATGGATGGCTGAAGTTGATGTTTTTCTGAAGGAGGAATGGCCTGCCCTTGGGGATTCAGAAATTCTAAAAAAGCAGCTGAAACAGTGCAGA
CTTTTAGTCAGTGATATTCAGACAATTCAGCCCAGTCTAAACAGTGTCAATGAAGGTGGGCAGAAGATAAAGAATGAAGCAGAGCCAGAGTTTGCTTCG
AGACTTGAGACAGAACTCAAAGAACTTAACACTCAGTGGGATCACATGTGCCAACAGGTCTATGCCAGAAAGGAGGCCTTGAAGGGAGGTTTGGAGAA
AACTGTAAGCCTCCAGAAAGATCTATCAGAGATGCACGAATGGATGACACAAGCTGAAGAAGAGTATCTTGAGAGAGATTTTGAATATAAAACTCCAGA
TGAATTACAGAAAGCAGTTGAAGAGATGAAGAGAGCTAAAGAAGAGGCCCAACAAAAAGAAGCGAAAGTGAAACTCCTTACTGAGTCTGTAAATAGTG
TCATAGCTCAAGCTCCACCTGTAGCACAAGAGGCCTTAAAAAAGGAACTTGAAACTCTAACCACCAACTACCAGTGGCTCTGCACTAGGCTGAATGGGA
AATGCAAGACTTTGGAAGAAGTTTGGGCATGTTGGCATGAGTTATTGTCATACTTGGAGAAAGCAAACAAGTGGCTAAATGAAGTAGAATTTAAACTTA
AAACCACTGAAAACATTCCTGGCGGAGCTGAGGAAATCTCTGAGGTGCTAGATTCACTTGAAAATTTGATGCGACATTCAGAGGATAACCCAAATCAGA
TTCGCATATTGGCACAGACCCTAACAGATGGCGGAGTCATGGATGAGCTAATCAATGAGGAACTTGAGACATTTAATTCTCGTTGGAGGGAACTACATG
AAGAGGCTGTAAGGAGGCAAAAGTTGCTTGAACAGAGCATCCAGTCTGCCCAGGAGACTGAAAAATCCTTACACTTAATCCAGGAGTCCCTCACATTCA
TTGACAAGCAGTTGGCAGCTTATATTGCAGACAAGGTGGACGCAGCTCAAATGCCTCAGGAAGCCCAGAAAATCCAATCTGATTTGACAAGTCATGAGA
TCAGTTTAGAAGAAATGAAGAAACATAATCAGGGGAAGGAGGCTGCCCAAAGAGTCCTGTCTCAGATTGATGTTGCACAGAAAAAATTACAAGATGTCT
CCATGAAGTTTCGATTATTCCAGAAACCAGCCAATTTTGAGCTGCGTCTACAAGAAAGTAAGATGATTTTAGATGAAGTGAAGATGCACTTGCCTGCATT
GGAAACAAAGAGTGTGGAACAGGAAGTAGTACAGTCACAGCTAAATCATTGTGTGAACTTGTATAAAAGTCTGAGTGAAGTGAAGTCTGAAGTGGAAA
TGGTGATAAAGACTGGACGTCAGATTGTACAGAAAAAGCAGACGGAAATCCCAAAGAACTTGATGAAAGAGTAACAGCTTTGAAATTGCATTATAATG
AGCTGGGAGCAAAGGTAACAGAAAGAAAGCAACAGTTGGAGAAATGCTTGAAATTGTCCCGTAAGATGCGAAAGGAAATGAATGTCTTGACAGAATG
GCTGGCAGCTACAGATATGGAATTGACAAAGAGATCAGCAGTTGAAGGAATGCCTAGTAATTTGGATTCTGAAGTTGCCTGGGGAAAGGCTACTCAAA
AAGAGATTGAGAAACAGAAGGTGCACCTGAAGAGTATCACAGAGGTAGGAGAGGCCTTGAAAACAGTTTTGGGCAAGAAGGAGACGTTGGTGGAAG
ATAAACTCAGTCTTCTGAATAGTAACTGGATAGCTGTCACCTCCCGAGCAGAAGAGTGGTTAAATCTTTTGTTGGAATACCAGAAACACATGGAAACTTT
TGACCAGAATGTGGACCACATCACAAAGTGGATCATTCAGGCTGACACACTTTTGGATGAATCAGAGAAAAAGAAACCCCAGCAAAAAGAAGACGTGC
TTAAGCGTTTAAAGGCAGAACTGAATGACATACGCCCAAAGGTGGACTCTACACGTGACCAAGCAGCAAACTTGATGGCAAACCGCGGTGACCACTGCA
GGAAATTAGTAGAGCCCCAAATCTCAGAGCTCAACCATCGATTTGCAGCCATTTCACACAGAATTAAGACTGGAAAGGCCTCCATTCCTTTGAAGGAATT
GGAGCAGTTTAACTCAGATATACAAAAATTGCTTGAACCACTGGAGGCTGAAATTCAGCAGGGGGTGAATCTGAAAGAGGAAGACTTCAATAAAGATA
TGAATGAAGACAATGAGGGTACTGTAAAAGAATTGTTGCAAAGAGGAGACAACTTACAACAAAGAATCACAGATGAGAGAAAGAGAGAGGAAATAAA
GATAAAACAGCAGCTGTTACAGACAAAACATAATGCTCTCAAGGATTTGAGGTCTCAAAGAAGAAAAAAGGCTCTAGAAATTTCTCATCAGTGGTATCA
GTACAAGAGGCAGGCTGATGATCTCCTGAAATGCTTGGATGACATTGAAAAAAAATTAGCCAGCCTACCTGAGCCCAGAGATGAAAGGAAAATAAAGG
AAATTGATCGGGAATTGCAGAAGAAGAAAGAGGAGCTGAATGCAGTGCGTAGGCAAGCTGAGGGCTTGTCTGAGGATGGGGCCGCAATGGCAGTGG
AGCCAACTCAGATCCAGCTCAGCAAGCGCTGGCGGGAAATTGAGAGCAAATTTGCTCAGTTTCGAAGACTCAACTTTGCACAAATTCACACTGTCCGTGA
AGAAACGATGATGGTGATGACTGAAGACATGCCTTTGGAAATTTCTTATGTGCCTTCTACTTATTTGACTGAAATCACTCATGTCTCACAAGCCCTATTAG

AAGTGGAACAACTTCTCAATGCTCCTGACCTCTGTGCTAAGGACTTTGAAGATCTCTTTAAGCAAGAGGAGTCTCTGAAGAATATAAAAGATAGTCTACA
ACAAAGCTCAGGTCGGATTGACATTATTCATAGCAAGAAGACAGCAGCATTGCAAAGTGCAACGCCTGTGGAAAGGGTGAAGCTACAGGAAGCTCTCT
CCCAGCTTGATTTCCAATGGGAAAAAGTTAACAAAATGTACAAGGACCGACAAGGGCGATTTGACAGATCTGTTGAGAAATGGCGGCGTTTTCATTATG
ATATAAAGATATTTAATCAGTGGCTAACAGAAGCTGAACAGTTTCTCAGAAAGACACAAATTCCTGAGAATTGGGAACATGCTAAATACAAATGGTATCT
TAAGGAACTCCAGGATGGCATTGGGCAGCGGCAAACTGTTGTCAGAACATTGAATGCAACTGGGGAAGAAATAATTCAGCAATCCTCAAAAACAGATG
CCAGTATTCTACAGGAAAAATTGGGAAGCCTGAATCTGCGGTGGCAGGAGGTCTGCAAACAGCTGTCAGACAGAAAAAAGAGGCTAGAAGAACAAAA
GAATATCTTGTCAGAATTTCAAAGAGATTTAAATGAATTTGTTTTATGGTTGGAGGAAGCAGATAACATTGCTAGTATCCCACTTGAACCTGGAAAAGAG
CAGCAACTAAAAGAAAAGCTTGAGCAAGTCAAGTTACTGGTGGAAGAGTTGCCCCTGCGCCAGGGAATTCTCAAACAATTAAATGAAACTGGAGGACC
CGTGCTTGTAAGTGCTCCCATAAGCCCAGAAGAGCAAGATAAACTTGAAAATAAGCTCAAGCAGACAAATCTCCAGTGGATAAAGGTTTCCAGAGCTTT
ACCTGAGAAACAAGGAGAAATTGAAGCTCAAATAAAAGACCTTGGGCAGCTTGAAAAAAAGCTTGAAGACCTTGAAGAGCAGTTAAATCATCTGCTGCT
GTGGTTATCTCCTATTAGGAATCAGTTGGAAATTTATAACCAACCAAACCAAGAAGGACCATTTGACGTTCAGGAAACTGAAATAGCAGTTCAAGCTAAA
CAACCGGATGTGGAAGAGATTTTGTCTAAAGGGCAGCATTTGTACAAGGAAAAACCAGCCACTCAGCCAGTGAAGAGGAAGTTAGAAGATCTGAGCTC
TGAGTGGAAGGCGGTAAACCGTTTACTTCAAGAGCTGAGGGCAAAGCAGCCTGACCTAGCTCCTGGACTGACCACTATTGGAGCCTCTCCTACTCAGAC
TGTTACTCTGGTGACACAACCTGTGGTTACTAAGGAAACTGCCATCTCCAAACTAGAAATGCCATCTTCCTTGATGTTGGAGGTACCTGCTCTGGCAGATT
TCAACCGGGCTTGGACAGAACTTACCGACTGGCTTTCTCTGCTTGATCAAGTTATAAAATCACAGAGGGTGATGGTGGGTGACCTTGAGGATATCAACG
AGATGATCATCAAGCAGAAGGCAACAATGCAGGATTTGGAACAGAGGCGTCCCCAGTTGGAAGAACTCATTACCGCTGCCCAAAATTTGAAAAACAAG
ACCAGCAATCAAGAGGCTAGAACAATCATTACGGATCGAATTGAAAGAATTCAGAATCAGTGGGATGAAGTACAAGAACACCTTCAGAACCGGAGGCA
ACAGTTGAATGAAATGTTAAAGGATTCAACACAATGGCTGGAAGCTAAGGAAGAAGCTGAGCAGGTCTTAGGACAGGCCAGAGCCAAGCTTGAGTCAT
GGAAGGAGGGTCCCTATACAGTAGATGCAATCCAAAAGAAAATCACAGAAACCAAGCAGTTGGCCAAAGACCTCCGCCAGTGGCAGACAAATGTAGAT
GTGGCAAATGACTTGGCCCTGAAACTTCTCCGGGATTATTCTGCAGATGATACCAGAAAAGTCCACATGATAACAGAGAATATCAATGCCTCTTGGAGAA
GCATTCATAAAAGGGTGAGTGAGCGAGAGGCTGCTTTGGAAGAAACTCATAGATTACTGCAACAGTTCCCCCTGGACCTGGAAAAGTTTCTTGCCTGGC
TTACAGAAGCTGAAACAACTGCCAATGTCCTACAGGATGCTACCCGTAAGGAAAGGCTCCTAGAAGACTCCAAGGGAGTAAAAGAGCTGATGAAACAA
TGGCAAGACCTCCAAGGTGAAATTGAAGCTCACACAGATGTTTATCAACCTGGATGAAAACAGCCAAAAAATCCTGAGATCCCTGGAAGGTTCCGAT
GATGCAGTCCTGTTACAAAGACGTTTGGATAACATGAACTTCAAGTGGAGTGAACTTCGGAAAAAGTCTCTCAACATTAGGTCCCATTTGGAAGCCAGTT
CTGACCAGTGGAAGCGTCTGCACCTTTCTCTGCAGGAACTTCTGGTGTGGCTACAGCTGAAAGATGATGAATTAAGCCGGCAGGCACCTATTGGAGGCG
ACTTTCCAGCAGTTCAGAAGCAGAACGATGTACATAGGGCCTTCAAGAGGGGAATTGAAAACTAAAGAACCTGTAATCATGAGTACTCTTGAGACTGTAC
GAATATTTCTGACAGAGCAGCCTTTGGAAGGACTAGAGAAACTCTACCAGGAGCCCAGAGAGCTGCCTCCTGAGGAGAGAGCCCAGAATGTCACTCGG
CTTCTACGAAAGCAGGCTGAGGAGGTCAATACTGAGTGGGAAAAATTGAACCTGCACTCCGCTGACTGGCAGAGAAAAATAGATGAGACCCTTGAAAG
ACTCCAGGAACTTCAAGAGGCCACGGATGAGCTGGACCTCAAGCTGCGCCAAGCTGAGGTGATCAAGGGATCCTGGCAGCCCGTGGGCGATCTCCTCA
TTGACTCTCTCCAAGATCACCTCGAGAAAGTCAAGGCACTTCGAGGAGAAATTGCGCCTCTGAAAGAGAACGTGAGCCACGTCAATGACCTTGCTCGCC
AGCTTACCACTTTGGGCATTCAGCTCTCACCGTATAACCTCAGCACTCTGGAAGACCTGAACACCAGATGGAAGCTTCTGCAGGTGGCCGTCGAGGACCG
AGTCAGGCAGCTGCATGAAGCCCACAGGGACTTTGGTCCAGCATCTCAGCACTTTCTTTCCACGTCTGTCCAGGGTCCCTGGGAGAGAGCCATCTCGCCA
AACAAAGTGCCCTACTATATCAACCACGAGACTCAAACAACTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGCTGACCTGAATAATG
TCAGATTCTCAGCTTATAGGACTGCCATGAAATCCGAAGACTGCAGAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTCAGCTGCATGTGATGCCTTGGA
CCAGCACAACCTCAAGCAAAATGACCAGCCCATGGATATCCTGCAGATTATTAATTGTTTGACCACTATTTATGACCGCCTGGAGCAAGAGCACAACAAT
TTGGTCAACGTCCCTCTCTGCGTGGATATGTGTCTGAACTGGCTGCTGAATGTTTATGATACGGGACGAACAGGGAGGATCCGTGTCCTGTCTTTTAAAA
CTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAGTACAGATACCTTTTCAAGCAAGTGGCAAGTTCAACAGGATTTTGTGACCAGCGCAGGCT
GGGCCTCCTTCTGCATGATTCTATCCAAATTCCAAGACAGTTGGGTGAAGTTGCATCCTTTGGGGGCAGTAACATTGAGCCAAGTGTCCGGAGCTGCTTC
CAATTTGCTAATAATAAGCCAGAGATCGAAGCGGCCCTCTTCCTAGACTGGATGAGACTGGAACCCCAGTCCATGGTGTGGCTGCCCGTCCTGCACAGA
GTGGCTGCTGCAGAAACTGCCAAGCATCAGGCCAAATGTAACATCTGCAAAGAGTGTCCAATCATTGGATTCAGGTACAGGAGTCTAAAGCACTTTAATT
ATGACATCTGCCAAAGCTGCTTTTTTTCTGGTCGAGTTGCAAAAGGCCATAAAATGCACTATCCCATGGTGGAATATTGCACTCCGACTACATCAGGAGA
AGATGTTCGAGACTTTGCCAAGGTACTAAAAAACAAATTTCGAACCAAAAGGTATTTTGCGAAGCATCCCCGAATGGGCTACCTGCCAGTGCAGACTGTC
TTAGAGGGGGACAACATGGAAACTCCCGTTACTCTGATCAACTTCTGGCCAGTAGATTCTGCGCCTGCCTCGTCCCCTCAGCTTTCACACGATGATACTCA
TTCACGCATTGAACATTATGCTAGCAGGCTAGCAGAAATGGAAAACAGCAATGGATCTTATCTAAATGATAGCATCTCTCCTAATGAGAGCATAGATGAT
GAACATTTGTTAATCCAGCATTACTGCCAAAGTTTGAACCAGGACTCCCCCCTGAGCCAGCCTCGTAGTCCTGCCCAGATCTTGATTTCCTTAGAGAGTGA
GGAAAGAGGGGAGCTAGAGAGAATCCTAGCAGATCTTGAGGAAGAAAACAGGAATCTGCAAGCAGAATATGACCGTCTAAAGCAGCAGCACGAACAT
AAAGGCCTGTCCCCACTGCCGTCCCCTCCTGAAATGATGCCCACCTCTCCCCAGAGTCCCCGGGATGCTGAGCTCATTGCTGAGGCCAAGCTACTGCGTC
AACACAAAGGCCGCCTGGAAGCCAGGATGCAAATCCTGGAAGACCACAATAAACAGCTGGAGTCACAGTTACACAGGCTAAGGCAGCTGCTGGAGCAA
CCCCAGGCAGAGGCCAAAGTGAATGGCACAACGGTGTCCTCTCCTTCTACCTCTCTACAGAGGTCCGACAGCAGTCAGCCTATGCTGCTCCGAGTGGTTG

	GCAGTCAAACTTCGGACTCCATGGGTGAGGAAGATCTTCTCAGTCCTCCCCAGGACACAAGCACAGGGTTAGAGGAGGTGATGGAGCAACTCAACAACT CCTTCCCTAGTTCAAGAGGAAGAAATACCCCTGGAAAGCCAATGAGAGAGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGC GATGGAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACTCCTGATTCCCGCATGGTTTTTATAATATTCATACAACAA AGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTG TTTTGTTAACAATGGCAGGTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAAGAAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCT TTATATGGAACGCATTTTGGGTTGTTTAAAAAATTTATAACAGTTATAAAGAAAGATTGTAAACTAAAGTGTGCTTTATAAAAAAAAGTTGTTTATAAAAAC CCCTAAAAACAAAACAAACACACACACACACACATACACACACACACACAAAACTTTGAGGCAGCGCATTGTTTTGCATCCTTTTGGCGTGATATCCATAT GAAATTCATGGCTTTTTCTTTTTTTGCATATTAAAGATAAGACTTCCTCTACCACCACACCCAAATGACTACTACACACTGCTCATTTGAGAACTGTCAGCTG AGTGGGGCAGGCTTGAGTTTTCATTTCATATATCTATATGTCTATAAGTATATAAATACTATAGTTATATAGATAAAGAGATACGAATTTCTATAGACTGA CTTTTTCCATTTTTTAAATGTTCATGTCACATCCTAATAGAAAGAAATTACTTCTAGTCAGTCATCCAGGCTTACCTGCTTGGTCTAGAATGGATTTTTCCCG GAGCCGGAAGCCAGGAGGAAACTACACCACACTAAAACATTGTCTACAGCTCCAGATGTTTCTCATTTTAAACAACTTTCCACTGACAACGAAAGTAAAG TAAAGTATTGGATTTTTTTAAAGGGAACATGTGAATGAATACACAGGACTTATTATATCAGAGTGAGTAATCGGTTGGTTGGTTGATTGATTGATTGATT GATACATTCAGCTTCCTGCTGCTAGCAATGCCACGATTTAGATTTAATGATGCTTCAGTGGAAATCAATCAGAAGGTATTCTGACCTTGTGAACATCAGAA GGTATTTTTTAACTCCCAAGCAGTAGCAGGACGATGATAGGGCTGGAGGGCTATGGATTCCCAGCCCATCCCTGTGAAGGAGTAGGCCACTCTTTAAGT GAAGGATTGGATGATTGTTCATAATACATAAAGTTCTCTGTAATTACAACTAAATTATTATGCCCTCTTCTCACAGTCAAAAGGAACTGGGTGGTTTGGTT TTTGTTGCTTTTTTAGATTTATTGTCCCATGTGGGATGAGTTTTTAAATGCCACAAGACATAATTTAAAATAAATAAACTTTGGGAAAAGGTGTAAGACAG TAGCCCCATCACATTTGTGATACTGACAGGTATCAACCCAGAAGCCCATGAACTGTGTTTCCATCCTTTGCATTTCTCTGCGAGTAGTTCCACACAGGTTTG TAAGTAAGTAAGAAAGAAGGCAAATTGATTCAAATGTTACAAAAAAAACCCTTCTTGGTGGATTAGACAGGTTAAATATATAAACAAACAAACAAAAATT GCTCAAAAAAGAGGAGAAAAGCTCAAGAGGAAAAGCTAAGGACTGGTAGGAAAAAGCTTTACTCTTTCATGCCATTTTATTTCTTTTTGATTTTTAAATC ATTCATTCAATAGATACCACCGTGTGACCTATAATTTTGCAAATCTGTTACCTCTGACATCAAGTGTAATTAGCTTTTGGAGAGTGGGCTGACATCAAGTG TAATTAGCTTTTGGAGAGTGGGTTTTGTCCATTATTAATAATTAATTAATTAACATCAAACACGGCTTCTCATGCTATTTCTACCTCACTTTGGTTTTGGGGT GTTCCTGATAATTGTGCACACCTGAGTTCACAGCTTCACCACTTGTCCATTGCGTTATTTTCTTTTTCCTTTATAATTCTTTCTTTTTCCTTCATAATTTTCAAA AGAAAACCCAAAGCTCTAAGGTAACAAATTACCAAATTACATGAAGATTTGGTTTTTGTCTTGCATTTTTTTCCTTTATGTGACGCTGGACCTTTTCTTTACC CAAGGATTTTTAAAACTCAGATTTAAAACAAGGGGTTACTTTACATCCTACTAAGAAGTTTAAGTAAGTAAGTTTCATTCTAAAATCAGAGGTAAATAGA GTGCATAAATAATTTTGTTTTAATCTTTTTGTTTTTCTTTTAGACACATTAGCTCTGGAGTGAGTCTGTCATAATATTTGAACAAAAATTGAGAGCTTTATTG CTGCATTTTAAGCATAATTAATTTGGACATTATTTCGTGTTGTGTTCTTTATAACCACCGAGTATTAAACTGTAAATCATAATGTAACTGAAGCATAAACAT CACATGGCATGTTTTGTCATTGTTTTCAGGTACTGAGTTCTTACTTGAGTATCATAATATATTGTGTTTTAACACCAACACTGTAACATTTACGAATTATTTT TTTAAACTTCAGTTTTTACTGCATTTTCACAACATATCAGACTTCACCAAATATATGCCTTACTATTGTATTATAGTACTGCTTTACTGTGTATCTCAATAAAG CACGCAGTTATGTTAC	
Pectinidae sp *(X99738.1)*	CATCCAATGATGACCGACCTTATGCAAGCCTTAGCGGACCTTAACAACATAAGATTTGCTGCCTACAGGACGGCCATGAAACTCAGGATGTTACAGAAGA AACTCCGGTTGGATCTTCTCGCCATCAATGTTGCTGTGGAAGCATTTGACCACCACAACTTACGGGGTCGTAACGACAGTCTTATGAACGTTATAGAGAT GATTAACTGTCTGTCCACTATGTACGACAATGTGTCAGTAGAACATCCGGATCTTGTCAATGTTCCACTCTGTGTCGATTTGGTGCTTAACTGGGTGCTCA ATGTGTACGACATAGCAAGGAGTGGGAAAGTGCGAGTTCTCTCCTTAAAAGTGGGTATCATCCTTATGTGTAACGGCCACTTAGATGAYAAATATAGATT TTTGTTCCGGTTAATAGCTGACACCAATGGTTTCTCCGACCAGCGGCGACTCGGCCTCCTCCTCCACGACAGCATGCAGGTACCTCGGCAGCTAGGAGAG GTGGCGTCCTTCGGAGGCAGCAATATAGAACCAAGTGTCCGCAGCTGTTTTGAAAAGGCCAATGGTCGACCAGAAATCCAGGCCACACACTTCCTAGAA TGGCTCAAGATGGAGCCACAGTCCCRGTGTGGCTTCCGGTCTTGCATCGCCTCGCAGCTTCCGAAACTGCCAAACACCAAGCCAAGTGCAATATTTGTA AAGAATTTCCAATAGTTGGCTTTAGATACCGATGCCTCAGGTGCTTTAACTTTGATGTTTGCCAAAACTGCTTTTTTCTCAGGAAGGAAAGCCAAGACCCAT GAACTTTCACACCCTATGCAAGAATACTGTACCGCAACAACCTCAGGAGAAGATGTCCGAGATTTTACTAAGATGTTAAGAAATAAGTTTAAGACAAAGA AACATTTTAAGAAACATCCTCGTCTTGGTTTCCTCCCAGTACAGACGGTGCTTGAAGGTGATGTTCTTGAAAGTCCATCACCATCACCCCAACATAGCATA TCATCACAGGACATGCATTCCCGCCTTGAACTTTATGCAAACAGGTTGGCTGAAGTGGAGCAGAGACAGGCATCGTCCACTCCCGACTCCGATGACGAG CATAACCTGATTGCCCAGTACTGCCAAAGTCTTAACGGAAATACTTCCCATACACTGAAGAGTCCAATGCAAATCATGATGGCAGTAGATTCAGATCAGA AGTCGGAACTGGAGGCCATGATCAAGGATCTCGAGGAGGAAAACAAGACACTACAGCAAGAGTACGACCGCCTGAAACAAGCTAATGAACTACGAGA AAACGGCAACGGATTGGAGGAGGTAGATGATGACAATAACAGCAATGTTGACATGGATGCTGAAATGATTGCTGAGGCCAAGTTACTGCGTCAACACA	1431

TCGTCTGGAGTCGCGTATGAAGATCCTAGAG

Scyliorhinus canicula (X99702.1)	CACCCTAAGATGACAGAGCTTTATCAGTCACTAGCCGACCTAAATAATGTAAGATTTTCTGCGTACAGAACAGCCATGAAGTTACGGAGATTGCAGAAAG CTCTTTGCTTGGATCTTCTCTCTCTACCAAGTGCTTGTGAAGCATTTGACCAGCACAATCTAAAGCAAAATGATCAACTAATGGATATTTTGGAGATAATTA ACTGCCTGACTTCAATCTATGATCGACTGGAACAGGAACACAGCAACCTCGTCAATGTGCCTCTCTGTGTGGACATGTGTCTCAACTGGCTTCTCAATGTC TATGACACTGGCCGAACGGGGAAGATTCGTGTTCGTGTCCTTTAAAAGTGGAATCATGTCACTGTGCAAAGCACATCTCGAGGACAAATATCGCTATCTGT TCAAGCAAGTGGCAAGTCCCACTGGATTCTGTGATCAGCGGAGGCTAGGGCTTCTGTTACATGATGCAATCCAGATTCCACGCCAGCTGGGAGAGGTTG CATCGTTTGGAGGCAGCAACATTGAACCAAGTGTACGAAGCTGCTTCCAGTTTGCCAACAACAAGCCGGAGGTTGAGGCAGCATTGTTCCTGGATTGGA TGAGGCTGGAGCCACAGTCTATGGTCTGGATGCCAGTGCTTCACCGAGTGGCAGCAGCAGAGACTGCCAAACACCAAGCTAAGTGCAATATCTGCAAG GAATGTCCTATCATCGGATTCAGGTACCGAAGTTTAAAGCACTTTAATTATGATGTCTGCCAAAGCTGCTTCTTTTCTGGTCGAACGGCAAAAGGACATAA AATGCATTACCCAATGGTGGAATACTGCACACCGACAACTTCAGGAGAAGATGTCCGGGATTTCGCCAAGGTATTAAAAAACAAGTTCCGAACAAAGAG ATACTTTGCAAAGCATCCCCGGATGGGATATCTTCCTGTGCAGACAGTTCTGGAGGGGGATAACTTAGAAACTCCTGTTACTCTGATCAACTTCTGGCCA GTAGATTATGAACCTGCCTCATCCCCTCAGTTGTCACACGATGATACTCATTCACGAATTGAACATTATGCTAGCAGGCTAGCAGAGATGGAAAACAGAA ATGGATCTTACTTGAATGACAGCATCTCACCCAATGAAAGCATAGATGATGAACATCTGCTGATTCAGCACTACTGTCAGAGTCTGAACCAGGAGTCACC ATTGAGTCAGCCTCGGAGCCCTGCTCARATTCTCATCTCATTGGAGAGTGAAGAAAGAGGAGAATTGGAAAGAATTCTAGCTGACCTGGAAGATGAAAA TAGAAATCTTCAGTCTGAGTATGAGAGGCTGAAGCAGCAGCATGACCACAAAAGGCCTGTCCCCCTTGCCATCCCCTCCTGAAATGATGCCCATTTCGCCG CAGAGTCCCAGGGATGCCGAGCTAATCGCGGAAGCCAAACTGCTCCGCCAGCACAAAGGTCGCCTTGAGGCNAGGATGCAGATCCTGGAAGATCACAA CAAGCAATTAGAATCCCAGCTGCACAGGCTGAGGCAGCTGCTGGAGCAGCCACAGACAGAGGTCAGGGTAAATGGTACTACAGTCTCCTCCCCTTCTAC TTCCTCTCAAAGATCAGACAGCAGTCAGCCAGTGCTACTCCATGGAGTCGGCAGTCAGACTTCAGGAACAATGGGTGAGGATGAATTGCTGAGTCCACC CCAGGGTACTACCACTGGATTGGAGGAAGTGATGGAGCAGCTCAACAATTCCTTTCCGAGCTCCCGACTGAGGGATTCCTATGGAAAGCAAGTGAAAGA GGTAACAATGTAGGCAGTCTCTTCCATATGGTGGACGACCTGGGCCGAGCAATGGAGACCTTGGTGACAATCATGACCGATGATGAGACAGCAGAGTA ACCAGATGTGCTTTATCACAACATTGGCTTCAGATGTCAAGCATGGTTTTTATATGTTCATACAAAAGATAAATGGATTGGACCGTACAGGTTTACNAGA AA	1993
Xenopus laevis (X99700.1)	CCTAAGATGACAGAATTATATCAATCCTTAGCGGACCTAAACAATGTGCGATTCTCAGCATACAGAACTGCCATGAAGCTAAGAAGACTACAAAAGGCCT TGTGTCTGGACTTGCTAGGGCTGTCTGCAGCTTGTGAAGCCTTGGACCAGTACAACCTTAAGCAGAATGACCAGCTGATGGACATCCTACAGATCATTAA TTGCTTGACCACAATTTACGATCGATTGGAGCAAGAGCACAATAACCTGGTGAACGTTCCTCTCTGCGTGGACATGTGCCTCAACTGGCTGCTGAATGTT TATGACACGGGTCGAACAGGACGTATACGTGTGTCTTATCTTTTAAAACTGGTGTCATTTCCCTGTGTAAAGCACATTTGGAAGATAAGTACAGATACCTATT CAAGCAGGTGGCAAGCTCCACGGGATACTGTGACCAGCGGAGACTGGGCCTGCTTTTACATGATGCAATTCAAATTCCCCGACAGCTGGGTGAAGTGGC CTCTTTTGGCGGCAGTAACATTGAACCCAGTGTCCGAAGCTGCTTTCAGTATGCCAATAACAAACCAGAAATTGAAGCAGCTCTGTTTCTGGATTGGATG AGGCTGGAACCGCAATCAATGGTGTGGCTGCCAGTGCTCCATAGAGTAGCTGCTGCTGAGACTGCCAAACACCAAGCCAAGTGCAACATCTGTAAGGA GTGTCCAATTATTGGATTCAGGTACCGAAGTTTRAAGCACTTTAACTATGACATCTGCCAAAGCTGCTTCTTTTCTGGCCGTGTTGCAAAAGGTCACAAAA TGCACTATCCAATGGTTGAATATTGCACTCCTACGACATCTGGAGAAGATGTTCGAGATTTTGCGAAAGTTTTGAAGAATAAATTCAGAACAAAAGATA CTTTGCTAAGCACCCAAGGATGGGATATTTGCCTGTTCAAACTGTTCTAGAAGGAGATAACATGGAAACTCCTGTTACTCTNATCAACTTCTGGCCAGTA GATTCTGCGCTNGCAGAAATGGAAAATACCAATGGATCCTACCTAAACGATAGCATTTCACCTAACGAAAGCATAGATGATGAACATTTGCTCATCCAGC ACTACTGCCAAAGCTTRAACCAAGAATCCCCTCTGAGTCAGCCTCGCAGCCCTGCTCAGATTCTTATTTCTTTGGAGAGTGAGGAACGAGGAGAGCTTGA GCGGATCCTAGCTGACCTGGAAGAAGAAAACAARAATTTGCARGCNGAATATGATAGACTGAAAGAGCAACATGATCACAAGGGTTTATCTCCTCTGCC AACACCACCAGAAATGATGCCCACCCTCCCCACAGAGCCAGCGGGATGCTGAGCTCATTGCTGAAGCCAAGCTGCTGCGCCAGCACAAGGGGCGCTTGG AAGCGAGGATGCAGATCTTAGAGGATCACAACAAACAACTAGAATCCCAGTTGCATCGACTGAGACAGCTGCTGGAGCAACCGCAAGCAGAGTCAAAG GTGAATGGTACAACATATTCTTCTCCTACTCCCTCCCTGCAGAGATCAGAGAGCAATCAGCCTATGCTGCTTCGTGTTGTTGGCAGTCAGACATCAGAAA GCATGGGTGAAGATGATCGTCTCCTCAGCCCTCATGACTCAAGCACAGGCCTAGAAGAAGTGATGGAACAGCTAAACAACTCCTTCCCAACTGCAAGAG GAAGAATTGTCCCTGGAAAGCAAATGAGAGAGGAACCAATGTAGGCAGCCTTTTCCACATGGTGGATGACTTGGGAAGAGCGATGGAAACCTTAGTTA CTGTTATGACTGATGACGAGGTGTTAGAATGACCGTTCCAATCACGACGCATGTGTTTTTATATATCACATTCGTACAAACAAAAGGGGATTAGACTGT AAGAGTTTACNAGAAAATATCTATTTTTNGNGAAGGGTAGTGGTACTATACTGTAGATTTCAGTAGTTTCTT	1961

REFERENCES

Allen, D. G., Gervasio, O. L., Yeung, E. W., & Whitehead, N. P. (2010). Calcium and the damage pathways in muscular dystrophy. *Canadian Journal of Physiology and Pharmacology, 88*(2), 83-91. doi: 10.1139/Y09-058

Anand, S. K., & Tikoo, S. K. (2013). Viruses as modulators of mitochondrial functions. *Advances in Virology, 2013*, 738794. doi: 10.1155/2013/738794

Arnold, K., Bordoli, L., Kopp, J., & Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics, 22*(2), 195-201. doi: 10.1093/bioinformatics/bti770

Arockiaraj, J., Easwvaran, S., Vanaraja, P., Singh, A., Othman, R. Y., & Bhassu, S. (2012). Immunological role of thiol-dependent peroxiredoxin gene in Macrobrachium rosenbergii. *Fish & Shellfish Immunology, 33*(1), 121-129. doi: 10.1016/j.fsi.2012.04.010

Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., . . . Stockinger, H. (2012). ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Research, 40*(Web Server issue), W597-603. doi: 10.1093/nar/gks400

Baker, D., & Sali, A. (2001). Protein structure prediction and structural genomics. *Science, 294*(5540), 93-96. doi: 10.1126/science.1065659

Bateman, K. S., Tew, I., French, C., Hicks, R. J., Martin, P., Munro, J., & Stentiford, G. D. (2012). Susceptibility to infection and pathogenicity of White Spot Disease (WSD) in non-model crustacean host taxa from temperate regions. *Journal of Invertebrate Pathology, 110*(3), 340-351. doi: 10.1016/j.jip.2012.03.022

Benkert, P., Biasini, M., & Schwede, T. (2011). Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics, 27*(3), 343-350. doi: 10.1093/bioinformatics/btq662

Benkert, P., Tosatto, S. C., & Schomburg, D. (2008). QMEAN: A comprehensive scoring function for model quality assessment. *Proteins, 71*(1), 261-277. doi: 10.1002/prot.21715

Berchtold, M. W., Brinkmeier, H., & Muntener, M. (2000). Calcium ion in skeletal muscle: its crucial role for muscle function, plasticity, and disease. *Physiological Reviews, 80*(3), 1215-1265.

Berg JM, T. J., Stryer L. . (2002). *Biochemistry* (W. H. Freeman Ed. 5th edition ed.). New York.

Bertorini, T. E., Cornelio, F., Bhattacharya, S. K., Palmieri, G. M., Dones, I., Dworzak, F., & Brambati, B. (1984). Calcium and magnesium content in fetuses at risk and prenecrotic Duchenne muscular dystrophy. *Neurology, 34*(11), 1436-1440.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., . . . Schwede, T. (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research, 42*(Web Server issue), W252-258. doi: 10.1093/nar/gku340

BIOVIA, D. S. (2015). Discovery Studio Modeling Environment (Version Release 4.5). San Diego.

Blake, D. J., Weir, A., Newey, S. E., & Davies, K. E. (2002). Function and genetics of dystrophin and dystrophin-related proteins in muscle. *Physiological Reviews, 82*(2), 291-329. doi: 10.1152/physrev.00028.2001

Bodensteiner, J. B., & Engel, A. G. (1978). Intracellular calcium accumulation in Duchenne dystrophy and other myopathies: a study of 567,000 muscle fibers in 114 biopsies. *Neurology, 28*(5), 439-446.

Bordoli, L., & Schwede, T. (2012). Automated protein structure modeling with SWISS-MODEL Workspace and the Protein Model Portal. *Methods in Molecular Biology, 857*, 107-136. doi: 10.1007/978-1-61779-588-6_5

Buchan, D. W., Minneci, F., Nugent, T. C., Bryson, K., & Jones, D. T. (2013). Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Research, 41*(Web Server issue), W349-357. doi: 10.1093/nar/gkt381

Campos, E. C., O'Connell, J. L., Malvestio, L. M., Romano, M. M., Ramos, S. G., Celes, M. R., . . . Rossi, M. A. (2011). Calpain-mediated dystrophin disruption may be a potential structural culprit behind chronic doxorubicin-induced cardiomyopathy. *European Journal of Pharmacology, 670*(2-3), 541-553. doi: 10.1016/j.ejphar.2011.09.021

Chai, Y. M., Yu, S. S., Zhao, X. F., Zhu, Q., & Wang, J. X. (2010). Comparative proteomic profiles of the hepatopancreas in Fenneropenaeus chinensis response to white spot syndrome virus. *Fish & Shellfish Immunology, 29*(3), 480-486. doi: 10.1016/j.fsi.2010.05.009

Chamberlain, J. S., & Benian, G. M. (2000). Muscular dystrophy: the worm turns to genetic disease. *Current Biology, 10*(21), R795-797.

Chandonia, J. M., & Brenner, S. E. (2006). The impact of structural genomics: expectations and outcomes. *Science, 311*(5759), 347-351. doi: 10.1126/science.1121018

Chen, W., Calvo, P. A., Malide, D., Gibbs, J., Schubert, U., Bacik, I., . . . Yewdell, J. W. (2001). A novel influenza A virus mitochondrial protein that induces cell death. *Nature Medicine, 7*(12), 1306-1312. doi: 10.1038/nm1201-1306

Chothia, C. (1992). Proteins. One thousand families for the molecular biologist. *Nature, 357*(6379), 543-544. doi: 10.1038/357543a0

Chung, W., & Campanelli, J. T. (1999). WW and EF hand domains of dystrophin-family proteins mediate dystroglycan binding. *Molecular Cell Biology Research Communications, 2*(3), 162-171. doi: 10.1006/mcbr.1999.0168

Claydon, K., Cullen, B., & Owens, L. (2004). OIE white spot syndrome virus PCR gives false-positive results in Cherax quadricarinatus. *Diseases of Aquatic Organisms, 62*(3), 265-268. doi: 10.3354/dao062265

Corteel, M., Dantas-Lima, J. J., Tuan, V. V., Thuong, K. V., Wille, M., Alday-Sanz, V., . . . Nauwynck, H. J. (2012). Susceptibility of juvenile Macrobrachium rosenbergii to different doses of high and low virulence strains of white spot syndrome virus (WSSV). *Diseases of Aquatic Organisms, 100*(3), 211-218. doi: 10.3354/dao02496

Cox, G. A., Cole, N. M., Matsumura, K., Phelps, S. F., Hauschka, S. D., Campbell, K. P., . . . Chamberlain, J. S. (1993). Overexpression of dystrophin in transgenic mdx mice eliminates dystrophic symptoms without toxicity. *Nature, 364*(6439), 725-729. doi: 10.1038/364725a0

Creighton, T. E. (1993). *Proteins: structures and molecular properties*: Macmillan.

Dahlmann, B., Kuehn, L., Reinauer, H., Kay, J., & Stauber, W. T. (1989). Muscle Protein Wasting in Diabetes mellitus: Role of Proteases. *73*, 127-138. doi: 10.1159/000417385

Dankert, J. R., Papadi, G. P., & Shields, R. P. (1992). Ultrastructural distribution of the M form of creatine phosphokinase in human muscle by immunogold labeling. *Microscopy Research and Technique, 20*(3), 281-287. doi: 10.1002/jemt.1070200308

De Aro, A. A., Guerra Fda, R., Esquisatto, M. A., Nakagaki, W. R., Gomes, L., & Pimentel, E. R. (2015). Biochemical and morphological alterations in the Achilles tendon of mdx mice. *Microscopy Research and Technique, 78*(1), 85-93. doi: 10.1002/jemt.22448

Defranchi, E., Bonaccurso, E., Tedesco, M., Canato, M., Pavan, E., Raiteri, R., & Reggiani, C. (2005). Imaging and elasticity measurements of the sarcolemma of fully differentiated skeletal muscle fibres. *Microscopy Research and Technique, 67*(1), 27-35. doi: 10.1002/jemt.20177

Delella, F. K., & Felisbino, S. L. (2010). Doxazosin treatment alters stromal cell behavior and increases elastic system fibers deposition in rat prostate. *Microscopy Research and Technique, 73*(11), 1036-1044. doi: 10.1002/jemt.20828

Dhar, A. K., Bowers, R. M., Licon, K. S., Veazey, G., & Read, B. (2009). Validation of reference genes for quantitative measurement of immune gene expression in shrimp. *Molecular Immunology, 46*(8-9), 1688-1695. doi: 10.1016/j.molimm.2009.02.020

Duan, Y., Liu, P., Li, J., Wang, Y., Li, J., & Chen, P. (2014). Molecular responses of calreticulin gene to Vibrio anguillarum and WSSV challenge in the ridgetail white prawn Exopalaemon carinicauda. *Fish & Shellfish Immunology, 36*(1), 164-171. doi: 10.1016/j.fsi.2013.10.024

Durand, S. V., & Lightner, D. V. (2002). Quantitative real time PCR for the measurement of white spot syndrome virus in shrimp. *Journal of Fish Diseases, 25*(7), 381-389. doi: 10.1046/j.1365-2761.2002.00367.x

Duret, L., Gasteiger, E., & Perriere, G. (1996). LALNVIEW: a graphical viewer for pairwise sequence alignments. *Computer Applications in the Biosciences, 12*(6), 507-510.

Emery, A. E. (2008). Resistance to infection and the muscular dystrophies--is there a molecular link? *Neuromuscular Disorders, 18*(5), 423-425. doi: 10.1016/j.nmd.2008.03.004

Escobedo-Bonilla, C. M., Alday-Sanz, V., Wille, M., Sorgeloos, P., Pensaert, M. B., & Nauwynck, H. J. (2008). A review on the morphology, molecular characterization, morphogenesis and pathogenesis of white spot syndrome virus. *Journal of Fish Diseases, 31*(1), 1-18. doi: 10.1111/j.1365-2761.2007.00877.x

FAO. (2009). FIGIS: Fisheries Global Information System. Retrieved 10 January, 2016

FAO. (2011). *Fishstat Plus (v, 2.32)*. Rome, Italy: FAO.

Finsterer, J., & Stollberger, C. (2003). The heart in human dystrophinopathies. *Cardiology, 99*(1), 1-19. doi: 68446

FL Garcia-Carreño, M. N. d. T., A Muhlia-Almazan. (2014). The role of lysosomal cysteine proteases in crustacean immune response. *Invertebrate Survival Journal, 11*, 10.

Fuchs, J. E., von Grafenstein, S., Huber, R. G., Margreiter, M. A., Spitzer, G. M., Wallnoefer, H. G., & Liedl, K. R. (2013). Cleavage entropy as quantitative measure

of protease specificity. *PLoS Computational Biology, 9*(4), e1003007. doi: 10.1371/journal.pcbi.1003007

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S. e., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. In J. M. Walker (Ed.), *The Proteomics Protocols Handbook* (pp. 571-607). Totowa, NJ: Humana Press.

Gieseler, K., Grisoni, K., & Segalat, L. (2000). Genetic suppression of phenotypes arising from mutations in dystrophin-related genes in Caenorhabditis elegans. *Current Biology, 10*(18), 1092-1097.

Goldstein, J. A., & McNally, E. M. (2010). Mechanisms of muscle weakness in muscular dystrophy. *Journal of General Physiology, 136*(1), 29-34. doi: 10.1085/jgp.201010436

Goonasekera, S. A., Davis, J., Kwong, J. Q., Accornero, F., Wei-LaPierre, L., Sargent, M. A., . . . Molkentin, J. D. (2014). Enhanced Ca(2)(+) influx from STIM1-Orai1 induces muscle pathology in mouse models of muscular dystrophy. *Human Molecular Genetics, 23*(14), 3706-3715. doi: 10.1093/hmg/ddu079

Gopalakrishnan, K., Sowmiya, G., Sheik, S. S., & Sekar, K. (2007). Ramachandran plot on the web (2.0). *Protein Pept Lett, 14*(7), 669-671.

Gordon, R. E. (2014). Electron Microscopy: A Brief History and Review of Current Clinical Application. In E. C. Day (Ed.), *Histopathology: Methods and Protocols* (pp. 119-135). New York, NY: Springer New York.

Gorman, M. J., Andreeva, O. V., & Paskewitz, S. M. (2000). Sp22D: a multidomain serine protease with a putative role in insect immunity. *Gene, 251*(1), 9-17.

Hnia, K., Zouiten, D., Cantel, S., Chazalette, D., Hugon, G., Fehrentz, J. A., . . . Winder, S. J. (2007). ZZ domain of dystrophin and utrophin: topology and mapping of a beta-dystroglycan interaction site. *Biochemical Journal, 401*(3), 667-677. doi: 10.1042/BJ20061051

Ho, B. K., Thomas, A., & Brasseur, R. (2003). Revisiting the Ramachandran plot: hard-sphere repulsion, electrostatics, and H-bonding in the alpha-helix. *Protein Science, 12*(11), 2508-2522. doi: 10.1110/ps.03235203

Honey, K., & Rudensky, A. Y. (2003). Lysosomal cysteine proteases regulate antigen presentation. *Nature Reviews: Immunology, 3*(6), 472-482. doi: 10.1038/nri1110

Huang, J., & Forsberg, N. E. (1998). Role of calpain in skeletal-muscle protein degradation. *Proceedings of the National Academy of Sciences of the United States of America, 95*(21), 12100-12105.

Huang, X., Poy, F., Zhang, R., Joachimiak, A., Sudol, M., & Eck, M. J. (2000). Structure of a WW domain containing fragment of dystrophin in complex with beta-dystroglycan. *Nature Structural Biology, 7*(8), 634-638. doi: 10.1038/77923

Ishikawa-Sakurai, M., Yoshida, M., Imamura, M., Davies, K. E., & Ozawa, E. (2004). ZZ domain is essentially required for the physiological binding of dystrophin and utrophin to beta-dystroglycan. *Human Molecular Genetics, 13*(7), 693-702. doi: 10.1093/hmg/ddh087

Jimenez-Vega, F., Vargas-Albores, F., & Soderhall, K. (2005). Characterisation of a serine proteinase from Penaeus vannamei haemocytes. *Fish & Shellfish Immunology, 18*(2), 101-108. doi: 10.1016/j.fsi.2004.02.001

Jin, H., Tan, S., Hermanowski, J., Bohm, S., Pacheco, S., McCauley, J. M., . . . Roberts, R. G. (2007). The dystrotelin, dystrophin and dystrobrevin superfamily: new paralogues and old isoforms. *BMC Genomics, 8*, 19. doi: 10.1186/1471-2164-8-19

Kalinowska, B., Alejster, P., Salapa, K., Baster, Z., & Roterman, I. (2013). Hypothetical in silico model of the early-stage intermediate in protein folding. *J Mol Model, 19*(10), 4259-4269. doi: 10.1007/s00894-013-1909-6

Kan, F. W., & Nanci, A. (1988). Backscattered electron imaging of lectin binding sites in tissues following freeze-fracture cytochemistry. *Journal of Electron Microscopy Technique, 8*(4), 363-370. doi: 10.1002/jemt.1060080405

Komuro, T. (1999). Comparative morphology of interstitial cells of Cajal: ultrastructural characterization. *Microscopy Research and Technique, 47*(4), 267-285. doi: 10.1002/(SICI)1097-0029(19991115)47:4<267::AID-JEMT5>3.0.CO;2-O

Koohmaraie, M. (1992). The role of Ca(2+)-dependent proteases (calpains) in post mortem proteolysis and meat tenderness. *Biochimie, 74*(3), 239-245.

Kou, G.-H., Peng, S.-E., Chiu, Y.-L., & Lo, C.-F. (1998). Tissue distribution of white spot syndrome virus (WSSV) in shrimp and crabs. *Advances in shrimp biotechnology*, 267-271.

Lessa, T. B., de Abreu, D. K., Rodrigues, M. N., Brolio, M. P., Miglino, M. A., & Ambrosio, C. E. (2014). Morphological and ultrastructural evaluation of the golden retriever muscular dystrophy trachea, lungs, and diaphragm muscle. *Microscopy Research and Technique, 77*(11), 857-861. doi: 10.1002/jemt.22408

Leu, J.-H., Chang, C.-C., Wu, J.-L., Hsu, C.-W., Hirono, I., Aoki, T., . . . Huang, H.-C. (2007). Comparative analysis of differentially expressed genes in normal and white spot syndrome virus infected Penaeus monodon. *BMC Genomics, 8*(1), 1-14. doi: 10.1186/1471-2164-8-120

Liu, Y. C., Li, F. H., Dong, B., Wang, B., Luan, W., Zhang, X. J., . . . Xiang, J. H. (2007). Molecular cloning, characterization and expression analysis of a putative C-type lectin (Fclectin) gene in Chinese shrimp Fenneropenaeus chinensis. *Molecular Immunology, 44*. doi: 10.1016/j.molimm.2006.01.015

Livak, K. J., & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods, 25*(4), 402-408. doi: 10.1006/meth.2001.1262

Lo, C.-F., S-E Peng, Y-S Chang and G-H Kou. (2005). White Spot Syndeome - What we have learned about the virus and the disease. In R. L. a. M. G. B.-R. P. Walker (Ed.), *Diseases in Asian Aquaculture* (pp. 13). Manila: Asian Fisheries Society.

Lo, C. F., Leu, J. H., Ho, C. H., Chen, C. H., Peng, S. E., Chen, Y. T., . . . Kou, G. H. (1996). Detection of baculovirus associated with white spot syndrome (WSBV) in penaeid shrimps using polymerase chain reaction. *Dis Aquat Org, 5*. doi: 10.3354/dao025133

Lovell, S. C., Davis, I. W., Arendall, W. B., 3rd, de Bakker, P. I., Word, J. M., Prisant, M. G., . . . Richardson, D. C. (2003). Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins, 50*(3), 437-450. doi: 10.1002/prot.10286

Ma, X. W., Li, Q., Xu, P. T., Zhang, L., Li, H., & Yu, Z. B. (2011). Tetanic contractions impair sarcomeric Z-disk of atrophic soleus muscle via calpain pathway. *Molecular and Cellular Biochemistry, 354*(1-2), 171-180. doi: 10.1007/s11010-011-0816-3

Mallouk, N., Jacquemond, V., & Allard, B. (2000). Elevated subsarcolemmal Ca2+ in mdx mouse skeletal muscle fibers detected with Ca2+-activated K+ channels.

Proceedings of the National Academy of Sciences, 97(9), 4950-4955. doi: 10.1073/pnas.97.9.4950

Mariol, M. C., & Segalat, L. (2001). Muscular degeneration in the absence of dystrophin is a calcium-dependent process. *Current Biology, 11*(21), 1691-1694. doi: Doi 10.1016/S0960-9822(01)00528-0

Mehler, M. F. (2000). Brain dystrophin, neurogenetics and mental retardation. *Brain Research: Brain Research Reviews, 32*(1), 277-307.

Mekata, T., Sudhakaran, R., Kono, T., Supamattaya, K., Linh, N. T., Sakai, M., & Itami, T. (2009). Real-time quantitative loop-mediated isothermal amplification as a simple method for detecting white spot syndrome virus. *Letters in Applied Microbiology, 48*(1), 25-32. doi: 10.1111/j.1472-765X.2008.02479.x

Mendell, J. R., Buzin, C. H., Feng, J., Yan, J., Serrano, C., Sangani, D. S., . . . Sommer, S. S. (2001). Diagnosis of Duchenne dystrophy by enhanced detection of small mutations. *Neurology, 57*(4), 645-650.

Mendoza-Cano, F., & Sanchez-Paz, A. (2013). Development and validation of a quantitative real-time polymerase chain assay for universal detection of the White Spot Syndrome Virus in marine crustaceans. *Virology Journal, 10*(186), 186. doi: 10.1186/1743-422X-10-186

Miyake, K., & McNeil, P. L. (2003). Mechanical injury and repair of cells. *Critical Care Medicine, 31*(8 Suppl), S496-501. doi: 10.1097/01.CCM.0000081432.72812.16

Muntoni, F., Torelli, S., & Ferlini, A. (2003). Dystrophin and mutations: one gene, several proteins, multiple phenotypes. *The Lancet Neurology, 2*(12), 731-740. doi: 10.1016/s1474-4422(03)00585-4

Muntoni, F., & Wells, D. (2007). Genetic treatments in muscular dystrophies. *Current Opinion in Neurology, 20*(5), 590-594. doi: 10.1097/WCO.0b013e3282efc157

Nathiga Nambi, K. S., Abdul Majeed, S., Sundar Raj, N., Taju, G., Madan, N., Vimal, S., & Sahul Hameed, A. S. (2012). In vitro white spot syndrome virus (WSSV) replication in explants of the heart of freshwater crab, Paratelphusa hydrodomous. *Journal of Virological Methods, 183*(2), 186-195. doi: 10.1016/j.jviromet.2012.04.013

Nayak, S., Ajay, K. M., Ramaiah, N., Meena, R. M., & Sreepada, R. A. (2011). Profiling of a few immune responsive genes expressed in postlarvae of Fenneropenaeus indicus challenged with Vibrio harveyi D3. *Journal of Invertebrate Pathology, 107*(2), 168-172. doi: 10.1016/j.jip.2011.04.001

New, M. (2002). Farming freshwater prawns: a manual for the culture of the giant river prawn (*Macrobrachium rosenbergii*). *FAO Fish Tech Pap 428*.

New, M. B., & Nair, C. M. (2012). Global scale of freshwater prawn farming. *Aquaculture Research, 43*(7), 960-969. doi: 10.1111/j.1365-2109.2011.03008.x

Nunan, L. M., & Lightner, D. V. (2011). Optimized PCR assay for detection of white spot syndrome virus (WSSV). *Journal of Virological Methods, 171*(1), 318-321. doi: 10.1016/j.jviromet.2010.11.015

Palade, G. E. (1952). A study of fixation for electron microscopy. *Journal of Experimental Medicine, 95*(3), 285-298.

Parkinson, G., Vojtechovsky, J., Clowney, L., Brunger, A. T., & Berman, H. M. (1996). New parameters for the refinement of nucleic acid-containing structures. *Acta Crystallographica. Section D: Biological Crystallography, 52*(Pt 1), 57-64. doi: 10.1107/S0907444995011115

Petrof, B. J., Shrager, J. B., Stedman, H. H., Kelly, A. M., & Sweeney, H. L. (1993). Dystrophin protects the sarcolemma from stresses developed during muscle contraction. *Proceedings of the National Academy of Sciences of the United States of America, 90*(8), 3710-3714.

Pirovano, W., & Heringa, J. (2010). Protein secondary structure prediction. *Methods in Molecular Biology, 609*, 327-348. doi: 10.1007/978-1-60327-241-4_19

Polican Ciena, A., Yokomizo De Almeida, S. R., De Sousa Bolina, C., De Sousa Bolina-Matos, R., Grassi Rici, R. E., Pereira Da Silva, M. C., . . . Watanabe, I. S. (2012). Ultrastructural features of the myotendinous junction of the sternomastoid muscle in Wistar rats: from newborn to aging. *Microscopy Research and Technique, 75*(9), 1292-1296. doi: 10.1002/jemt.22063

Pradeep, B., Rai, P., Mohan, S. A., Shekhar, M. S., & Karunasagar, I. (2012). Biology, Host Range, Pathogenesis and Diagnosis of White spot syndrome virus. *Indian Journal of Virology, 23*(2), 161-174. doi: 10.1007/s13337-012-0079-y

Prior, T. W., & Bridgeman, S. J. (2005). Experience and strategy for the molecular testing of Duchenne muscular dystrophy. *J Mol Diagn, 7*(3), 317-326. doi: 10.1016/S1525-1578(10)60560-0

R B Pramod Kiran, K. V. R., S J Jung, M J Oh. (2002). Experimental susceptibility of different life-stages of the giant freshwater prawn, Macrobrachium rosenbergii (de Man), to white spot syndrome virus (WSSV). *Journal of Fish Diseases*(25).

Rafael, J. A., & Brown, S. C. (2000). Dystrophin and utrophin: genetic analyses of their role in skeletal muscle. *Microscopy Research and Technique, 48*(3-4), 155-166. doi: 10.1002/(sici)1097-0029(20000201/15)48:3/4<155::aid-jemt4>3.0.co;2-0

Rahman, M. M., Escobedo-Bonilla, C. M., Corteel, M., Dantas-Lima, J. J., Wille, M., Sanz, V. A., . . . Nauwynck, H. J. (2006). Effect of high water temperature (33 °C) on the clinical and virological outcome of experimental infections with white spot syndrome virus (WSSV) in specific pathogen-free (SPF) Litopenaeus vannamei. *Aquaculture (Amsterdam, Netherlands), 261*(3), 842-849. doi: http://dx.doi.org/10.1016/j.aquaculture.2006.09.007

Ramachandran, G. N., Ramakrishnan, C., & Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology, 7*, 95-99.

Rao, R., Bhassu, S., Bing, R. Z., Alinejad, T., Hassan, S. S., & Wang, J. (2016). A Transcriptome Study on Macrobrachium rosenbergii Hepatopancreas Experimentally Challenged with White Spot Syndrome Virus (WSSV). *Journal of Invertebrate Pathology*. doi: 10.1016/j.jip.2016.01.002

Rentschler, S., Linn, H., Deininger, K., Bedford, M. T., Espanel, X., & Sudol, M. (1999). The WW domain of dystrophin requires EF-hands region to interact with beta-dystroglycan. *Biological Chemistry, 380*(4), 431-442. doi: 10.1515/BC.1999.057

Richardson, J., Keedy, D., Richardson, D., Bansal, M., & Srinivasan, N. (2013). Biomolecular Forms and Functions: A Celebration of 50 Years of the Ramachandran Map.

Richardson, J. S., & David, C. (2012). MORE DATA, MORE DIMENSIONS, MORE USES. *Biomolecular Forms and Functions: A Celebration of 50 Years of the Ramachandran Map*, 46.

Roberts, R. G., & Bobrow, M. (1998). Dystrophins in vertebrates and invertebrates. *Human Molecular Genetics, 7*(4), 589-595. doi: Doi 10.1093/Hmg/7.4.589

Roterman, I., Konieczny, L., Banach, M., & Jurkowski, W. (2011). Intermediates in the protein folding process: a computational model. *Int J Mol Sci, 12*(8), 4850-4860. doi: 10.3390/ijms11084850

Royuela, M., Paniagua, R., Rivier, F., Hugon, G., Robert, A., & Mornet, D. (1999). Presence of invertebrate dystrophin-like products in obliquely striated muscle of the leech, Pontobdella muricata (Annelida, Hirudinea). *Histochemical Journal, 31*(9), 603-608.

Ruegg, U. T., & Gillis, J.-M. (1999). Calcium homeostasis in dystrophic muscle. *Trends in Pharmacological Sciences, 20*(9), 351-352. doi: http://dx.doi.org/10.1016/S0165-6147(99)01377-2

Rybakova, I. N., Patel, J. R., & Ervasti, J. M. (2000). The Dystrophin Complex Forms a Mechanically Strong Link between the Sarcolemma and Costameric Actin. *The Journal of Cell Biology, 150*(5), 1209-1214. doi: 10.1083/jcb.150.5.1209

Scothern, A., & Garrod, D. (2008). Visualization of desmosomes in the electron microscope. *Methods in Cell Biology, 88*, 347-366. doi: 10.1016/S0091-679X(08)00418-4

Segalat, L. (2002). Dystrophin and functionally related proteins in the nematode Caenorhabditis elegans. *Neuromuscular Disorders, 12 Suppl 1*, S105-109.

Segalat, L. (2007). Invertebrate animal models of diseases as screening tools in drug discovery. *ACS Chemical Biology, 2*(4), 231-236. doi: 10.1021/cb700009m

Shekhar, M. S., & Ponniah, A. G. (2015). Recent insights into host-pathogen interaction in white spot syndrome virus infected penaeid shrimp. *Journal of Fish Diseases, 38*(7), 599-612. doi: 10.1111/jfd.12279

Shi, X. Z., Ren, Q., Zhao, X. F., & Wang, J. X. (2009). Expression of four trypsin-like serine proteases from the Chinese shrimp, Fenneropenaeus chinensis, as regulated by pathogenic infection. *Comparative Biochemistry and Physiology. Part B: Biochemistry and Molecular Biology, 153*(1), 54-60. doi: 10.1016/j.cbpb.2009.01.011

Shin, J., Tajrishi, M. M., Ogura, Y., & Kumar, A. (2013). Wasting mechanisms in muscular dystrophy. *International Journal of Biochemistry and Cell Biology, 45*(10), 2266-2279. doi: 10.1016/j.biocel.2013.05.001

Sorimachi, H., Hata, S., & Ono, Y. (2011). Calpain chronicle--an enzyme family under multidisciplinary characterization. *Proceedings of the Japan Academy. Series B: Physical and Biological Sciences, 87*(6), 287-327.

Sparrow, J., Hughes, S. M., & Segalat, L. (2008). Other Model Organisms for Sarcomeric Muscle Diseases. *Sarcomere and Skeletal Muscle Disease, 642*, 192-206.

Sritunyalucksana, K., Jiraporn, S., McColl, K., Nielsen, L., Flegel, T.W. (2006). Comparison of PCR testing methods for white spot syndrome virus (WSSV) infections in *penaeid* shrimp. *Aquaculture (Amsterdam, Netherlands), 255*.

Stentiford, G. D., Bonami, J. R., & Alday-Sanz, V. (2009). A critical review of susceptibility of crustaceans to Taura syndrome, Yellowhead disease and White Spot Disease and implications of inclusion of these diseases in European legislation. *Aquaculture (Amsterdam, Netherlands), 291*(1–2), 1-17. doi: http://dx.doi.org/10.1016/j.aquaculture.2009.02.042

Supungul, P., Klinbunga, S., Pichyangkura, R., Jitrapakdee, S., Hirono, I., Aoki, T., & Tassanakajon, A. (2002). Identification of immune-related genes in hemocytes of

black tiger shrimp (Penaeus monodon). *Marine Biotechnology (New York, N.Y.), 4*(5), 487-494. doi: 10.1007/s10126-002-0043-8

Takahashi, A., Camacho, P., Lechleiter, J. D., & Herman, B. (1999). Measurement of intracellular calcium. *Physiological Reviews, 79*(4), 1089-1125.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution, 30*(12), 2725-2729. doi: 10.1093/molbev/mst197

van der Plas, M. C., Pilgram, G. S., de Jong, A. W., Bansraj, M. R., Fradkin, L. G., & Noordermeer, J. N. (2007). Drosophila Dystrophin is required for integrity of the musculature. *Mechanisms of Development, 124*(7-8), 617-630. doi: 10.1016/j.mod.2007.04.003

Vaseeharan, B., Lin, Y. C., Ko, C. F., & Chen, J. C. (2006). Cloning and characterisation of a serine proteinase from the haemocytes of mud crab Scylla serrata. *Fish & Shellfish Immunology, 21*(1), 20-31. doi: 10.1016/j.fsi.2005.09.006

Verburg, E., Dutka, T. L., & Lamb, G. D. (2006). Long-lasting muscle fatigue: partial disruption of excitation-contraction coupling by elevated cytosolic Ca2+ concentration during contractions. *American Journal of Physiology: Cell Physiology, 290*(4), C1199-1208. doi: 10.1152/ajpcell.00469.2005

Vijayan, K. K., Stalin Raj, V., Balasubramanian, C. P., Alavandi, S. V., Thillai Sekhar, V., & Santiago, T. C. (2005). Polychaete worms--a vector for white spot syndrome virus (WSSV). *Diseases of Aquatic Organisms, 63*(2-3), 107-111. doi: 10.3354/dao063107

Vriend, G. (1990). WHAT IF: a molecular modeling and drug design program. *Journal of Molecular Graphics, 8*(1), 52-56, 29.

Vulin, A., Wein, N., Strandjord, D. M., Johnson, E. K., Findlay, A. R., Maiti, B., . . . Flanigan, K. M. (2014). The ZZ domain of dystrophin in DMD: making sense of missense mutations. *Human Mutation, 35*(2), 257-264. doi: 10.1002/humu.22479

Wang, J., Pansky, A., Venuti, J. M., Yaffe, D., & Nudel, U. (1998). A sea urchin gene encoding dystrophin-related proteins. *Human Molecular Genetics, 7*(4), 581-588. doi: Doi 10.1093/Hmg/7.4.581

Welsh, R. M., Selin, L. K., & Szomolanyi-Tsuda, E. (2004). Immunological memory to viral infections. *Annual Review of Immunology, 22*, 711-743. doi: 10.1146/annurev.immunol.22.012703.104527

White, S. J., & den Dunnen, J. T. (2006). Copy number variation in the genome; the human DMD gene as an example. *Cytogenet Genome Res, 115*(3-4), 240-246. doi: 10.1159/000095920

Wilson, S. C., Hooft RWW, Vriend G (1998). Who checks the checkers? Four validation tools applied to eight atomic resolution structures. EU 3-D Validation Network. *Journal of Molecular Biology, 276*(2), 417-436.

Xue, S., Yang, W., & Sun, J. (2013). Role of chymotrypsin-like serine proteinase in white spot syndrome virus infection in Fenneropenaeus chinensis. *Fish & Shellfish Immunology, 34*(2), 403-409. doi: http://dx.doi.org/10.1016/j.fsi.2012.10.017

Yeung, E. W., Whitehead, N. P., Suchyna, T. M., Gottlieb, P. A., Sachs, F., & Allen, D. G. (2005). Effects of stretch-activated channel blockers on [Ca2+]i and muscle damage in the mdx mouse. *Journal of Physiology, 562*(Pt 2), 367-380. doi: 10.1113/jphysiol.2004.075275

Yuan, F. H., Chen, Y. G., Zhang, Z. Z., Yue, H. T., Bi, H. T., Yuan, K., . . . Chen, Y. H. (2016). Down-regulation apoptosis signal-regulating kinase 1 gene reduced the Litopenaeus vannamei hemocyte apoptosis in WSSV infection. *Fish & Shellfish Immunology, 50*, 109-116. doi: 10.1016/j.fsi.2015.12.003

Yuan, L., Zhang, X., Chang, M., Jia, C., Hemmingsen, S. M., & Dai, H. (2007). A new fluorescent quantitative PCR-based in vitro neutralization assay for white spot syndrome virus. *Journal of Virological Methods, 146*(1-2), 96-103. doi: 10.1016/j.jviromet.2007.06.009

Zeng, Y., & Lu, C. P. (2009). Identification of differentially expressed genes in haemocytes of the crayfish (Procambarus clarkii) infected with white spot syndrome virus by suppression subtractive hybridization and cDNA microarrays. *Fish & Shellfish Immunology, 26*(4), 646-650. doi: 10.1016/j.fsi.2008.11.005

Zhang, Y. (2008). Progress and challenges in protein structure prediction. *Current Opinion in Structural Biology, 18*(3), 342-348. doi: 10.1016/j.sbi.2008.02.004

Zhou, A. Q., O'Hern, C. S., & Regan, L. (2011). Revisiting the Ramachandran plot from a new angle. *Protein Science, 20*(7), 1166-1171. doi: 10.1002/pro.644