Transcriptomic approaches to identify genes and markers for production traits in giant freshwater prawn, (*Macrobrachium rosenbergii*)

Hyungtaek Jung

Earth, Environmental and Biological Sciences

Science and Engineering Faculty

Queensland University of Technology

Brisbane, Australia

A thesis submitted in fulfilment of requirements for the degree of

Doctor of Philosophy

2013

Keywords

Giant freshwater prawn, GFP, *Macrobrachium rosenbergii*, transcriptomes, next generation sequencing, 454 pyrosequencing, ABI re-resequencing, genetic effects, growth-related candidate genes, muscle development genes, muscle degradation genes, growth traits, single nucleotide polymorphisms, SEQUENOM, gene-gene interactions, genetic associations, marker-assisted selection, genomic selection.

Statement of Original Authorship

This work has not previously been submitted for a degree or diploma at any other educational institution. To the best of my knowledge, this thesis contains no material from any other source, except where due reference is made.

Hyungtaek Jung

14 June, 2013

Works Published or Submitted for Publication by the Author Incorporated into the Thesis:

Statement of Contribution to Jointly Authored Works in the Thesis

1. **Jung H**, Lyons RE, Hurwood DA, Mather PB (2013) Genes and growth performance in crustacean species: A review of relevant genomic studies in crustaceans and other taxa. *Reviews in Aquaculture* **Online Version**.

This manuscript is incorporated as Chapter 2 of this thesis. The first author was responsible for the research, analysis and interpretation of data, and written work in this manuscript. The co-authors provided conceptual, logistical and editorial support.

2. **Jung H**, Lyons RE, Dinh H, Hurwood DA, McWilliam S, Mather PB (2011) Transcriptomics of a giant freshwater prawn (*Macrobrachium rosenbergii*): *de novo* assembly, annotation and marker discovery. *PLoS One* **6**: e27938.

This manuscript is incorporated as Chapter 3 of this thesis. The first author was responsible for the research, analysis and interpretation of data, and written work in this manuscript. The co-authors provided conceptual, logistical and editorial support.

3. **Jung H**, Lyons RE, Li Y, Thanh NM, Dinh H, Hurwood DA, Mather PB (2013) A candidate gene association study for growth performance in an improved giant freshwater prawn (*Macrobrachium rosenbergii*) culture line. *Accepted in Marine Biotechnology*.

This manuscript is incorporated as Chapter 4 of this thesis. The first author was responsible for the research, analysis and interpretation of data, and written work in this manuscript. The co-authors provided conceptual, logistical and editorial support.

Acknowledgements

This thesis is dedicated to my father, Wonchoon Jung, who has always supported me in heaven, and my mother, Soonae Lim, who has always showed endless love and support. I love you.

First, I would like to thank to my supervisors, Prof. Peter Mather, Dr. Russell Lyons, and Dr. David Hurwood, for all their help and support over the years. Thanks to their excellent supervision, advice, guidance and suggestions throughout the study program, I could finally have this blessing moment. In particular, thanks Peter for letting me have a superb project as a PhD student, thanks Russell for a fantastic project concept and for allowing me to do all RNA works in Commonwealth Scientific and Industrial Research Organisation (CSIRO), and thanks to David for continuous comments and friendly attitude.

Second, I would also like to thank Dr. Woo-Jin Kim at National Fisheries Research and Development Institute in South Korea and Dr. Peter Prentis in QUT for their constructive comments on this work. In particular, thanks to Woo-Jin, for his excellent comments and a wide range of information related to applications of next generation sequencing and single nucleotide polymorphism (SNP).

Third, I would like to acknowledge Australian Genome Research Facility (AGRF) for 454 Roche sequencing and SEQUENOM genotyping. I also thank CSIRO Brisbane for letting me do all RNA works in their laboratories. In particular, Mrs. Leanne Dierens and Mr. Barney Hines for assisting with RNA isolation, Dr. Sean McWilliam for assisting in bioinformatics analyses, and Dr. Yutao Li and Dr. Nguyen Minh Thanh (Intentional University in Vietnam) for assisting with SNP association studies. Thanks to all in the QUT Molecular Genetics Research Facility (MGRF) for assistance. In particular, thanks to Mr. Vincent Chand for his support as a friend and as a lab manager, and thanks to Mr. Hung Dinh for collecting and transporting samples in Vietnam. I would also like to thank my friends for their support and understanding. In particular, thanks to Hanbada members for their support over the years. During all the years, they showed belief that I can complete this journey.

Fourth, I would like to express my gratitude to an Australian government based on an International Postgraduate Research Scholarship program and a QUT based Vice Chancellor

Initiative Scholarship program for providing a postgraduate scholarship to me to conduct research in Australia. Additional funding for this work was provided by a QUT grant awarded to Peter Mather (QUT VC NPGS award). I am also grateful to the World Aquaculture Society who judged my oral presentation as second best at the AQUA2012 meeting in Praque, Czech Republic. Extra thanks for faith of Peter Mather in supporting me to represent the group at this prestigious meeting.

Finally, I would like to thank my family in Australia (Su-Jin, Dong-Hyun and Dong-Ju) and Korea (Jaetaek, Mikyung, Hyejin and nephews). They have showed love, support and understanding over the years. Without their unconditional love and patient support during this journey, I would have not reached this monumental moment. Thanks to their valuable sacrifices, this thesis could have been completed.

Abstract

Global aquaculture is expanding rapidly to address increasing demand for aquatic protein needs and an uncertain future for wild fisheries. To date most farmed aquatic stocks are essentially wild and little is known about their genomes or the genes that affect important economic traits in culture. Biologists have recognized that recent technological advances including next generation sequencing technologies (NGST) have opened up the possibility for generating genome wide sequence data sets rapidly for non-model organisms at a reasonable cost. Giant freshwater prawn (*Macrobrachium rosenbergii* or GFP), is currently the most economically important farmed freshwater crustacean species. Many factors can influence individual growth rate in target species but of particular importance in agriculture and aquaculture will be identification and characterization of the specific gene loci that contribute important phenotypic variation to growth because the information can be applied to speed up genetic improvement programs and increase productivity via marker-assisted selection (MAS) and genomic selection or genome-wide selection (GS). While currently there is only limited genomic information available for any crustacean species including GFP, a number of putative candidate genes have been identified or implicated in growth and muscle development in other crustacean species. In general, four approaches are suggested here in an effort to stimulate increased research on identification of growth-related genes in crustacean species. Survey the available information on; 1) associations between genes and growth reported in crustaceans, 2) identify growth-related genes involved with moulting, 3) identify muscle development and degradation genes involved in moulting, and 4) test for correlations between DNA sequences that have confirmed growth trait effects in farmed terrestrial and aquatic animal species and related sequences in other crustacean species (e.g. penaeid shrimps). This information in concert can provide a foundation for increasing the rate at which knowledge about key genes affecting growth traits in crustacean species is gained.

As an initial step, 454 pyrosequencing of cDNA was employed to characterise the transcriptome from GFP and to identify genes related to growth. A collection of 787,731 sequence reads (244.37 Mb) obtained from 454 pyrosequencing analysis of cDNA prepared from muscle, ovary and testis tissues were collected from 18 adult prawns and assembled into 123,534 expressed sequence tags (ESTs). Of these, 46% (8,411) of contigs and 19% (115,123) of singletons respectively possessed high similarity to sequences in the GenBank

non-redundant database, with most significant (E value < $1e^{-5}$) contig (80%) and singleton (84%) matches occurring with crustacean and insect sequences. KEGG analysis of contig open reading frames and InterproScan domains identified several biological pathways and domains with potentially important roles in growth and muscle development. Transcripts derived from genes including actin, myosin heavy and light chain, tropomyosin and troponin with fundamental roles in muscle development and muscle construction were abundant. Among the contigs, 834 single nucleotide polymorphisms (SNPs), 1198 indels and 658 simple sequence repeats (SSRs) motifs were also identified. The *M. rosenbergii* transcriptome data reported provides an invaluable resource for improving our understanding of this species' genome structure and biology.

In order to better understand the genetic basis of growth traits influenced strongly by multi-gene interactions, a large number of candidate gene markers are required. Since SNPs are abundant across the genome and are anchored in virtually all genomic regions including coding sequences, a candidate gene approach (CGA) using SNP markers can provide an effective means for detecting genes and regions that underlie phenotypic variation in adaptively significant traits. A total of 342 SNPs were identified in GFP ESTs after re-sequencing 13,834 bp from 47 EST loci. This equates to a remarkably high SNP density, of one SNP per 40 bp. Re-sequencing confirmed 42 putative SNPs from a preliminary set of 61 SNPs identified using 454 pyrosequencing. A total of 28 SNPs in 23 growth-related candidate genes, identified from ABI and 454 sequencing, were re-genotyped using SEQUENOM Mass Array to assess correlations between SNP markers in growth-related candidate genes with individual growth performance in an improved GFP culture line. A number of exonic and intronic SNPs in several candidate genes had a significant (Ankyrin-like repeats protein, Glycogen phosphorylase, Heat shock protein90, Peroxidasin, Rolling pebbles, Transforming growth factor β-induced precursor and Utp-glucose1-phosphate uridylyltransferase 2) ($P < 0.05$) and/or marginal (Glyceraldehyde-3-phosphate dehydrogenase, Inhibitor of growth and Non-histone chromosomal protein) ($0.05 < P < 0.07$) association with EBVs showing measurable genetic variation that ranged from 0.026 ~ 0.048 under a dominance and/or an additive model. A number of loci and SNP marker pairings also showed significant differences in SNP allele distributions, linkage disequilibrium (LD), and violation of Hardy-Weinberg Equilibrium (HWE) in comparisons between fast and slow growth performance groups. This is the largest set of candidate gene SNP markers (CGSMs) developed for *M. rosenbergii* to date and will be useful in future genetic analyses including for dissection of

complex traits, for tracing genetic relationships in breeding programs, and for monitoring genetic diversity in commercial and wild populations of *M. rosenbergii*. In particular, some unique SNPs identified here, may also be useful in other *Macrobrachium* species and provide nuclear markers for systematic studies and for detection of interspecific hybridization among congeners. Confirmation of CGSMs identified here in different culture lines could expedite potential application of these markers in MAS/GS of GFP breeding programs in the future.

Keywords: next generation sequencing, giant freshwater prawn, *Macrobrachium rosenbergii*, 454 pyrosequencing, transcriptome, genetic effects, growth-related genes, candidate gene SNP markers, correlations, growth traits, marker-assisted selection, genomic selection.

Note on Thesis Preparation

Chapters 2 to 4 of this thesis are presented as either published papers or as accepted manuscripts for publication. As such, there will be some necessary repetition of information in the General Introduction and General Discussion and Conclusions when compared against the Introductions and Discussions of individual data chapters. Chapters 2 to 4, Figures and Tables in each chapter are re-initialized to maintain each chapter as an independent research paper. All chapters also include their respective reference lists.

Table of Contents

CHAPTER 1: General Introduction and Thesis Structure

Many aquatic species are considered natural resources to be exploited for human consumption and profit. If current rates of wild stock depletion continue, however, the future of marine and freshwater wild fisheries will be increasingly uncertain (Worm *et al*. 2006). Over the past century, applications of genetics in aquaculture, (in particular to hatchery production and breed development programs) have assisted some industries to become intensive, highly productive systems, efficient for direct production of food. This has caused a consequent paradigm shift that has changed our perceptions of aquatic environments and the species therein (Hauser & Carvalho 2008). Without doubt, as human populations continue to increase, human dependence on farmed aquatic food resources will also increase and improved culture stocks will play an increasingly important role in world food security and in meeting increasing demand for food in the future (Hulata 2001; Hauser & Seeb 2008; Hutchings & Fraser 2008; Gjedrem *et al*. 2012; Lind *et al*. 2012).

Decapod crustaceans, in particular, marine species including prawns, lobsters and crabs contribute significantly to modern global aquaculture production with their combined value totalling US$ 70 billion in 2011 produced from over 200 cultured species (FAO 2012). While modern genetic technologies have been applied to several farmed marine penaeid prawn species to increase genetic gains, to date only a few species have benefited from this approach (Li *et al*. 2006; O'Leary *et al*. 2006; Tassnakajon *et al*. 2006; Ibarra *et al*. 2007; Lyons *et al*. 2007; Zhang *et al*. 2007; Hamasaki & Kitada 2008; Staelens *et al*. 2008; Wu *et al*. 2009). According to recent reports, giant freshwater prawn (*Macrobrachium rosenbergii* or GFP), is now considered to be one of the most important farmed crustacean species contributing to the global prawn aquaculture industry and great interest is now focused on improving the productivity of this important culture species (Nhan *et al*. 2009; Thanh *et al*. 2009; Thanh *et al*. 2010a, b; Aflalo *et al*. 2012; Kitcharoen *et al*. 2012; Luan *et al*. 2012). Recent statistical analyses indicate that average world annual GFP production has surpassed 500,000 tonnes annually with a value of US$ 2.5 billion and that the culture industry for *M. rosenbergii* now exceeds US$ 1.4 billion per year in Asia alone (Phuong *et al*. 2006; Chareontawee *et al*. 2007; FAO 2012). As a consequence, there is growing interest in GFP culture, particularly in Asia (Schwantes *et al*. 2009; Nhan *et al*. 2010). To increase

productivity of farmed GFP stocks there is a need to better understand this species basic biology, ecology and production traits to allow development of more productive culture strains for the expanding global industry.

Since the 1980s, *M. rosenbergii* has been disseminated widely around the world for commercial farming (New 2000, 2005) and GFP is now used for both research and commercial culture purposes (Thanh *et al*. 2009; Thanh *et al*. 2010a, b; Aflalo *et al*. 2012; Kitcharoen *et al*. 2012; Luan *et al*. 2012). Potential expansion of culture industries has, however, been limited by the relative low productivity of farmed strains compounded by the limited genetic information available for the species. After many decades of farming essentially unimproved stocks, many GFP culture industries have reported low and declining culture stock productivity (Chareontawee *et al*. 2007; Nhan *et al*. 2009; Thanh *et al*. 2009; Thanh *et al*. 2010a, b; Dinh *et al*. 2012; Dinh *et al*. 2013). This is likely the result of a decline in natural levels of genetic diversity in many GFP culture lines where in some places there is a shortage of wild spawners (Wilder *et al*. 1999) and an associated increase in inbreeding rates that results from exposure to repeated genetic bottlenecks that have eroded genetic diversity in many GFP stocks as they were developed for aquaculture. Most broodstock used in culture to date, have been developed from the 'western' form of GFP (de Bruyn *et al*. 2004a), a strain collected originally from the wild in Malaysia and then translocated subsequently to Hawaii for domestication (Thanh *et al*. 2009). The history of domestication and serial translocation among sites likely produced culture stocks with low levels of genetic variation and high inbreeding rates, characteristics that are not optimal for culture. While crossbreeding can be a relatively simple and inexpensive approach for improving the productivity of farmed stocks (Fjalestad 2005), selective breeding programs are likely to provide greater economic benefits over the longer term (Ponzoni *et al*. 2007; Thanh *et al*. 2009). While a number of stock improvement programs have been initiated recently for GFP around the world (Charoenatawee *et al*. 2007; Divu *et al*. 2008; Min *et al*. 2009; Nhan *et al*. 2009; Thanh *et al*. 2009; Thanh *et al*. 2010a, b; Aflalo *et al*. 2012; Dinh *et al*. 2012; Kitcharoen *et al*. 2012; Luan *et al*. 2012), to date, little attention has been paid to applying some of the new genomic technologies that have become available over the last five years to developing better GFP culture lines.

A lack of basic information about the genome of any target species can be a major obstacle when developing improved culture lines for production industries. This is because identifying genes that affect phenotypic variation in important production traits can be a very

difficult and challenging task where only limited public DNA sequence information is available for the target species. In 2010, when this project was initiated, onlypartial genetic information from mitochondrial (Miller *et al*. 2005) and nuclear DNA sequences including microsatellite markers (Chand *et al*. 2005; Charoentawee *et al*. 2006; Bhassu *et al*. 2008; Divu *et al*. 2008; Min *et al*. 2009) had been investigated in GFP to address phylogeographic questions (de Bruyn *et al*. 2004a, b; de Bruyn *et al*. 2005; de Bruyn & Mather 2007), gene characterization related to immune defence (Hsieh *et al*. 2006; Baruah *et al*. 2009; Cam *et al*. 2009; Sung *et al*. 2009) and for determining sexual maturation (Ngernsoungnern *et al*. 2009) focusing on two geographical regions; Asia and northern Australia. These studies in concert have provided, however, only limited genomic data. A more comprehensive genomic dataset, therefore, is required for GFP that can be applied to stock improvement programs.

Application of the emerging science of genomics in prawn species used in culture is a new development designed to elucidate the genetic basis of economically important traits or to characterise important phenotypic and genotypic variation (Li *et al*. 2006; O'Leary *et al*. 2006; Tassanakajon *et al*. 2006; Lyons *et al*. 2007; Staelens *et al*. 2008; Robalino *et al*. 2009; Wu *et al*. 2009; Du *et al*. 2010). When the current project commenced, genomic studies of GFP have not been conducted anywhere. In theory, a genomics approach can provide new tools for discovery and characterisation of novel alleles and genes. In addition, marker-assisted selection (MAS) and genomic selection or genome-wide selection (GS) developed via transcriptomic approaches can be a cost-effective way of providing a large number of genetic markers for identification and development of economically important quantitative trait loci (QTL) in target organisms. In particular, advances in GS can enhance the precision and efficiency of breeding programs. Therefore, understanding the genetic/genomic basis of variation in phenotypic traits can enhance effective breeding programs and this approach has been accepted as a powerful tool for estimation of breeding values (Morell *et al*. 2012; Hayes *et al*. 2013; Taylor 2013). While applications of genomic-assisted breeding were only in the early stages development in 2010, transcriptomic selection approaches using expressed sequence tag (EST) data and genetic linkage maps had been initiated for a small number of aquatic species, notably some salmonid species (Moen *et al*. 2008; McClelland & Naish 2008; Rexroad *et al*. 2008), common carp (Sun & Liang 2004), tilapia (Lee *et al*. 2005), channel catfish (Quiniou *et al*. 2007) and certain penaeid shrimp species (Li *et al*. 2003; Li *et al*. 2006; Maneeruttanarungroj *et al*. 2006; Zhang *et al*. 2007; Staelens *et al*. 2008; Du *et al*. 2010). A fine genetic map, however, requires a large number of molecular markers with

markers spread widely across the genome and is an essential prerequisite for mapping of QTLs to complement GS, which is a form of MAS, and for comparative genome mapping (Liu & Cordes 2004; Goddard & Hayes 2009). Compared with more traditional linkage maps based on random amplified polymorphic DNA (RAPD), restriction fragment length polymorphisms (RFLP), amplified fragment length polymorphisms (AFLP) or microsatellite markers, EST-simple sequence repeat (SSR) and single nucleotide polymorphism (SNP) markers provide a novel approach for studying variation in important traits in target organisms (Lander & Botstein 1989; Maneeruttanarungroj *et al*. 2006; Kucuktas *et al*. 2009; Du *et al*. 2010). While SSR and SNP markers are comparatively expensive to develop, they are very reliable and highly informative as they provide co-dominant markers that can often be transferred easily to new crosses or new populations (Bouck & Vision 2007; Moen *et al*. 2008; Gorbach *et al*. 2009; Du *et al*. 2010). Recent advances in livestock genomics and associated breeding practices have provided new opportunities, and the advent of genomic/transcriptomic selection promises a bright future for improving important traits in agriculture (Meuwissen *et al*. 2001; Liu & Cordes 2004; Goddard & Hayes 2009; Börner *et al*. 2012; Gebgardt 2013) and the pace of their application are likely to increase expontentially as a result the rapid improvement in cutting-edge next generation sequencing technologies (NGSTs) (Liu 2011; Eggen 2012; Hayes *et al*. 2013).

Molecular markers provide a practical approach for assisting MAS/GS breeding technologies developed in GFP stock enhancement programs. Introduction of NGST has revolutionised several genetic applications including generation of expressed sequence tag (EST) markers, allowed non-model species to be studied including aquaculture species with unprecedented genomic and transcriptomic coverage, and allowed biologists to address questions previously far beyond the reach of earlier technologies (Helyar *et al*. 2012; Ma *et al*. 2012; Ventura *et al*. 2013). Most studies, however, applied quantitative genetic approaches to improving production traits using traditional molecular markers in aquaculture species up to 2010, notably in Atlantic salmon (Quinton *et al*. 2005; Houston *et al*. 2008; Gheyas *et al*. 2009), tilapia (Eknath *et al*. 2007), penaeid prawns (Gitterle *et al*. 2005; Li *et al*. 2006; Ibarra *et al*. 2007) and freshwater crayfish (Jerry *et al*. 2005). While actual genetic response for improved growth rate in these species is high and has ranged from 10 to 25% per generation, MAS/GS employing genomic data, can in theory increase the rate of genetic gains by up to 25 to 50% per generation compared with outcomes that result from conventional animal

selective breeding approaches in isolation (Dekkers & Hospital 2002; Villanueva *et al*. 2005; Ansari-Mahyari *et al*. 2008; Morrell *et al*. 2012; Hayes *et al*. 2013; Taylor 2013).

The basic idea behind MAS/GS is that variants in specific coding gene loci that influence variation in important QTLs may be targeted directly in selection programs. For successful application of MAS/GS, a large number of samples are required to measure phenotypic values in the training population and also over successive generations to avoid the loss of association between phenotypes and markers due to genetic recombination (Gjedrem *et al*. 2012; Lind *et al*. 2012; Morrell *et al*. 2012; Taylor 2013). While traits can be controlled by single genes, most QTL's of economic importance (including growth rate) are likely to be controlled by a set of gene loci encoded epistatically (Fiehn 2001). The genes may likely interact in several gene pathways. Despite the fact that there is potential for epistatic interactions among genes to contribute to heritable variation (phenotypic variation), this is still a controversial issue because response to environmental variation is unpredictable, a number of studies have provided evidence that this genetic interaction can constitute substantially to qualtitative traits (Xu *et al*. 2011; Springer 2013). Therefore, effort has been directed in modern genetic/genomic studies to understand the stability and heritability of epistatic variation and to determine whether this variation can effectively be captured in breeding programs, or whether it may be too unstable for practical use (Jiang *et al*. 2011; Park & Lehner 2013). Assuming that multiple genes contribute variation to a target trait, certain candidate genes can also often constitute large effects. Candidate genes that show a large effect on preferred phenotypes are referred to as 'major' genes. Information about QTLs can therefore add to accuracy when estimating breeding values (Hulata 2001; Liu & Cordes 2004). Where genetic effects on QTLs are relatively large, such candidate genes can be exploited in breeding programs (e.g. growth performance) to target specific production traits or to identify specific high-performing individuals (Dunham 2004; Liu & Cordes 2004). Using these approaches, commercial breed improvement programs have reported up to twice the rate of genetic gain compared with simple phenotypic selection in beef cattle (meat marbling) (Lee *et al*. 2010; Sukegawa *et al*. 2010) and some plant crops (weight at harvest) (Crosbie *et al*. 2006; Ragot & Lee 2007). Application of MAS/GS to GFP stock improvement is new, however, and functional genomic and proteomic information on this species is essentially lacking. Therefore, comprehensive genomic data that comes primarily from EST collections can help promote GFP culture industries worldwide and allow rapid development of improved stocks for culture industries.

Optimizing genetic diversity in stocks in combination with MAS/GS approaches that apply genomic data can assist rapid development of improved culture lines. In 2010, when the current project was initiated, had been developed only limited genetic information is available for GFP and only a few SNP markers (Thanh *et al*. 2010b) have been developed for GFP that can contribute to the study of production trait responses (e.g. growth rate). In addition, only a limited number of microsatellite markers (Karaket *et al*. 2011) have been used to assess early growth performance among different GFP strains. Earlier genetic studies of GFP have been directed primarily at addressing evolutionary questions (de Bruyn *et al*. 2004a, b; de Bruyn *et al*. 2005; de Bruyn & Mather 2007), gene characterization of possible immune responses (Hsieh *et al*. 2006; Baruah *et al*. 2009; Cam *et al*. 2009; Sung *et al*. 2009) or at factors determining sexual maturation (Ngernsoungnern *et al*. 2009: Ventura *et al*. 2011). While significant technological developments in molecular genetics and genomics can offer a means to improve selection accuracies (i.e. estimating breeding values [EBVs]) and estimating selection intensities for many traits, to date only a single study has focused on actual identification and isolation of genetic markers linked to important phenotypic traits in GFP. This SNP study examined the correlation between specific polymorphisms in two candidate genes with individual growth performance. A strong positive association was identified between certain SNPs in the crustacean hyperglycaemic hormone (CHH) gene and growth rate (Thanh *et al*. 2010b). This SNP marker, if confirmed to directly influence growth phenotype in GFP, has potential applications in breeding programs designed to optimize individual growth rate. In general, growth-related traits with moderate to high heritabilities respond well to selection in aquaculture organisms if traditional approaches are applied. In addition, extensive genetic variation is also still available in wild stocks of many aquatic species (Wringe *et al*. 2010; Sánchez-Molano *et al*. 2011; Sauvage *et al*. 2012; Chiasson *et al*. 2013). For GFP, the availability of growth performance to respond positively to conventional selection was confirmed as had been reported previously in a number of aquaculture species (Thanh *et al*. 2009; Dinh *et al*. 2012; Dinh *et al*. 2013). Growth is, however, a complex genetic trait and many loci can interact with environmental factors and they are also likely to contribute to overall phenotypic variation as well. Applicable data on gene loci affecting growth in GFP is lacking so there is an urgent need to expand genomic studies in this species because the first attempt to develop improved GFP culture lines were cultured in a single pond to minimize environmental effects (Dinh *et al*. 2012; 2013).

Recent technological advances in high throughput DNA sequencing that allow rapid, cost-efficient genome and/or transcriptomic analysis have the potential to revolutionize genomics research on GFP and will enable researchers to focus on a large number of functional genomics questions that previously could not be easily addressed (Vera *et al*. 2008; Wheat 2008; Gilad *et al*. 2009). In particular, Roche 454 Genome Sequencing (GS) FLX, one of the next generation sequencing technologies, offers great potential as this technology can provide a rapid shotgun approach for screening large components of the genomes of target non-model species without the need for any prior knowledge about the genome. The technology increases throughput of DNA sequencing while reducing time, labour and costs (Margulies *et al*. 2005; Emrich *et al*. 2007; Vera *et al*. 2008) and has proven to be a versatile platform for producing long DNA sequence reads, with exceptional accuracy, and ultra-high throughput sequencing for model and non-model organisms compared with earlier sequencing technologies (Droege & Hill 2008; Vera *et al*. 2008; Wheat *et al*. 2008). Today there are a number of NGS technologies (i.e. SOLiD, Illumina, Ion-Torrent) that can provide comprehensive genome coverage using pair-end and made-pairs strategies for any model and non-model species. The pace of technological development has been very rapid and capacity increases expontentially. When the currnet project was initiated, however, alternative NGSTs were better suited to organisms where a reference genome was already available because they could provide only relatively short sequence read lengths and would this was a major obstacle in GFP transcriptome studies. The current project sought therefore, to accelerate the pace of gene discovery in GFP, where no reference genome currently exists, and to contribute to genetic improvement programs for GFP culture lines by significantly increasing genetic data on expressed transcripts in the target species. Application of this approach in GFP can provide baseline data for development of genomics toolkits for GFP in the future and can assist stock improvement of this species via MAS/GS.

The structure of the following sections in this study is as follows:

Chapter 2: The aim was to identify potentially important putative or candidate genes with functional roles in growth-related traits relevant for production of GFP in culture. Essentially, this study was a review of work conducted on other species (both aquatic and non-aquatic) that identified and characterized gene loci with demonstrated or putative roles in growth and growth-related phenotypic variation as a basis for refining the list of potential target genes in GFP.

Chapter 3: The aim was to develop and characterise a transcriptome dataset for GFP from fast and slow growing GFP individuals available from a breeding program in Vietnam. Here we sought to capture sequence information from hundreds of thousands of random gene products (mRNAs) that are expressed in target tissues (muscle, testis and ovary) in fast and slow growing female GFP individuals using 454 GS-FLX. A functional and computational genomics approach was used for gene discovery and gene annotation in GFP, including identification of a number of potentially important putative or candidate genes related to quantitative traits of interest.

Chapter 4: The aim was to develop a set of SNP markers in potential growth-related candidate genes that could then be tested for correlations with growth performance in phenotypically divergent GFP from a current stock improvement program in Vietnam. Thereafter, the aim was to perform association studies with the SNPs in selected candidate genes by assessing SNP frequency distribution (genotyping) in fast and slow growth GFP families. This included identifying correlations between SNP and individual growth performance in fast and slow growth GFP individuals and families.

Chapter 5: The aim was to discuss the general results, outcomes and potential applications of the data presented in earlier sections of the thesis, and to identify and discuss potential limitations. The value of markers and other resources generated as part of this study, and how they can best be utilized by researchers and industry in future GFP stock improvement programs are also discussed.

Hence, the specific aims of the study were to: (1) identify genes with a potential role in growth and growth-related traits in GFP [Chapter 2]; (2) develop a transcriptomic dataset for GFP that can be used to mine sequences related to individual growth performance [Chapter 3]; (3) characterise mutations potentially affecting individual growth rate in GFP, develop a set of SNP markers for screening in GFP culture lines, and examine correlations between allelic variation in the set of SNPs in candidate growth genes and growth phenotypes in fast and slow growth GFP families available from a stock improvement program in Vietnam [Chapter 4].

Account of research progress linking the research papers:

The thesis was designed as a set of "stand alone" scientific papers to be accepted and published independently before they were integrated into a comprehensive thesis to address the overall aim that was to develop and trial new genomic resources for GFP that could assist current and future stock improvement efforts for cultured lines, specifically to develop fast growth lines for the culture industry. At the time of submission of the thesis the status of the research papers were as follows:

Paper 1 (Chapter 2) Title: Genes and growth performance in crustacean species: A review of relevant genomic studies in crustacean and other taxa. (*Reviews in Aquaculture*, **Online Version**)

Paper 2 (Chapter 3) Title: Transcriptomics of a giant freshwater prawn (*Macrobrachium rosenbergii*): *de novo* assembly, annotation and marker discovery. (*PLoS One* **6**: e27938)

Paper 3 (Chapter 4) Title: A candidate gene association study for growth performance in an improved giant freshwater prawn (*Macrobrachium rosenbergii*) culture line. *Accepted in Marine Biotechnology.*

Please note: Figures and Tables are re-initialised in each chapter to maintain the independence of each published research paper.

REFERENCES

Aflalo ED, Raju DVSN, Bommi NA, Verghese JT, Samraj TYC, Hulata G *et al*. (2012) Toward a sustainable production of genetically improved all-male prawn (*Macrobrachium rosenbergii*): Evaluation of production traits and obtaining neo-females in three Indian strains. *Aquaculture* **338-341**: 197–207.

Ansari-Mahyari S, Sørensen AC, Lund MS, Thomsen H, Berg P (2008) Across-family marker-assisted selection using selective genotyping strategies in dairy cattle breeding schemes. *Journal of Dairy Science* **91**: 1628–1639.

Bairoch A (2000) The ENZYME database. *Nucleic Acids Research* **28**: 304–305.

Baruah K, Cam DTV, Dierckens K, Wille M, Defoirdt T, Sorgeloos P *et al*. (2009) In vivo effects single or combined N-acyl homoserine lactone quorum sensing signals on the performance of *Marcrobrachium rosenbergii* larvae. *Aquaculture* **288**: 233–238.

Bhassu S, See LM, Hassan R, Sira SS, Tan SG (2008) Isolation and characterization of microsatellite loci in the Malaysian giant freshwater prawn, *Marcrobrachium rosenbergii*. *Molecular Ecology Resources* **8**: 983–985.

Börner V, Teuscher F, Reinsch N (2012) Optimum multistage genomic selection in dairy cattle. *Journal of Dairy Science* **95**: 2097–2107.

Bouck A, Vision T (2007) The molecular ecologist's guide to expressed sequence tags. *Molecular Ecology* **16**: 907–924.

de Bruyn M, Wilson JC, Mather PB (2004a) Huxley's line demarcates extensive genetic divergence between eastern and western forms of the giant freshwater prawn, *Marcrobrachium rosenbergii*. *Molecular Phylogenetics and Evolution* **30**: 251–257.

de Bruyn M, Wilson JC, Mather PB (2004b) Reconciling geography and genealogy: phylogeography of giant freshwater prawns from the Lake Carpentaria region. *Molecular Ecology* **13**: 3515–3526.

de Bruyn M, Nugroho E, Mokarrom Md, Wilson JC, Mather, PB (2005) Phylogeographic evidence for the existence of an ancient biogeographic barrier: the Isthmus of Kra Seawy. *Heredity* **94**: 370–378.

de Bruyn M, Mather PB (2007) Molecular signatures of Pleistocene sea-level changes that affected connectivity among freshwater shrimp in Indo-Australian waters. *Molecular Ecology* **16**: 4295–4307.

Cam DTV, Nhan DT, Ceuppens S, Hao NV, Dierckens K, Wille M *et al*. (2009) Effect of N-acyl homoserine lactone-degrading enrichment cultures on *Marcrobrachium rosenbergii* larviculture. *Aquaculture* **294**: 5–13.

Chand V, de Bruyn M, Mather PM (2005) Microsatellite loci in the eastern form of the giant freshwater prawn (*Marcrobrachium rosenbergii*). *Molecular Ecology Notes* **5**: 308–310.

Charoenatawee K, Poompuang S, Na-Nakorn U (2006) Isolation and characterization of microsatellites in giant freshwater prawn *Marcrobrachium rosenbergii*. *Molecular Ecology Notes* **6**: 823–825.

Chareontawee K, Poompuang S, Na-Nakorn U, Kamonrat W (2007) Genetic diversity of hatchery stocks of the giant freshwater prawn (*Marcrobrachium rosenbergii*) in Thailand. *Aquaculture* **271**: 121–129.

Chiasson MA, Quinton CD, Danzmann RG, Ferguson MM (2013) Comparative analysis of genetic parameters and quantitative trait loci for growth traits in Fraser strain

Arctic charr (*Salvelinus alpinus*) reared in freshwater and brackish water environments. *Journal of Animal Science* **91**: 2047–2056.

Crosbie TM, Eathington SR, Richard Johnsn Sr G, Edwards M, Reiter R, Stark S *et al*. (2006) Plant breeding: Past, present, and future. P. 3–50. In K.R. Lamkey and M. Lee (ed.) Plant breeding: The Arnel R. Hallauer International symponisium. Blackwell, Ames, IA.

Dekkers JCM, Hospital F (2002) The use of molecular genetics in the improvement of agricultural populations. An overview of applications of MAS in plants and livestock. *Nature Reviews Genetics* **3**: 22–32.

Dinh H, Coman G, Hurwood D, Mather P (2012) Experimental assessment of the utility of VIE tags in a stock improvement program for giant freshwater prawn (*Macrobrachium rosenbergii*) in Vietnam. *Aquaculture Research* **43**: 1471–1479.

Dinh H, Vu NT, Nguyen NH, Ponzoni RW, Hurwood DA, Mather PM (2013) Quantitative genetic parameter estimates for body and carcass traits in a cultured stock of giant freshwater prawn (*Macrobrachium rosenbergii*) selected for harvest weight in Vietnam. *Aquaculture* **404-405**: 122–129.

Divu D, Khushiramani R, Malathi S, Karunasagar I, Karunasagar I (2008) Isolation, characterization and evaluation of microsatellite DNA markers in the giant freshwater prawn, *Marcrobrachium rosenbergii*, from South India. *Aquaculture* **284**: 481–284.

Droege M, Hill B (2008) The Genome Sequencer FLX^{TM} system-Longer reads, more applications, straight forward bioinformatics and more complete data sets. *Journal of Biotechnology* **136**: 3–10.

Du ZQ, Ciobanu DC, Onteru SK, Gorbach D, Mileham AJ, Jaramillo, G *et al*. (2009) A gene-based SNP linkage map for pacific white shrimp, *Litopenaeus vannamei*. *Animal Genetics* **41**: 286–294.

Dunham RA (2004) Aquaculture and Fisheries Biotechnology: Genetic approaches. CABI Publishing, UK. 372 pp.

Eggen A (2012) The development and application of genomic selection as a new breeding paradigm. *Animal Frontiers* **2**: 10–15.

Eknath AE, Bentsen HB, Ponzoni RW, Rye M, Nguyen NH, Thodesen J *et al*. (2007) Genetic improvement of farmed tilapias: composition and genetic parameters of a synthetic base population of *Oreochromis niloticus* for selective breeding. *Aquaculture* **273**: 1–14

Emrich SJ, Barbazuk WB, Li L, Schnable PS (2007) Gene discovery and annotation using LCM-454 transcriptome sequencing. *Genome Research* **17**: 69–73.

FAO (2012) Fisheries Statistical Database, Global Aquaculture Production (Fisheries Global Information System, online query).

Fiehn O (2001) Combining genomics, metabolome analysis, and biochemical modeling to understand metabolic networks. *Comparative and Functional Genomics* **2**: 155–168.

Fjalestad KT (2005) Breeding strategies. In: Gjerdrem, T. (Ed.), Selection and Breeding Programs in Aquaculture. Springer, the Netherlands, pp. 145–158.

Gebhardt C (2013) Bridging the gap between genome analysis and precision breeding in potato. *Trends in Genetics* **29**: 248–256.

Gheyas AA, Guy DR, Mota-velasco J, Tinch AE, Woolliams JA (2009) Effect of a major QTL affecting IPN resistance on production traits in Atlantic salmon. Abstract book, the 10th International Symposium on Genetics in Aquaculture 'Role of aquaculture genetics in addressing global food crisis', 22-26 June 2009, Bangkok, Thailand, pp. 24.

Gjedrem T (2012) Genetic improvement for the development of efficient global aquaculture: A personal opinion review. *Aquaculture* **344-349**: 12–22.

Gilad Y, Pritchard JK, Thornton K (2009) Characterizing natural variation using next-generation sequencing technologies. *Trends in Genetics* **25**: 463–471.

Gitterle T, Rye M, Salte R, Cock J, Johansen H, Lozano C *et al*. (2005) Genetic (co)variation in harvest body weight and survival in *Penaeu* (*Litopenaeus*) *vannamei* under standard commercial conditions. *Aquaculture* **243**: 83–92.

Goddard ME, Hayes BJ (2009) Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Review Genetics* **10**: 381–391.

Gorbach DM, Hu ZL, Du ZQ, Rothschild MF (2009) SNP discovery in *Litopenaeus vannamei* with a new computational pipeline. *Animal Genetics* **40**: 106–109.

Hamasaki K, Kitada S (2008) Potential of stock enhancement for decapods crustaceans. *Reviews in Fisheries Science* **16**: 164–174.

Hauser L, Seeb JE (2008) Advances in molecular technology and their impact on fisheries genetics. *Fish and Fisheries* **9**: 473–486.

Hauser L, Carvalho GR (2008) Paradigm shifts in fisheries genetics: ugly hypotheses slain by beautiful facts. *Fish and Fisheries* **9**: 333–362.

Hayes BJ, Lewin HA, Goddard ME (2013) The future of livestock breeding: genomic selection for efficiency, reduced emissions intensity, and adaptation. *Trends in Genetics* **29**: 206–214.

Helyar SJ, Limborg MT, Bekkevold D, Babbucci M, van Houdt J, Maes GE *et al*. (2012) SNP discovery using next generation transcriptomic sequencing in Atlantic herring (*Clupea harengus*). *PLoS One* **7**: e42089.

Houston RD, Haley CS, Hamilton A, Guy DR, Tinch AE, Taggart JB *et al*. (2008) Major quantitative trait loci affect resistance to infectious pancreatic necrosis in Atlantic salmon (*Salmo salar*). *Genetics* **178**: 1109–1115.

Hsieh CY, Chuang PC, Chen LC, Tu C, Chien MS, Huang KC *et al*. (2006) Infectious hypodermal and haematopoietic necrosis virus (IHHNV) infections in giant freshwater prawn, *Marcrobrachium rosenbergii*. *Aquaculture* **25**: 73–79.

Hulata G (2001) Genetic manipulations in aquaculture: a review of stock improvement by classical and modern technologies. *Genetica* **111**: 155–173.

Hutchings JA, Fraser DJ (2008) The nature of fisheries-and farming-induced evolution. *Molecular Ecology* **17**: 294–313.

Ibarra AM, Racotta IS, Arcos FG, Palacios E (2007) Progress on the genetics of reproductive performance in penaeid shrimp. *Aquaculture* **268**: 23–43.

Jiang H, Xu L, Gu Z (2011) Growth of novel epistatic interactions by gene duplication. *Genome Biology and Evolution* **3**: 295–301.

Jerry DR, Purvis IW, Piper LR, Dennis CA (2005) Selection for faster growth in the freshwater crayfish Cherax destructor. *Aquaculture* **247**: 169–176.

Karaket T, Poompuang S, Na-Nakorn U, Kamonrat W, Hallerman E (2011) DAN macrosatellite-based evaluation of early growth performance among strains of freshwater prawn *Macrobrachium rosenbergii* de Man. *Aquaculture* **311**: 115–122.

Kitcharoen N, Rungsin W, Koonawootrittriron S, Na-Nakorn U (2012) Heritability for growth traits in giant freshwater prawn, *Macrobrachium rosenbergii* (de Mann 1879) based on best linear unbiased prediction methodology. *Aquaculture Research* **43**: 19–25.

Kucuktas H, Wang S, Li P, He C, Xu P, Sha Z *et al*. (2009) Construction of genetic linkage maps and comparative genome analysis of catfish using gene-associated markers. *Genetics* **181**: 1649–1660.

Lander ES, Botstein D (19890 Mapping Mendelian factors underlying quantitative traits using RFLP linkage map. *Genetics* **121**: 185–199.

Lee BY, Lee WJ, Streelman JT, Carleton KL, Howe AE, Hulata G *et al*. (2005) A second-generation genetic linkage map of tilapia (*Oreochromis* spp.). *Genetics* **170**: 237–244.

Lee SH, van der Werf JH, Lee SH, Park EW, Oh SJ, Gibson JP *et al*. (2010) Genetic polymorphisms of the bovine Fatty acid binding protein 4 gene are significantly associated with marbling and carcass weight in Hanwoo (Korean Cattle). *Animal Genetics* **41**: 442–444.

Li Y, Byrne K, Miggiano E, Whan V, Moore S, Keys S *et al*. (2003) Genetic mapping of the kuruma prawn *Penaeus japomicus* using AFLP markers. *Aquaculture* **219**: 143–156.

Li YL, Dierens L, Byrne K, Miggiano E, Lehnert S, Preston N *et al*. (2006) QTL detection of production traits for the Kuruma prawn *Penaeus japonicus* (Bate) using AFLP markers. *Aquaculture* **258**: 198–210.

Li Z, Li J, Wang Q, He Y, Liu P (2006) AFLP-based genetic linkage map of marine shrimp *Penaeus* (*Fenneropenaeus*) *chinensis*. *Aquaculture* **261**: 463–472.

Lind CD, Ponzoni RW, Nguyen NH, Khaw HL (2012) Selective breeding in fish and conservation resources for aquaculture. *Reproduction in Domestic Animals* **47(Suppl. 4)**: 255–263.

Liu ZJ, Cordes JF (2004) DNA marker technologies and their applications in aquaculture genetics. *Aquaculture* **242**: 735–736.

Liu Z (2011) Development of genomic resources in support of sequeicng, assembly, and annotation of the catfish genome. *Comparative Biochemistry and Physiology D* **6**: 11–17.

Luan S, Yang G, Wang J, Luo K, Zhang Y, Gao Q *et al*. (2012) Genetic parameters and response to selection for harvest body weight of the giant freshwater prawn *Macrobrachium rosenbergii*. *Aquaculture* **362-363**: 88–96.

Lyons RE, Dierens LM, Tan SH, Preston NP, Li Y (2007) Characterization of AFLP markers associated with growth in the Kuruma prawn, *Marsupenaeus japonicus*, and identification of a candidate gene. *Marine Biotechnology* **9**: 712–721.

Ma K, Qiu G, Feng J, Li J (2012) Transcriptome analysis of the Oriental river prawn, *Macrobrachium nipponense* using 454 pyrosequencing for discovery of genes and markers. *PLoS One* **7**: e39727.

Maneeruttanarungroj C, Pongsomboon S, Wuthisuthimethavee S, Klinbunga S, Wilson KJ, Swan J *et al*. (2006) Development of polymorphic expressed sequence tag-derived microsatellites for the extension of the genetic linkage map of the black tiger shrimp (*Penaeus monodon*). *Animal Genetics* **37**: 363–368.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA et al, 2005. Genome sequencing in microfabricated high-density picolitre reactions. *Nature* **437**: 376–380.

McClelland EK, Naish KA (2008) A genetic linkage map for coho salmon (*Oncortynchus kisutch*). *Animal Genetics* **39**: 214–218.

Meglécz E, Constedoat C, Dubut V, Gilles A, Malausa T, Pech N *et al*. (2010) QDD: a user-friendly program to select microsatellite markers and design primers from large sequencing projects. *Bioinformatics Application Note* **26**: 403–404.

Meuwissen TH, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**: 1819–1829.

Miller AD, Murphy NP, Burridge CP, Austin CM (2005) Complete mitochondrial DNA sequences of the Decapod Crustaceans *Pseudocarcinus gigas* (Menippidae) and *Marcrobrachium rosenbergii* (Palaemonidae). *Marine Biotechnology* **7**: 339–349.

Min SL, Tan SG, Hassan R, Siraj SS, Bhassu S (2009) Development of microsatellite markers from an enriched genomic library for the genetic analysis of the Malaysian giant freshwater prawn, *Macrobrachium rosenbergii*. *Biochemical Genetics* **47**: 722–726.

Moen T, Hayes B, Baranski M, Berg PR, Kjøglum S, Koop BF *et al*. (2008) A linkage map of the Atlantic salmon (*Salmo salar*) based on EST-derived SNP markers. *BMC Genomics* **9**: 233.

Morrell PL, Buckler ES, Ross-Ibarra J (2012) Crop genomics: advances and applications. *Nature Review Genetics* **13**: 85–96.

Nhan DT, Wille M, Hung LT, Sorgeloos P (2009) Comparison of reproductive performance and offspring quality of giant freshwater prawn (*Macrobrachium rosenbergii*) broodstock from different regions. *Aquaculture* **298**: 36–42.

Nhan DT, Wille M, Hung LT, Sorgeloos P (2010) Effects of larval stocking density and feeding regime on larval rearing of giant freshwater prawn (*Macrobrachium rosenbergii*). *Aquaculture* **300**: 80–86.

New MB (2000) Commercial freshwater prawn farming around the world. In: New, M.B, Valenti, W.C. (Eds.), Freshwater Prawn Culture–The farming of *Macrobrachium rosenbergii*. Blackwell Science Ltd, Oxford, London, pp. 290–323.

New MB (2002) Farming freshwater prawns–a manual for the culture of giant river prawn (*Macrobrachium rosenbergii*). FAO Fisheries Technical Paper, vol. 428. FAO, Rome.

New MB (2005) Freshwater prawn farming: global status, recent research and a glance at the future. *Aquaculture Research* **36**: 210–230.

Ngernsoungnern P, Ngernsoungnern A, Weerachatyanukui, Meeratana P, Hanna P, Sobhon P *et al*. (2009) Abalone egg-laying hormone induces rapid ovarian maturation and early spawning of giant freshwater prawn, *Macrobrachium rosenbergii*. *Aquaculture* **296**: 143–149.

O'Leary NA, Trent III HF, Robalino J, Peck MET, McKillen DJ, Gross PS (2006) Analysis of multiple tissue-specific cDNA libraries from the Pacific whiteleg shrimp, *Litopenaeus vannamei*. *Integrative and Comparative Biology* **46**: 931–939.

Park S, Lehner B (2013) Epigenetic epistatic interactions constrain the evolution of gene expression. *Molecular Systems Biology* **9**: 645.

Phuong NT, Hai TN, Hien TTT, Bui TV, Huong DTT, Son VN *et al*. (2006) Current status of freshwater prawn culture in Vietnam and the development and transfer of seed production technology. *Fisheries Science* **72**: 1–12.

Ponzoni RW, Nguyen NH, Khaw HL (2007) Investment appraisal of genetic improvement programs in Nile tilapia (*Oreochromis niloticus*). *Aquaculture* **269**: 187–199.

Quinton CD, McMillan I, Glebe B (2005) Development of an Atlantic salmon (*Salmo salar*) genetic improvement program: genetic parameters of harvest body weight and carcass quality traits estimated with animal models. *Aquaculture* **247**: 211–217.

Ragot M, Lee M (2007) Marker-assisted selection in maize: Current status, potential, limitations and perspective from the private and public sectors. P. 117–150. In E.P. Guimarães et al. (ed.) Marker-assisted selection, current status and future perspectives in crops, livestock, forestry, and fish. FAO, Rome.

Rexroad III CE, Palti Y, Gahr SA, Vallejo RL (2008) A second generation genetic map for rainbow trout (*Oncorhynchus mykiss*). *BMC Genetics* **9**: 74.

Robalino J, Camegie RBm O'Leary N, Ouvry-Patat SA, de la Vega E, Gross PS *et al*. (2009) Contributions of functional genomics and proteomics to the study of immune responses in the Pacific white leg shrimp *Litopenaeus vannamei*. *Veterinary Immunology and Immunopathology* **128**: 110–118.

Sánchez-Molano E, Cerna A, Toro MA, Bouza C, Hermida M, Pardo BG, Cabaleiro S *et al*. (2011) Detection of growth-related QTL in turbot (*Scophthalmus maximus*). *BMC Genomics* **12**: 473.

Sauvage C, Vagner M, Derôme, Audet C, Bernatchez L (2012) Coding gene SNP mapping reveals QTL linked to growth and stress response in brook charr (*Salvelinus fontinalis*). *Genes Genomes Genetics* **2**: 707–720.

Schwantes VS, Diana JS, Yi Y (2009) Social, economic, and production characteristics of giant freshwater prawn *Macrobrachium rosenbergii* culture in Thailand. *Aquaculture* **287**: 120–127.

Springer NM (2013) Epigenetics and crop improvement. *Trends in Genetics* **29**: 241–247.

Staelens J, Rombaut D, Vercauteren I, Argue B, Benzie J, Vuylsteke M (2008) High-density linkage maps and sex-linked markers for the black tiger shrimp (*Penaeus monodon*). *Genetics* **179**: 917–925.

Sukegawa S, Miyake T, Takahagi Y, Murakami H, Morimatsu F, Yamada T *et al*. (2010) Replicated association of the single nucleotide polymorphisms in EDG1 with marbling in three general populations of Japanese Black beef cattle. *BMC Research Notes* **3**: 66.

Sun X, Liang L (2004) A genetic linkage map of common carp (*Cyprinus carpio* L.) and mapping of a locus associated with cold tolerance. *Aquaculture* **238**: 165–172.

Sung HH, Yang CW, Lin YH, Chang PT (2009) The effect of two CpG oligodeoxynucleotides with different sequences on haemocytic immune responses of giant freshwater prawn, *Macrobrachium rosenbergii*. *Fish and Shellfish Immunology* **26**: 256–263.

Tassanakajon A, Klinbunga S, Paunglarp N, Rimphanitchayakit V, Udomkit A, Jitrapakdee S *et al*. 2006. Penaeus monodon gene discovery project: The generation of an EST collection and establishment of a database. *Gene* **384**: 104–112.

Taylor JF (2013) Implementation and accuracy of genomic selection. *Aquaculture* **Online version**.

Thanh NM, Ponzoni RW, Nguyen NH, Vu NT, Barnes A, Mather PB (2009) Evaluation of growth performance in a diallel cross of three strains of giant freshwater prawn (*Macrobrachium rosenbergii*) in Veitnam. *Aquaculture* **287**: 75–83.

Thanh NM, Nguyen NH, Ponzoni RW, Vu NT, Barnes AC, Mather PB (2010a) Estimates of strain additive and non-additive genetic effects for growth traits in a diallel cross of three strains of giant freshwater prawn (*Macrobrachium rosenbergii*) in Veitnam. *Aquaculture* **299**: 30–36.

Thanh NM, Barnes AC, Mather PB, Yutao L, Lyons RE (2010b) Single nucleotide polymorphisms in the actin and crustacean hyperglycemic hormone genes and their correlation with individual growth performance in giant freshwater prawn *Macrobrachium rosenbergii*. *Aquaculture* **301**: 7–15.

Ventura T, Rosen O, Sagi A (2011) From the discovery of the crustacean androgenic gland to the insulin-like hormone in six decades. *General Comparative Endocrinology* **173**: 381–388.

Ventura T, Manor R, Aflalo ED, Chalifa-caspi V, Weil S, Sharabi O, Sagi A (2013) Post-embryonic transcriptomes of the prawn Macrobrachium rosenbergii: Multigenic succession through metamorphosis. *PLoS One* **8**: e55322.

Vera JC, Wheat CW, Fescemyer HW, Frilander MJ, Crawford DL, Hanski I *et al*. (2008) Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. *Molecular Ecology* **17**: 1636–1647.

Villanueva B, Pong-Wong R, Fernández J, Toro MA (2005) Benefits from marker-assisted selection under an additive polygenic genetic model. *Journal of Animal Sciences* **83**:1747–1752.

Wheat CW (2008) Rapidly developing functional genomics in ecological model systems via 454 transcriptome sequencing. *Genetica* **138**: 433–451.

Wilder MN, Yang WJ, Huong DTT, Maeda M (1999) Reproductive mechanisms in the giant freshwater prawn. *Macrobrachium rosenbergii* and cooperative research to improve seed production technology in the Mekong delta region of Vietnam, *UJNR Technical Report* No. **28**.

Worm B, barbier EB, Beaumont N, Duffy JE, Folke C, Halpern BS *et al*. (2006) Impacts of biodiversity loss on ocean ecosystem services. *Science* **314**: 787–790.

Wringe BF, Devlin RH, Ferguson MM, Moghadam HK, Sakhrani D, Danzmann RG (2010) Growth-related quantitative trait loci in domestic and wild rainbow trout (*Oncorhynchus mykiss*). *BMC Genetics* **138**: 433–451.

Wu P, Qi D, Chen L, Zhang H, Zhang X, Qin JG *et al*. (2009) Gene discovery from an ovary cDNA library of oriental river prawn *Macrobrachium nipponense* by ESTs annotation. *Comparative Biochemistry and Physiology Part D*, **4**: 111–120.

Xu L, Jiang H, Chen H, Gu Z (2011) Genetic architecture of growth traits revealed by global epistatic interactions. *Genome Biology and Evolution* **3**: 909–914.

Zhang L, Yang C, Zhang Y, Li L, Zhang X, Zhang Q *et al*. (2007) A genetic map of Pacific white shrimp (*Litopenaeus vannamei*): sex-linked microsatellite markers and high recombination rates. *Genetica* **131**: 37–49.

CHAPTER 2: Genes and growth performance in crustacean species: A review of relevant genomic studies in crustaceans and other taxa

- Presented at the 2012 World Aquaculture Society 'Global aquaculture securing our future' 1-5 September 2012, Prague, Czech Republic.

- Published in Reviews in Aquaculture (2013). Reviews in Aquaculture, Online Version.

Preface to Chapter 2

The current study was initiated to understand and identify possible growth-related genes in crustacean species because currently there is only limited genomic information available for any crustacean species. In an effort to stimulate increased research on identification of growth-related genes in crustacean species, we reviewed the available information on; 1) associations between genes and growth reported in crustaceans, 2) growth-related genes involved with moulting, 3) muscle development and degradation genes involved in moulting, and 4) correlations between DNA sequences that have confirmed growth trait affects in farmed animal species used in terrestrial agriculture and related sequences in crustacean species. The information in concert can provide a foundation for increasing the rate at which knowledge about key genes affecting growth traits in crustacean species is gained.

Genes and growth performance in crustacean species: A review of relevant genomic studies in crustaceans and other taxa.

Jung H[1*], Lyons RE[2], Hurwood DA[1], Mather PB[1]

[1]Earth, Environmental and Biological Sciences, Science and Engineering Faculty, Queensland University of Technology, Brisbane, Queensland 4001, Australia

[2]Commonwealth Scientific and Industrial Research Organisation, Livestock Industries, St Lucia, Queensland 4067, Australia

ABSTRACT

Global aquaculture has expanded rapidly to address increasing demand for aquatic protein needs and an uncertain future for wild fisheries. To date, however, most farmed aquatic stocks are essentially wild and little is known about their genomes or the genes that affect important economic traits in culture. Biologists have recognized that recent technological advances including next generation sequencing (NGS) have opened up the possibility of generating genome wide sequence data sets rapidly from non-model organisms at a reasonable cost. In an era when virtually any study organism can 'go genomic', understanding gene function and genetic effects on expressed quantitative trait locus phenotypes will be fundamental to future knowledge development. Many factors can influence individual growth rate in target species but of particular importance in agriculture and aquaculture will be identification and characterization of the specific gene loci that contribute important phenotypic variation to growth because the information can be applied to speed up genetic improvement programs and increase productivity via marker-assisted selection (MAS). While currently there is only limited genomic information available for any crustacean species, a number of putative candidate genes have been identified or implicated in growth and muscle development in some species. In an effort to stimulate increased research on identification of growth-related genes in crustacean species, here we review the available information on; 1) associations between genes and growth reported in crustaceans, 2) growth-related genes involved with moulting, 3) muscle development and degradation genes involved in moulting, and 4) correlations between DNA sequences that have confirmed growth trait affects in farmed animal species used in terrestrial agriculture and related sequences in crustacean species. The information in concert can provide a foundation for increasing the rate at which knowledge about key genes affecting growth traits in crustacean species is gained.

Keywords: next generation sequencing, genomes, quantitative trait locus, crustaceans, genetic effects, growth-related genes, muscle development/degradation genes, growth traits.

All species' names used in this article are as reported in the references cited. Some species' names (and/or scientific authorities) may have changed since publication of the articles cited.

INTRODUCTION

More than 200 aquatic taxa (including crustaceans, fish, molluscs and some algae) are now considered resources to be exploited for human consumption and profit. If current fishing practices and rates of wild fish population exploitation continue, the future of most marine ecosystems and wild fisheries will be uncertain due mainly to overfishing and habitat degradation (Worm *et al*. 2006). As a consequence, in order to meet rising global demand for aquatic animal food, aquaculture has developed rapidly and is now practiced widely around the world. In a resource-constrained world where the redirection of edible foods as feedstock for animal production is coming under increased scrutiny, achieving maximum productivity in aquaculture systems will be an imperative as the global community strives to achieve sustainable food security. Over the past 50 years, genetics has played an increasingly important role in fisheries management (aquaculture, hatchery production and breed development programs) allowing some industries to develop into intensive, high-input production systems that now supply large international markets (e.g. Atlantic salmon in Norway). This paradigm shift has changed perceptions about marine and freshwater environments and the species therein (Hauser & Carvalho 2008). It is increasingly obvious that, as human populations continue to expand, the proportion of seafood that comes from aquaculture must also increase and genetically improved breeds will play a major role in addressing food security issues and contribute to economic growth in the future (Hulata 2001; Hauser & Seeb 2008; Hutchings & Fraser 2008). A number of decapod crustacean taxa constitute important aquaculture species including; crabs, crayfish, lobsters, shrimps and prawns that contribute substantially to the modern US$60 billion global aquaculture industry (FAO 2009). Most cultured crustaceans are high value species and research is now directed at stock enhancement programs to improve their culture performance (Hartnoll 2001; Hamasaki & Kitada 2008). To date, however a lack of genomic resources for most aquatic species, and associated poor understanding of molecular and biochemical processes underpinning growth and other economically-important traits, has hindered efforts to increase productivity of crustacean species in aquaculture.

Genomics is a new science that analyses the whole or partial genomes of target species, and genomic approaches are now being applied in a wide range of species to elucidate genetic factors that contribute to economically important traits and/or phenotypes and to manage genetic diversity in crop and livestock species (Miller 2010; Wheeler *et al*.

2010; Morrell *et al*. 2012). New technologies that allow rapid cost-effective high-throughput sequencing, often referred to as next generation sequencing (NGS), offer significant opportunities to improve our understanding of the domestication process, contribute to the evolution of modern quantitative genetic theory, and can accelerate stock improvement of target populations in agriculture and aquaculture (Morrell *et al*. 2012). NGS has opened up the possibility of rapidly and cost-effectively collating genome wide sequence data sets, even in non-model organisms (Ekblom & Galindo 2011). Since the technical aspects of NGS have been reviewed extensively in both model and non-model species elsewhere [See Shendure & Ji (2008); Gilad *et al*. (2009); Ekblom & Galindo (2011)], the current review will concentrate primarily on information relevant to growth and growth-related traits in crustacean species.

In an era when virtually any study organism can 'go genomic', understanding gene function and genetic effects on expressed phenotypes will be fundamental to future production systems. Modern DNA-sequencing technologies have replaced the once painstaking hunt for individual genes associated with particular phenotypes (Stillman *et al*. 2008; Clark *et al*. 2011; Jung *et al*. 2011; Kawahara-Miki *et al*. 2011) and this development will likely revolutionize genetic improvement programs for cultured crustaceans in the future (Jung *et al*. 2011; Kawahara-Miki *et al*. 2011). Application of genomic approaches to stock improvement in aquaculture species has already resulted in some significant production gains in some aquatic organisms including carp, rainbow trout, Atlantic salmon, tilapia, and catfish (McAndrew & Napier 2011). This is, however, a new horizon for crustacean species and to date studies have been confined mostly to only a small number of penaeid species (Leekitcharoenphon *et al*. 2010; Leu *et al*. 2011) and to a few non-penaeids (Jung *et al*. 2011; Kawahara-Miki *et al*. 2011). While genomic analysis of crustaceans using NGS technologies is at an early developmental stage, it is obvious that the amount and quality of data available will grow rapidly. A major focus of this development will be identification and characterisation of genes and markers that affect variation in economically important traits. These markers, including those affecting growth related traits, can then be applied strategically to speed up genetic improvement programs for cultured crustacean species via marker-assisted selection (MAS) programs. Despite the limited genetic information that is available, a number of putative candidate genes related to growth and muscle development have already been identified in crustacean species. In an effort to stimulate more research on identification of growth-related genes in crustacean species, this review will address the following topics; 1) recent genome and functional genomics (transcriptomics) projects that

have been undertaken since 2008 on five major crustacean groups (crabs, crayfish, lobsters, shrimps and prawns) (Stillman *et al*. 2008), 2) associations between individual genes and growth traits in crustaceans, 3) growth-related genes including muscle build-up and degradation genes involved with moulting, and 4) DNA sequences that have confirmed growth trait affects in farmed terrestrial species and where homologous sequences also exist in crustacean species. Synthesis of this information will provide a baseline for increasing the rate at which knowledge about key genes that affect growth traits in crustacean species is gained.

ADVANCES IN CRUSTACEAN GENOMICS SINCE 2008

A paper by Stillman et al. in 2008 surveyed DNA sequence data (genomic and expressed sequence tags [ESTs]) that were available from public databases in crustacean taxa. The amount of DNA sequence in public databases has grown exponentially since this time as costs and speed of NGS analysis have improved greatly. A draft genome was recently published for the water flea (*Daphnia pulex*) a microcrustacean that is tolerant of a wide variety of different environmental conditions (Colbourne *et al*. 2011). While the *Daphnia* genome has potential to be used as a model genome and scaffold for other crustacean species, the authors noted that, however, data were collected under laboratory environmental conditions, and traits observed and functions demonstrated experimentally (functional annotations) may only represent a subset of those expressed in natural ecosystems. Future availability of empirical data, including new genome sequences from a wider diversity of species can address this problem. To this end, an overview of useful EST, cDNA and transcriptome sequences published for various crustacean species and that is available in public databases are summarized in Table 1.

Combining small EST collections from a number of related taxa can provide the most efficient and powerful method for predicting conserved expressed protein components encoded by uncharacterized genomes (Leekitcharoenphon *et al*. 2010; Leu *et al*. 2011). The purpose of this section of the review, therefore, is to report findings from the most recent genomic and transcriptomic projects undertaken on major crustacean taxa (crabs, crayfish, krill, lobsters, shrimps, prawns, and water flea). All literature surveys conducted for the current study were based on the following criteria: 1) published EST, gDNA, cDNA, and transcriptome data after Stillman *et al*. (2008), 2) data consists of more than 2,000 ESTs and

clones, and 3) excluding microarray expression studies and proteomic analyses. A summary of the genomic literature available on commercially important crustacean species is presented in Table 2. We have focused most attention in this review on only the major crustacean species for which substantial databases have been compiled. Expanded surveys of other Arthropod and/or Crustacea genomic resources (including Amphipoda, Branchiopoda, Copepoda, and Maxillopoda), however, will allow researchers to explore the diversity of functionally important genes within a comparative genomic framework.

MESSAGES FROM GENE INTERACTOIN IN CELLS

The genetics underpinning many phenotypes are complex with contributions often coming from multiple loci as epistatic interactions, a process referred to as gene-gene interactions (Horn *et al*. 2011; Xu *et al*. 2011). In addition, a single gene or a single mutation can affect multiple phenotypic traits via pleiotropy (Wagner & Zhang 2011). Understanding the functional interactions among genes and mutations has become a successful tool for illuminating the genetics of complex traits, and understanding organization within cells and their dynamics (Crombach & Hogeweg 2008; Carter *et al*. 2009). Organisms can be viewed as information processors, highly evolved to recognize environmental conditions and programmed to respond in a way that maximizes their fitness (Carter *et al*. 2009). Growth traits are determined by the timely orchestration of complex interactions among gene variants and environmental factors that are integrated in terms of activation and inhibition of genes. Recent large-scale genetic analyses of model organisms have enabled systematic screening of epistatic interactions as well as the pleiotropic structure of the genotype-phenotype map and provide valuable insights into the functional organization of a eukaryotic cell (Costanzo *et al*. 2010; Wagner & Zhang 2011), including the genetic architecture of growth traits (Szappanos *et al*. 2011; Xu *et al*. 2011; Loehlin & Werren 2012). A simplified diagram describing gene-gene interactions, deduced from model organism studies (Carter *et al*. 2009; Szappanos *et al*. 2011; Wagner & Zhang 2011), is presented in Figure 1.

While detecting and classifying epistatic and/or pleiotropic gene (or mutation) interactions is a powerful approach for deciphering the roles of individual genes and mapping functional relationships among pathways, application of this approach to crustacean species constitutes a real challenge because of the limited genetic and/or genomic information currently available. An alternative approach is to identify and characterise those genes or

gene families that have been implicated in the trait of interest in other organisms, often referred to as a gene candidate approach. By performing association studies to directly test the effects of genetic variants against a phenotypic trait, one can rapidly and inexpensively assess the relative importance of a gene or groups of genes. A number of studies have shown that functionally identified genes (or mutations) in one species are likely to play a similar role in other target species (Fortuna *et al*. 2009; Goddard & Hayes 2009). Identification of those key genes previously associated with phenotypic regulation (e.g. growth, disease, reproduction) in other organisms, and represented across crustacean taxa is therefore potentially a useful tool for bridging functional and structural genomics. While the mitochondrial genome is essential for regulation of energy metabolism as the powerhouse of eukaryotic cells (Hess *et al*. 2009), the current review will focus on the nuclear genome mainly due to a lack of empirical mitochondrial data for crustacean species.

CANDIATE GENES CORRELATED WITH GROWTH

Most eukaryotic genes contain coding (amino acid translated regions-exons) and non-coding regions (introns, and 5' & 3' untranslated regions that include promoter regions). Allelic variation in DNA sequence in any of these regions can potentially alter the structure and activity of the protein they encode, or induce changes in regulation of mRNA transcription levels (De Santis & Jerry 2007). Genetic variants and allele specific gene expression within a functional gene can determine phenotypic variation and hence influence expression of complex traits, growth traits and genetic diseases via regulation of gene expression and mRNA stability (Palacios *et al*. 2009; Ciobanu *et al*. 2010b). While a variety of molecular markers and techniques have been developed to screen allelic polymorphisms in genes, Single Nucleotide Polymorphisms (SNPs) and mRNA expression profiles are currently considered to be the marker and technique of choice, respectively for studying how allelic variation in a gene can affect its regulation by identifying its location in the genome (Palacios *et al*. 2009; Ciobanu *et al*. 2010b; Garvin *et al*. 2010; Seeb *et al*. 2011).

SNPs represent type I markers and are inherited in a co-dominant fashion, they are simple to score and are relatively inexpensive to screen compared with some alternative genetic markers (Garvin *et al*. 2010; Kim *et al*. 2011; Seeb *et al*. 2011). The recent development of technologies that allow the simultaneous genotyping of thousands of SNPs in a single assay has resulted in a number of genome-wide association studies (GWAS) being

completed for model species (McClure *et al*. 2010; Bolormaa *et al*. 2011; Pausch *et al*. 2011). This approach allows researchers to map regions with significant (Quantitative Trait Loci, QTLs) association with the phenotype of interest more accurately (Goddard & Hayes 2009; McClure *et al*. 2010; Pausch *et al*. 2011). For many complex traits, the next step in identifying genes and mutations that underpin phenotypic variation has proven to be a non-trivial exercise even with well annotated genomes. With the exception of Daphnia, there are no annotated genomes for a crustacean and this makes GWAS impossible in these organisms. Although many internal and external environmental factors can influence growth rate (e.g. diet, food ration, stocking density, temperature, salinity, dissolved oxygen, pathogens and water chemistry etc.), here we will focus the discussion on the roles of genetic factors (SNPs and mRNA expression) in specific candidate genes that affect growth rate in crustaceans. A recent study applied 418 SNP markers in Pacific white shrimp, *Litopenaeus vannamei*, to develop genetic maps and to identify QTLs for economically important traits (Du *et al*. 2010). A number of SNPs in anonymous genes were reported to show significant associations with weight gain, growth rate, survival, and/or pathogen resistance in these populations (Ciobanu *et al*. 2010a), demonstrating the value of having markers located within functional genes.

In the following section we will highlight specific candidate genes that potentially may affect growth rate in crustacean species based on previous studies in these organisms or in model species. By way of justifying our supposition, we further describe the functional roles of these genes or gene families on growth, moulting and muscle production and degradation. A summary of described DNA polymorphisms and their relationship to relevant phenotypes involved in regulation of growth in crustaceans is presented in Table 3.

1. 5-Hydroxytryptamine receptor

Several lines of evidence suggest that serotonin (5-hydroxytryptamine, 5-HT), a biogenic amine neurotransmitter found in the nervous systems of all organisms, can affect a diverse array of physiological, behavioural and cognitive functions (Vázquez-Acevedo *et al*. 2009; Zhang *et al*. 2011d). In crustaceans, serotonin has been shown to stimulate the release of several neurohormones including crustacean hyperglycemic hormone (CHH), red pigment-dispersing hormone, neurodepressing hormone and moult-inhibiting hormone (MIH) by interacting with various receptor subtypes (Gerhardt *et al*. 1997; Ongvarrasopone *et al*. 2006). Fourteen 5-HT receptor subtypes have been identified, cloned and then classified into 7 families (5-HT1 to 5-HT7) according to their amino acid sequence homology,

pharmacological binding properties, and coupling to secondary messengers in invertebrates (Escamilla-Chimal *et al*. 2002; Ongvarrasopone *et al*. 2006). A number of 5-HT receptors have been cloned and characterized for their function in crustaceans (Clark *et al*. 2004; Sosa *et al*. 2004; Komali *et al*. 2005). These include 5-HT1, 5-HT2, and 5-HT7-like receptors that show similar pharmacological and signalling properties to their mammalian counterparts (Tierney 2001).

In Pacific white shrimp, *L*. *vannamei*, 2 exonic synonymous SNPs in the 5-HT1 receptor, have been linked to high body weight (Martin-Marti *et al*. 2010). In addition, up-regulation of 5-HT1 in the eyestalk and brain of *Metapenaeus ensis* may be important for understanding the mechanism of 5-HT1 action and the role that its receptor plays in control of reproduction (Tiu *et al*. 2005). Ongvarrasopone *et al*. (2006) suggested that receptor localization plays a critical role in regulating ovarian maturation and spawning in giant tiger shrimp, *Penaeus monodon*. A better understanding of the 5-HT receptor will undoubtedly provide important information about the mechanisms of 5-HT regulation in different aspects of crustacean physiology.

2. Alpha-amylase

In crustaceans, maintenance of appropriate levels of glucose in the hemolymph is essential for supporting several key physiological functions and for responding to a variety of environmental stressors (Lorenzon 2005; Asaro *et al*. 2011). Digestion of glycogenic carbohydrates (i.e. complex polysaccharides that include starch and other disaccharides) provides one of the main sources of hemolymph glucose (Verri *et al*. 2001). Ability to digest starch depends on levels of key enzymes (i.e. amylase and maltase) in the hepatopancreas, the main site of digestive enzyme production in crustaceans.

Amylase (AMY) hydrolyses starch molecules to produce a variety of products including dextrins, and progressively smaller polymers composed of glucose units (Windish *et al*. 1965). α-AMY (α-1,4-glucan-4-glucanohydrolase) is a member of glycoside hydrolase family 13, that catalyses hydrolysis of α-(1,4) glycosidic linkages in starch and related compounds (Janecek 1997; Van der Maarel *et al*. 2002). This enzyme is a major glucosidase that is associated with digestive function in the hepatopancreas (Van Wormhoudt & Sellos 1996; Pavasovic *et al*. 2007; Asaro *et al*. 2011). A number of studies have suggested that high growth rate, lower mortality rate and final body weight are closely linked to digestive capacity (Brito *et al*. 2000; Gamboa-Delgado *et al*. 2003; Pavasovic *et al*. 2007). Among the

full range of digestive enzymes that are present in the crustacean gut, AMY may be an important component that maximizes assimilation rate thereby influencing growth rate (Brito *et al*. 2000; Wang 2007; Simon & Jeffs 2011). Preliminary analysis in Pacific white shrimp, *L. vannamei*, has identified a number of intronic and exonic SNPs in the AMY gene (Glenn *et al*. 2005; Peng *et al*. 2008). To date association studies involving this gene have been limited to 1 intronic SNP correlated with individual body weight in small populations (Glenn *et al*. 2005). While no association was detected, the authors suggested that the chance of identifying significant associations with variation in the AMY gene and body weight were not high (Glenn *et al*. 2005). While variants in the AMY gene have yet to show positive correlations with growth traits in crustaceans, AMY should not be discounted because it has a functional role in digestion and hence potential to influence growth. These enzymes have also been shown to be much more active in freshwater crustaceans (red-claw crayfish) than in the saltwater penaeid species, perhaps reflecting an increased utilisation of plant matter in the diet of many freshwater species (Pavasovic *et al*. 2006). Since feed can contribute to more than 60% of production costs in aquatic species in culture, optimising efficiency of digestive enzymes to deal with artificial feeds can in theory, contribute significantly to improving growth rates in crustacean species (Williams 2007).

3. Cathepsin L

Cysteine proteinases (EC 3.4.22) in the papain superfamily, known as lysosomal cathepsins, include a large number of enzymes that play an essential role in several highly regulated life processes (McGrath 1999). Eleven cathepsins have been characterised (Turk *et al*. 2000) and are divided into two families: the cathepsin L family and the cathepsin B family (Karrer *et al*. 1993; Berti & Storer 1995),

Cathepsin L (CatL, EC 3.4.22.15) is regarded conventionally as a major component of the lysosomal proteolytic system. It is involved in intracellular protein turnover and acts as a scavenger in the cell to clear unwanted proteins (McGrath 1999; Turk *et al*. 2000) and also has diverse physiological roles in antigen presentation (Nakagawa *et al*. 1998), matrix modelling (Yamada *et al*. 2000) and maturation of exopeptidase (Dahl *et al*. 2001). Recent data reported from insects and crustaceans have revealed that CatL plays roles in; food digestion (Hu & Leung 2007), the nucleus of the oocyte (Hu & Leung 2004), the innate immune system (Li *et al*. 2010), and moulting (Le Boulary *et al*. 1996, 1998; Liu *et al*. 2006). In crustaceans, cDNA encoding CatL-like enzymes have been isolated from the digestive

gland (hepatopancreas) of American lobster (*Homarus americanus*) (Laycock *et al*. 1989), and two shrimp species *Penaeus vannamei* (Le Boulay *et al*. 1996) and *M. ensis* (Hu & Leung 2004). According to Aoki *et al*. (2004), an enzyme found in *Pandalus borealis* shares homology with mammalian cathepsins S, L, and K, but shows differences in both enzymatic and structural properties. Although CatL is expressed ubiquitously in a shrimp (extremely high in the hepatopancrease) (Hu & Leung 2004), indirect evidence for a food digestive role for crustacean CatL will require further analysis.

In shrimp, a single intronic SNP in CatL was correlated with high individual body weight in one sampled population (Glenn *et al*. 2005). In addition, genetic studies in a shrimp (Le Boulay *et al*. 1996) and cotton bollworm *Helicoverpa armigera* (Liu *et al*. 2006) suggested an essential role for CatL in moulting. More studies of CatL could provide important information on mechanisms of regulation in crustacean physiology.

4. Cyclophilin

Cyclophilins (Cyps) contain a single conserved peptidylprolyl cis-trans isomerase (PPIase domain) and are abundant and ubiquitous cytosolic proteins expressed constitutively in bacteria to vertebrates (Galat 1999; Opiyo & Moriyama 2009). Due to their PPIase activity, Cyps have been reported to possess diverse regulatory functions in protein folding (Schmid *et al*. 1993), receptor complex stabilization (Leverson & Ness 1998), apoptosis (Lin & Lechleiter 2001) and receptor signalling (Brazin *et al*. 2002). Cyps are also known to be involved in pathogen responses (Fidantsef *et al*. 1999), immunological system activation (Hamilton & Steiner 1998) and as modulators of response to oxidative stress (Guedes *et al*. 2005). In addition, Cyps are involved potentially in promoting endothelial cell proliferation, migration, invasive capacity, and tubulogenesis at low concentrations (Kim *et al*. 2004b).

Studies of Cyps in aquatic animals have suggested that they play an important role in the innate immune system (especially isoform CypA). For example, CypA was identified as being involved in early infection of the bacterium *Edwardsiella ictaluri* in channel catfish *Ictalurus punctatus* (Yeh & Klesius 2008). In a clam (*Venerupis philippinarum*), an isoform of CypA was expressed in haemocytes, suggesting that it could be involved in the defence response against bacterial infections (Chen *et al*. 2011). In giant tiger shrimp *Penaeus monodon*, expression of CypA in the hepatopancreas was up-regulated after stimulation with lipopolysaccharide, indicating that CypA probably contributed to the defence response against bacterial infections (Qiu *et al*. 2009). In *P. monodon*, Tangprasittipap *et al*. (2010) reported a negative correlation between cyclophilin-like expression and individual body

weight. In a study of expression profiles in two size classes of Atlantic pink shrimp (*Farfantepenaeus paulensis*), cyclophilin, hamemocyanin and myosin were strongly expressed in high weight individuals but were only weakly expressed or absent in low weight individuals, (Kaminura *et al*. 2008). This result suggests that these genes may have some effect on individual growth performance, and warrants further study in crustacean species. Improved understanding of the immune response and growth retardation in crustaceans linked to pathogen challenge could contribute to developing new approaches for maximising growth rate in crustaceans in culture.

5. Fatty acid-binding protein

Fatty acid-binding proteins (FABPs) are small (~15-kDa) cytosolic proteins that bind non-covalently to hydrophobic ligands, primarily fatty acids (Esteves & Ehrlich 2006) and are important genes involved in the development of fatness traits (McArthur *et al*. 1999). The physiological roles of FABPs include; uptake and utilization of fatty acids, intracellular fatty acid transport to specific organelles and engagement in metabolic pathways, cell growth and differentiation, cellular signalling, gene transcription, and protection of cellular structures from detergent effects of fatty acids (Ho *et al*. 2002; Storch *et al*. 2002; Zimmerman & Veerkamp 2002; Corsico *et al*. 2004). In FABP families, 12 FABP isoforms have been identified in vertebrates (Liu *et al*. 2008) that can be divided into 2 main groups (Glatz & Van der Vusse 1996): those associated with the plasma membrane ($FABP_{PM}$) and those with intracellular or cytoplasmic proteins ($FABP_C$).

Many studies have focused on finding economically important polymorphic sites in FABP genes in livestock species. In association studies between polymorphisms in FABP orthologs and fat deposition traits in cattle, several SNPs were identified in both exonic and intronic gene regions in the adipocyte FABP gene (A-FABP) (Wang *et al*. 2006) and heart FABP gene (H-FABP) (Wang *et al*. 2007; You *et al*. 2007). Both genes have significant effects on several growth traits including body weight (BW), intramuscular fat content (IMF), abdominal fat weight (AFW), and abdominal fat percentage (AFP) in chickens (Wang *et al*. 2006; You *et al*. 2007).

To date, approximately only 40 FABPs have been identified in invertebrate species (Zimmerman & Veerkamp 2002; Esteves & Ehrlich 2006). In crustaceans, evidence for several putative FABPs have been detected in the crayfish (*Pacifastacus leniusculus*), giant tiger prawn (*P. monodon*), Kuruma prawn (*Marsupenaeus japonicus*), Chinese white shrimp (*F. chinensis*) and Chinese mitten crab (*E. sinensis*) (Söderhäll *et al*. 2006; Lyons *et al*. 2007;

Ren *et al*. 2009; Gong *et al*. 2010; Li *et al*. 2011a). Gene expression analysis of FABPs in *E. sinensis* using RT-PCR, revealed the presence of Es-FABP transcripts in a variety of tissues and Es-FABP expression analysis revealed participation of haemocytes, hepatopancreas, ovary, and testis in lipid nutrient absorption and utilization processes. Es-FABP expression was dependent on the status of ovarian and testis development as evidenced by a role for Es-FABP in lipid transport (Gong *et al*. 2010; Li *et al*. 2011a). In *M. japonicus*, characterisation of an amplified fragment length polymorphism (AFLP) band that resided under a previously identified QTL peak that influenced weight, total length, and carapace length (Li *et al*. 2006), revealed a gene homologous to the elongation of very long chain fatty acids-like (ELOVL) protein family (Lyons *et al*. 2007). An examination of weight differences between individuals showed that individuals carrying isoform A (regardless of whether A- or AB) showed poor growth phenotypes while individuals without isoform A (B- or –) had a positive growth phenotype that contributed 16% to total phenotypic variation. This represents a suitable gene candidate for future growth studies in crustaceans (Lyons *et al*. 2007). ELOVLs have also been shown to have significant effects on growth in yeast (Rossler *et al*. 2003) and play an important role in lipid recruitment in adipose tissue in mice (Westerberg *et al*. 2006). It is not unreasonable, therefore, to predict that changes in FABP activity may have significant biochemical and physiological effects on growth in crustaceans and constitute potential gene candidates that influence growth traits in crustaceans.

6. Fibrillarin

Fibrillarin is a 40-kDa protein that is located in the dense fibrillar component of the nucleolus (Ochs *et al*. 1985) and has essential functions related to RNA splicing and ribosomal RNA processing (Staněk & Neugebauer 2006; Deryusheva & Gall 2009)

Comparisons of amino acid alignments of fibrillarin-like ESTs' from vertebrates and invertebrates have indicated that these proteins are highly conserved for several cellular functions including as an essential protein required for rRNA methylation, pre-rRNA cleavage and ribosome assembly (Venema & Tollervey 1999; Fatica & Tollervey 2002; Brown *et al*. 2003). In addition, fibrillarin is involved potentially in processing and modification of pre-rRNA, regulation of cell proliferation and individual growth by controlling ribosome biosynthesis and p53 functions (Du & Stillman 2002; Michael & Oren 2002). Knockdown of fibrillarin in HeLa cells using an RNAi technique had no effect on nuclear envelope assembly, but it did cause abnormal (aberrant) nuclear morphology, decreased cell proliferation and reduced cell growth (Amin *et al*. 2007). Fibrillarin has also

been reported to stimulate HeLa cell growth, but high endogenous expression of a fibrillarin-like EST in the optic lobe of female shrimp showed a negative correlation with body weight (Tangprasittipap *et al*. 2010). This apparent difference in effect may result from the cell types involved, given that HeLa cells are tumor cells while the shrimp cells were not (Tangprasittipap *et al*. 2010).

7. Glyceradehyde-3-phosphate dehydrogenase

Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) is a glycolytic enzyme that converts glyceraldehyde-3-phosphate reversibly to 1,3-bisphosphoglycerate by coupling with the reduction of NAD+ to NADH (Sirover 2011; Tristan *et al*. 2011). GAPDH was once considered to be a simple housekeeping protein but has recently been shown to be involved in many cellular processes in addition to glycolysis (Sirover 2011; Tristan *et al*. 2011). The multifunctional role of GAPDH has also been expanded to include; DNA repair (Meyer-Siegler *et al*. 1991), tRNA export (Singh & Green 1993), membrane fusion and transport (Tisdale 2001), cytoskeletal dynamics (Kumagai & Sakai 1983), and cell death (Hara *et al*. 2005). While GAPDH is used commonly as a control gene for gene expression profiling studies in crustaceans (Dhar *et al*. 2009), Wade *et al*. (2012) suggested that it should not be used for normalization of gene expression in prawns because GAPDH showed a strong spike in expression during the pre-moult stage in a colour adaptation response experiment. In *L. vannamei* populations, SNPs were identified in the GAPDH gene and some showed potential associations with weight gain, survival, and pathogen resistance (Ciobanu *et al*. 2010a). Such diverse GAPDH functions may be important in complex interactions associated with growth, disease and reproduction in crustacean species.

8. Growth hormone and insulin-like growth factor

The somatotrophic axis and transforming growth factor superfamily have been often targeted as possible candidate genes because of their key physiological role in growth regulation, development, metabolism, and lactation in model taxa (De Santis & Jerry 2007; Lucy 2008; Duan *et al*. 2010). The somatotropic axis essentially consists of growth hormone-releasing hormone (GHRH), growth hormone inhibiting hormone (GHIH or somatostatin), growth hormone (GH), insulin-like growth factors (IGF-I and -II), and associated carrier proteins and receptors. In general, the insulin signalling pathway can modulate uptake of glucose, fatty acids and amino acids into adipose tissue, muscle and liver, and promotes storage of nutrients in the form of glycogen, lipid and protein, respectively (Hwa & Rosenfeld 1999; Adam *et al*.

2005). Evidence has accumulated that components of the IGF axis, IGFs, IGFRs, and a superfamily of IGF binding proteins (IGFBPs), act together to control a number of crucial biological processes including cellular growth, cell proliferation, cell differentiation, survival from apoptosis and cell migration (Khandwala *et al*. 2000; Pollak *et al*. 2004). These processes are related to tissue formation and remodelling, bone growth, brain development and energy metabolism, which all ultimately influence organism size and affect individual longevity. Together they are believed to stimulate anabolic processes including cell proliferation, skeletal growth and protein synthesis (De Santis & Jerry 2007). The range and roles of the many different IGFs and GFs, and their receptors in vertebrates have been addressed extensively in a number of studies [See Hwa & Rosenfeld (1999), Adam *et al*. (2005), Jiang *et al*. (2007), Ola *et al*. (2008), Kocour & Kohlmann (2011), and Mullen *et al*. (2011)].

The action of GH is mediated by the transmembrane GH receptor (GHR) (Kopchick & Andry 2000). Hence, GH and GHR genes have been considered as potential targets for studies of genetic variation related to growth traits. GH effects on growth have been observed in a number of livestock species (Gupta *et al*. 2009; Tahmoorespur *et al*. 2011; Zhang *et al*. 2011a; Zhao *et al*. 2011) and variation in GHR gene sequences have been associated with a number of performance traits particularly in cattle (Viitala *et al*. 2006; Garrett *et al*. 2008; Waters *et al*. 2010) including milk yield (Blott *et al*. 2003; Viitala *et al*. 2006), as well as feed intake, feed conversation efficiency and body energy traits (Banos *et al*. 2008). The potential effect of GH and GHR genes on growth traits have also been reported in fish (Tao & Boulding 2003; Reinecke *et al*. 2005) with the majority of polymorphisms identified in introns (De Santis & Jerry 2007).

While putative GH and GHR genes have been reported in crustaceans (Toullec & Van Wormhoudt 1987), there is a lack of direct evidence for their role in promoting growth. Subsequently, pratices that involve direct administration of vertebrate GH to promote crustaceans' growth have become commonplace (Arenal *et al*. 2008; Santiesteban *et al*. 2010). Studies have shown that administration of vertebrate GH in crustaceans can promote growth (Charmantier *et al*. 1989; Xu *et al*. 2000; Sun *et al*. 2005; Yazawa *et al*. 2005; Arenal *et al*. 2008). An investigation of enhancement of growth in *L. vannammei* using recombinant tilapia GH showed that shrimp postlarvae treated with tilapia GH were 42.4% heavier and 5.2% longer than controls (Santiesteban *et al*. 2010). This suggests the presence of GH receptor-like proteins that recognise and mediate GH activity. While a number of ecdysteroidal neuropeptides have been reported in crustaceans including MIH and CHH, to date little work

has been done to explore their pathways or identity. Clearly much more work is needed to elucidate these pathways and to identify gene homologues in crustaceans.

The insulin-like growth factor (IGF) system is a complex network of peptide hormones (IGF1 and IGF2), cell surface receptors and circulating binding proteins. As a promising candidate gene for marker-assisted selection of growth traits, IGF1 and IGF2 are well documented in mammalian species and fish. Several genetic polymorphisms in the IGF1 gene associated with growth traits have been reported in chickens (Zhou *et al*. 2005; Bennett *et al*. 2006), in swine (Estany *et al*. 2007; Hao *et al*. 2011),bovines (Curi *et al*. 2005; Mullen *et al*. 2011) and in fish (Li *et al*. 2009). These studies have suggested strong additive genetic control of IGF1, a gene that is regulated at both the transcriptional and translational level (Wang *et al*. 2003; Mullen *et al*. 2011). There is also growing interest in the role(s) that IGF2, a peptide hormone well known for its mitogenic activity that can stimulate cell growth and differentiation in many tissues (Greene *et al*. 1997), plays in livestock and fish growth. A series of studies in livestock and fish have shown that DNA sequence polymorphisms within the IGF2 gene contribute to variation in complex production traits notably in muscle mass, fat deposition, carcass traits, body weight, fertility, and survival traits in pigs (Van Laere *et al*. 2003; Hou *et al*. 2010), bovines (Sherman *et al*. 2008; Berkowicz *et al*. 2011), and in fish (Juhua *et al*. 2010; Li *et al*. 2012).

Evidence for the GH and IGF axis has also accumulated in invertebrates that contain peptides that share substantial sequence homologies and/or biological actions with mammalian genes and presence of insulin, IGFs, IGFRs, and, IGFBPs. Single insulin binding domain protein (SIBD) has been identified in invertebrates including in some crustacean species (Garofalo 2002; Okuno *et al*. 2002; Wu & Brown 2006; Castellanos *et al*. 2008; Gai *et al*. 2010). In a study that attempted to elucidate the functional significance of insulin-like peptides in the white shrimp *P*. *vannamei*, juveniles injected with a single dose of recombinant human IGF-1 or bovine insulin at the intermoult stage showed a significant elevation in glucose and glycogen levels in the gills and hemolymph (Gutiérrez *et al*. 2007). This study also suggested the presence of endogenous *Penaeus* insulin(s) that, just like their vertebrate counterparts, were likely to be involved in regulation of carbohydrate metabolism. Recently, a SIBD, assumed to be a member of the IGFB superfamily, was identified in a haemocyte cDNA library from the shrimp *L*. *vannamei* and it was reported that it may be involved in the shrimp immune response after bacterial challenge (Castellanos *et al*. 2008). Another putative SIBD, associated with the endocrine and immune systems, has also been identified at the mRNA level in Chinese mitten crab *E*. *sinensis* (Gai *et al*. 2010), while there

is an interesting report of putative insulin-like peptides in androgenic gland (AG) of crustaceans suggesting a possible role in controlling sex differentiation (Manor *et al*. 2007; Ventura *et al*. 2011). A number of studies have indicated that AG can play an important role in sexual differentiation, development of the appendix masculinae and secondary male sex characteristics in crustaceans [See Manor *et al*. (2007), Ventura *et al*. (2009), Rosen *et al*. (2010), Mareddy *et al*. (2011), and Ventura *et al*. (2011)]. Much more work will be required to understand the effect(s) of IGFs on crustacean growth traits and to elucidate how insulin-like peptides regulate sexual shifts (e.g. appropriate intervention) in crustaceans because sexually dimorphic growth between genders has been reported in a wide array of malacostracans (Gitterle *et al*. 2005; Buřič *et al*. 2010; Gopal *et al*. 2010).

9. Myostatin and growth differentiation factor 8/11

Myostatin (MSTN), also known as growth differentiation factor-8 (GDF-8), is a critical autocrine/paracrine inhibitor of skeletal muscle growth and is a member of the transforming growth factor-β (TGF-β) superfamily (McPherron *et al*. 1997). GDF-11 is also a closely related member of this superfamily that is thought to have originated from a single ancestral gene following gene duplication (Xu *et al*. 2003). In vertebrates, MSTN, principally controls growth of muscle cells as a negative regulator of muscle development (MePherron *et al*. 1997), and GDF-11, regulates neurogenesis in the olfactory epithelium as well as development of the axial skeleton (McPherron *et al*. 1999; Wu *et al*. 2003). Both are inhibitors of specific cellular functions. Due to its important role, MSTN has become the best known example of an economically important gene. It has been studied intensively and its roles have been reviewed widely [See Lee (2004), Dominique & Gérard (2006), Rodgers & Garikipati (2008), and Huang *et al*. (2011b)].

The importance of MSTN in regulating phenotypic growth was highlighted initially in mice, where null mutations at the MSTN locus resulted in up to 30% faster growth rate (McPherron *et al*. 1997). Interest in this molecule associated with growth differences has since been extended to a number of other species. Investigation of MSTN developmental expression in bovine skeletal muscle (Shibata *et al*. 2003) and its function in myogenesis and adipogenesis showed MSTN expression to be involved directly with growth (Lin *et al*. 2002; Joulia *et al*. 2003; Rebbapragada *et al*. 2003; Wagner *et al*. 2005). In cattle, null mutations that are present naturally in the MSTN gene have been associated with the double-muscling phenotype (McPherron & Lee 1997). A SNP identified in the inhibitory domain of propeptide has also been associated with a moderate increase in silverside percentage, eye muscle area,

and total meat percentage (Sellick *et al*. 2007). In chickens, positive correlations have been observed between specific SNPs within the MSTN gene and production traits including abdominal fat percentage, birth weight, breast muscle percentage and breast muscle weight (Gu *et al*. 2004; Gill *et al*. 2009; Zhang *et al*. 2011b). Several mutations have also been identified that may contribute to breast meat and abdominal fat percentage in ducks (Lu *et al*. 2011) and for growth and relative fat content in sheep (Kijas *et al*. 2007; Hickford *et al*. 2009). Although a specific role for MSTN in skeletal muscle is confined to mammalian models, the major functions of MSTN, GDF-8, and GDF-11 in lower vertebrates including fish are also conserved (Medeiros *et al*. 2009; Stinckens *et al*. 2010). In contrast to the pattern in mammals where MSTN is primarily limited to skeletal muscle, MSTN occurs in many fish tissues including muscle, brain, eye, stomach, skin, testis, ovary, kidney, intestine, gill, liver, spleen, and heart (Pan *et al*. 2007; Ye *et al*. 2007; De Santis *et al*. 2008), with expression levels varying with developmental stage (Delgado *et al*. 2008). Ubiquitous expression suggests that functions of MSTN in fish could be diverse (De Santis *et al*. 2008; Nadjar-Boger & Funkenstein 2011).

MSTN-like genes have also been identified in arthropods and molluscs and this has prompted new investigations into their functions, in particular with regard to their ability to regulate muscle growth in invertebrates. cDNAs encoding MSTN-like proteins have been characterized from bay scallop, *Argopecten irradians* (Kim *et al*. 2004a), black land crab, *Gecarcinus lateralis* (Covi *et al*. 2008), Chinese mitten crab, *E. sinensis* (Kim *et al*. 2009b), and Morotoge shrimp, *Pandalopsis japonica* (Kim *et al*. 2010). These studies have indicated that the MSTN gene has been conserved across evolution and that it may also play a major role in muscle growth and development in invertebrates, as it does in mammals. Two non-synonymous SNPs have been found in the MSTN gene in Japanese scallop (*Chlamys farrei*) that showed significant associations with growth traits (Wang *et al*. 2010). In adult scallops, MSTN is predominately expressed in striated muscle, with different expression levels observed in other tissues, this result suggests that the function of MSTN may not only be restricted to muscle growth and development, but that it may also be involved in other developmental and physiological processes (Hu *et al*. 2010). Since MSTN is a relatively specialized member of a diverse group of cytokines (the TGF-β superfamily), that are involved in developmental patterning, organogenesis and tissue homeostasis (Tsuchida *et al*. 2006), MSTN's major role(s) in crustaceans will need to be more carefully considered with regard to the moulting event because skeletal muscle in crustaceans is highly plastic and is able to undergo dramatic remodelling during development and growth (Mykles 1997). Recent

studies in land crab and American lobster demonstrated that MSTN turnover has roles to play in muscle atrophy during moulting where it remodels the contractile apparatus (Covi *et al*. 2008; MacLea *et al*. 2010). In a giant tiger shrimp (*P. monodon*), reduced levels of MSTN and GDF-11 transcripts resulted in a dramatic slowing of growth rate compared with control groups (De Santis *et al*. 2011). Unlike in vertebrates where MSTN and GDF-11 show negative regulation of growth, expression of these two genes may regulate growth positively in crustaceans (De Santis *et al*. 2011). Hence, further characterisation of MSTN and GDF genes in crustaceans could provide valuable gene targets for application in future stock improvement programs.

10. Signal transducer and activator of transcription

The Janus kinase (Jak) signal transducer and activator of transcription (STAT) pathway is one of the main signalling pathways in eukaryotic cells (Dearolf 1999; Hou *et al*. 2002; Arbouzova & Zeidler 2006). This pathway is used in a variety of growth and developmental processes in multiple tissues to control cell; proliferation, differentiation, survival, and apoptosis (Luo & Dearolf 2001; Hou *et al*. 2002; Arbouzova & Zeidler 2006). Seven STAT proteins (STAT1-4, 5A, 5B, and 6) have been identified in mammals (Li 2008) but to date, only a single STAT gene has been found in invertebrates (Barillas-Mury *et al*. 1999; Chen *et al*. 2008). In crustaceans STAT sequences have been cloned from both *P. monodon* and *F. chinensis* (Chen *et al*. 2008; Sun *et al*. 2011a).

Evidence is also mounting for a role of the Jak/STAT pathway in antiviral immunity in crustaceans and insects, (Luo & Dearolf 2001; Agaisse & Perrimon 2004; Chen *et al*. 2008; Flegel & Sritunyalucksana 2011) as well as during embryonic development (Cheng *et al*. 2010) and eye development (Wang & Huang 2010). One study attempted to find SNPs in the STAT gene from Pacific white shrimp that were associated with individual body weight. Studied animals, however, were all homozygous, so potentially larger and more diverse populations may need to be examined to confirm a functional role (Martin-Marti *et al*. 2010).

11. Secreted protein acidic and rich in cysteine

Secreted protein acidic and rich in cysteine (SPARC), also known as osteonectin or basement membrane protein 40 (BM40), is a collagen-binding, counter-adhesive protein belonging to the matricellular group of proteins (McCurdy *et al*. 2010). SPARC is conserved in a wide variety of evolutionarily diverse organisms, suggesting a fundamental role for SPARC in multicellular organisms (Bradshaw & Sage 2001; Tanaka *et al*. 2001; Rotllant *et al*. 2008;

Bradshaw 2009; McCurdy *et al*. 2010). SPARC participates in a wide range of biological processes (development, tissue repair, and tissue remodelling) and several physiological functions (assembly and organization of extracellular matrix, modulating multiple intracellular signalling pathways, cell migration, adhesion, proliferation and differentiation) and has been reviewed extensively [See Termine *et al*. (1981), Sage & Bornstein (1991), Bradshaw & Sage (2001), Murphy-Ullrich (2001), Motamed *et al*. (2003), Bornstein (2009), Bradshaw (2009), and Chlenski & Cohn (2010)]. In SPARC-deficient mammals, there is a failure to generate a robust fibrotic response to a number of different stimuli (Bradshaw 2009). The infarct scar formed in SPARC-null mice exhibited a higher incidence of immature collagen fibres than did scars from wildtype mice suggesting SPARC was associated with collagen concentration, muscle stiffness and cell proliferation (Bradshaw *et al*. 2009; Harris *et al*. 2011). In crustaceans, comparisons of different size variants in shrimp showed a higher expression of SPARC-like mRNA in the optic lobe of small compared with large individuals. A statistically significant negative correlation was evident with shrimp body weight after RT-PCR expression analysis (Tangprasittipap *et al*. 2010). A possible physiological role for exogenous SPARC is as a modulator of cell shape, and as an inhibitor of cell adhesion and spreading (Sage *et al*. 1989; Lane & Sage 1990) in cell cycle progression and high expression of a SPARC-like EST in small shrimp individuals could inhibit cell progression from the G1 to the S stage (Tangprasittipap *et al*. 2010). This hypothesis is consistent with previous reports in model species (Funk & Sage 1991; Barker *et al*. 2005). Therefore, understanding the precise mechanisms that govern SPARC activity in extracellular matrix formation via modulation receptor binding may provide novel and exciting ways of better understanding gene regulation of growth in crustaceans.

12. Translin-associated factor X

Translin and its binding partner protein, translinassociated factor-X (TRAX), are components of an evolutionarily conserved RNA binding complex, which has been implicated in a broad spectrum of biological activities including cell growth regulation, mRNA processing, spermatogenesis, neuronal development/function, genome stability regulation and carcinogenesis (Li *et al*. 2008; Jaendling & McFarlane 2010). Furthermore, translin (with or without TRAX) has nucleic-acid-binding activity and it is apparent that control of nucleic acid metabolism and distribution are central biological roles for this protein and its partner TRAX (Jaendling & McFarlane 2010). Schröer *et al*. (2007) suggested that TRAX may be a molecular switch for growth-associated protein (GAP)-43 that controls axonal regeneration in

rats. In *L. vannamei* populations, haplotype SNPs were located in TRAX and they showed some associations with weight gain, survival, and pathogen resistance (Ciobanu *et al*. 2010a).

CANDIATE GROWTH GENES EFFECT ON MOULTING

Neuropeptides are the largest and most diverse group of endocrine signalling molecules present in the crustacean nervous system (Lago-Lestón *et al*. 2007). Signalling peptides are critical for initiation and regulation of a diverse array of physiological processes including feeding, reproduction and development (Schwartz *et al*. 2000; Sweedler *et al*. 2002). In crustaceans, periodic shedding of the exoskeleton is one of the most important physiological processes essential for crustacean growth and postembryonic development including moulting and regeneration. This event is induced by synthesis of ecdysteroids (Skinner 1985; Lachaise *et al*. 1993). Although the functions of many of the hormones and genes involved in this process are still not well defined, ecdysteroids are believed to be produced by paired Y-organs (YO) in crustaceans. YOs are ectodermally-derived endocrine glands located in the cephalothorax (Skinner 1985; Lachaise *et al*. 1993). They are negatively regulated by two neuropeptides; moult-inhibiting hormone (MIH) and crustacean hyperglycaemic hormone (CHH) that are produced in the X-organ sinus gland complex (XOSG) located in the optic ganglia of the eyestalk (Skinner 1985; Lachaise *et al*. 1993). A number of physiological processes in crustaceans are known to be regulated by diverse neuropeptides synthesized by the XOSG (De Kleijn & Van Herp 1995). A number of excellent reviews of crustacean neuroendocrinology, ecdysteroids, MIH, and CHH describe regulation of moulting [See Chang *et al*. (2001), Covi *et al*. (2009), Nakatsuji *et al*. (2009), Chang & Mykles (2011), Hopkins (2012), and Webster *et al*. (2012)]. While a large number of hormones and genes are known to be involved in crustacean moulting events, here we will confine the discussion to aspects of putative crustacean growth genes involved in moulting and later we will briefly address candidate genes in relation to muscle development and/or muscle degradation during moulting.

13. Actin

Actins are highly conserved structural proteins that are expressed ubiquitously in eukaryotic cells. They constitute the most abundant proteins in a cell and comprise up to 10% of all cellular protein and they play key roles in a number of cellular functions in the organism

including muscle contraction, cell motility, cell division and cytokinesis, vesicle and organelle movement, cell signaling, establishment and maintenance of cell junctions and cell shape, and phagocytosis (Pollard & Cooper, 1986; De Loof *et al*. 1996; Cadoret *et al*. 1999). Actin filaments provide strength, connections to other cells and the extracellular matrix, paths for intracellular transport, and act as a scaffold for generating force (De Loof *et al*. 1996; Sutherland & Witke 1999). Actins comprise a multigene family with diverse genomic organization and expression patterns that vary with species, tissue, and developmental stage (Hooper & Thuma 2005). They have been categorized broadly into two classes according to their N-terminal sequences and specific tissue distributions that are referred to as either cytoplasmic or muscle actins (Hightower & Meagher 1986; Herman 1993). Most actins in invertebrates and plants belong to the cytoplasmic group (Zhu *et al*. 2005). To date, more than 20 actins have been identified in crustaceans (crabs, crayfish, lobsters, prawns, and shrimp) (Macias & Sastre 1990; Varadaraj *et al*. 1996; Zhu *et al*. 2005; Sun *et al*. 2007; Kim *et al*. 2009a). In a recent SNP association analysis study, four synonymous polymorphisms were identified in an actin fragment but SNP allele distributions were not related significantly to individual growth performance in the two studied groups of giant freshwater prawn *M. rosenbergii* (Thanh *et al*. 2010). As evidence for a role for actin in muscle build-up during the moult cycle in crustaceans, Cesar and Yang (2007) reported that muscle structural α-actin and cytoskeletal β-actin increased during the intermoult and premoult stages, a phase where high muscle growth occurred in the abdominal muscle of *L. vannamei*. Since crustacean muscle protein synthesis and degradation is modulated during each moult cycle, further studies will be required to investigate regulation of structural and regulatory proteins in the muscles of crustaceans.

14. Crustacean hyperglycaemic hormone

CHH belongs to a peptide family, the CHH family, which also includes MIH, mandibular organ-inhibiting hormone (MOIH), gonad/vitellogenesis-inhibiting hormone (G/VIH), and insect ion transport peptide (ITP) (Chang 2001; Chen *et al*. 2005). CHH, is one of a unique group of neuropeptides in arthropods and is the most abundant neuroendocrine, synthesized by the X-organ (XO) and is stored in the sinus gland (SG) prior to being released directly into the hemolymph (Böcking *et al*. 2002; Fanjul-Moles 2006). A number of studies have indicated that moulting and reproduction in crustaceans is regulated by the eyestalk derived CHH gene family which is one of the major groups of peptide hormones produced in the

XOSG and a wide range of biochemical, functional, structural, and evolutionary aspects of the CHH family have been discussed and reviewed [See Fanjul-Moles (2006), Chung *et al*. (2010), Montagné *et al*. (2010), Chang & Mykles (2011), Hopkins (2012), and Webster *et al*. (2012)]. CHH is involved primarily in regulation of glucose levels in the hemolymph, as well as in metabolism of carbohydrates and lipids. In addition, it has a major role in moulting (De Kleijn & Van Herp 1995), reproduction (De Kleijn & Van Herp 1998), and osmoregulation (Serrano et al. 2003). In a study that assessed correlations between SNPs in the CHH gene with individual growth performance in giant freshwater prawn *M. rosenbergii*, four intronic SNPs showed positive associations with three growth traits (body weight, carapace length, and standard length) (Thanh *et al*. 2010). A second haplotype-trait association analysis confirmed that the four SNPs were in linkage disequilibrium, and that a single haplotype showed a significant association with fast growth (Thanh *et al*. 2010). CHH gene has high potential to impact body weight variation in crustaceans and should, therefore, be considered as a primary gene of interest in growth studies.

15. Eyestalk factors

Eyestalks in crustaceans contain neurosecretory cells that have been considered to be involved in regulation of moulting, and this structure can affect ecdysteroid secretion from the YO (Hopkins 2012). Since moulting is directly related to muscle development and growth in crustaceans, experimental approaches have been trialled to affect growth rate and eyestalk ablation appears to be the most effective way of influencing moulting and growth because it affects the endocrine system directly (Chen *et al*. 1995). Several studies have also shown that eyestalk factors (removal of eyestalk) can impact individual growth rates (Chen *et al*. 1995; Venkitraman *et al*. 2004; Allayie *et al*. 2011). Other studies have reported that eyestalk ablation can also be used to shorten the moult interval and to stimulate gonad development in crustaceans (Okumura & Katsumi 2001; Venkitraman *et al*. 2004). Applying bilateral eyestalk ablation, however, should be approached cautiously since it can cause mortality (Chen *et al*. 1993; Chen *et al*. 1995), and is not considered realistic at an industrial scale. Characterisation of the gene products expressed within the eyestalk, however, may yield important gene candidates for future studies.

16. Moult inhibiting hormone

MIH is an important regulator of steroidogenesis in the YO, and this hormone is produced by neurosecretory cell somata located at the medulla terminalis of the XO in the eyestalks and MIG is transported to the neurohemal SG, after which it is released into the hemolymph (Skinner 1985; Lachaise *et al*. 1993; Watson *et al*. 2001). MIH is responsible for maintaining animals in the intermoult stage, while both MIH and CHH are involved in regulation of the length of the intermoult period (Gu *et al*. 2000; Gu *et al*. 2002; Yodmuang *et al*. 2004).

SNPs have been identified in MIH genes (MIH1 & MIG2) in penaeid prawns (Yu *et al*. 2006; Li *et al*. 2011b), and in a study designed to assess the correlation between SNPs in MIH genes with individual growth performance in white shrimp *L. vannamei*, two exonic SNPs were reported and a SNP in MIH exon-1 showed significant association with body weight (Li *et al*. 2011b). Analysis of larger populations would be useful to estimate more precisely the effect of individual SNPs on production traits in *P. monodon*. Apart from growth performance, MIH has also been identified as a key endocrine regulator that co-ordinates moulting and reproduction in mature female blue crab *C. sapidus*. MIH in blue crab inhibits moult and stimulates vitellogenesis simultaneously (Zmora *et al*. 2009a). MIH can also regulate vitellogenesis and this is associated with appearance of MIH specific membrane binding sites in the hepatopancreas after the pubertal/final moult (Zmora *et al*. 2009b). Therefore, mutations in the MIH gene are prime targets as candidate genes that impact growth performance and reproduction in crustaceans.

CANDIDATE MUSCLE BUILD-UP OR DEGRADATION GENES INVOLVED IN MOULTING

Crustacean muscle growth is not continuous and is strongly influenced by the moulting cycle. During the moult, muscles regenerate, and energy reserves including glycogen and lipids are accumulated in the hemolymph and the midgut for the next moult (Devaraj & Natarajan 2006; Kuballa & Elizur 2007). Overall muscle protein synthesis is very important for growth, reproduction and other metabolic activities in crustaceans. Recent studies of invertebrates have highlighted the importance of muscle specific genes and proteins in crustaceans (Hooper & Thuma 2005; Hooper *et al*. 2008; Jung *et al*. 2011). While several genes and hormones are

actually involved in a complex moulting process, here we will focus on genes related to muscle build-up or degradation during the moulting event.

17. Methyl farnesoate and farneosoic acid *O*-methyltransferase

Methyl farnesoate (MF), the immediate precursor of insect juvenile hormone III (JHIII), is a sesquiterpenoid that is synthesised in the crustacean mandibular organ (MO) (Tamone *et al*. 1997). MF potentially regulates juvenile characteristics in crustaceans in a manner similar to JHIII in insects (Borst *et al*. 1987; Laufer *et al*. 1987; Borst *et al*. 2001; Burtenshaw *et al*. 2008). Recent studies have suggested that sequiterpenoid compounds may serve an evolutionarily conserved role as regulators or modulators of moulting, reproduction, larval development, growth, morphogenesis, behaviour, osmotic stress, and general protein synthesis in crustaceans (Nagaraju 2007; Li *et al*. 2010; Tiu *et al*. 2012). Recent articles by Williamson *et al*. (2001), Holford *et al*. (2004), and Kuballa *et al*. (2007) provide detailed information regarding this biochemical pathway. Widespread activity of farnesoic acid *O*-methyltransferase (FAMeT) suggests that it may have a broad spectrum of action in many tissues that contribute to the function and regulation of MF synthesis in crustaceans (Silva Gunawardene *et al*. 2003). The potential role of MF and FAMeT as a morphogen has been studied in several crustacean species. When MF was administered to *M. rosenbergii*, higher MF levels caused earlier retardation of late larval growth, and at the highest dose provided, it retarded larval development and affected patterns of metamorphosis (Abdu *et al*. 1998). In addition, MF and ecdysteroids have been shown to play a role in controlling morphogenesis in allometric growth in spider crab, *Libinia emarginata* (Laufer *et al*. 2002). As further evidence for a role for FAMeT in growth and moulting, it was found to be highly up-regulated in the intermoult stage during the moult cycle in blue swimmer crab, *P. pelagicus* (Kuballa & Elizur 2007). In addition, RNA interference (RNAi) studies indicate that shrimps injected with double stranded RNA (dsRNA) for FAMeT knock-down, did not advance to the final stage of the moult cycle and showed 100% mortality while control shrimps completed their moult (Hui *et al*. 2008). Furthermore, expression of moult-related genes encoding cathepsin-L and the hemocyanin gene were disrupted (Hui *et al*. 2008). While to date no studies have attempted to associate genetic variation in these genes with economically-important traits including growth, these genes represent important targets for future study.

18. Heat shock proteins

Heat shock proteins (HSPs) are ubiquitous and are the most abundant and highly conserved multigene superfamily found in all organisms (Kregel 2002). HSPs have received great attention for their roles in protein chaperoning (Gusev *et al*. 2005; Mayer & Bukau 2005), developmental processes (Christians *et al*. 2003) and tolerance mechanisms to chemical, physical and pathophysiological stress (Macario & Conway de Macario 2007; Rinehart *et al*. 2007). A variety of HSP forms are also involved in many essential cellular functions, including metabolism, growth, differentiation and programmed cell death (Ranford *et al*. 2000; Christians *et al*. 2003; Multhoff 2006). In crustaceans, a number of studies of HSP60 and HSP70 have been conducted in relation to innate immune responses and environmental stress (Rungrassamee *et al*. 2010; Zhou *et al*. 2010; Huang *et al*. 2011a). As evidence for their involvement in crustacean muscle atrophy, heat shock cognate 70 (HSC70) and HSP90 mRNA levels were significantly induced in the premoult relative to the intermoult phase in American lobster claw muscle (Spees *et al*. 2003). In a study of the marine shrimp *L. vannamei*, muscle atrophic protein HSP70 expression levels were not significantly changed during the premoult stages, suggesting that muscle protein degradation was reduced in abdominal muscle over the moult cycle, a stage that is typically associated with claw muscle degradation (Cesar & Yang 2007).

19. Myosin heavy chain

Myosins are a major component of the contractile apparatus and consist of two heavy and four associated light chains (Harrington & Rogers 1984). As a large superfamily of proteins, myosins share a common domain that has been shown to interact with actin, to hydrolyze ATP and to produce movement including the ability of muscle to develop and maintain force (Gauvry *et al*. 1997; Mermall *et al*. 1998; Sellers 2000). Myosin isoforms are expressed ubiquitously in all eukaryotic cells and are the most abundant contractile protein present in skeletal muscle (Denardi *et al*. 1993; Jung *et al*. 1998). Basic skeletal muscle fibre types that power locomotion differ significantly in their contractile properties (Perry *et al*. 2009), largely because of the differences in myofibrillar protein isoforms present among the various fibres (Schiaffino & Reggiani 1996). Using a similar approach, crustacean muscles have been grouped based on their structural criteria, ATPase histochemistry and the specific assemblage of myofibrillar isoforms present in a particular fibre type (Silverman *et al*. 1987; Mykles

1988, 1997; L_AFramboisea *et al*. 2000). As a consequence, crustacean muscles can be classified as fast, slow twitch or phasic (S_1), or slow tonic (S_2) muscle types. More detailed information about biochemical features of crustacean muscle fibres are provided by Silverman *et al*. (1987), Neil *et al*. (1993), Mykles (1997), and Medler *et al*. (2007a, b).

MHC isoforms have received more attention than other myofibrillar proteins, as differences in shortening velocity are directly correlated with rate of ATP hydrolysis by the myosin head (Schiaffino & Reggiani 1996). As such, this type of MHC is considered to be the primary determinant of contractile properties among different muscle types (Schiaffino & Reggiani 1996). Morphological changes and biochemical composition of abdominal and skeletal muscles in crustaceans over the moult cycle have been observed and expression of actin and myosin transcripts suggest that muscle fibres are rearranged as part of various moult stages (Medler & Mykles 2003; De Oliveira Cesar *et al*. 2006). In addition, during myogenesis in particular, MHC binds strongly to muscle precursor cells in the thoracic limbs of American lobster (Harzsch & Kreissl 2010) and in isopods (Kreissl *et al*. 2008). This results in muscle development in crayfish (Jirikowski *et al*. 2010). If differences in weight and size are related to production of muscle tissue, it is likely that myosin will be one of the genes expressed (Kamimura *et al*. 2008). In south-western Atlantic pink shrimp *Farfantepenaeus paulensis*, higher MHC expression was observed in a high weight shrimp group a result that identified MHC as a possible growth candidate gene (Kamimura *et al*. 2008).

20. Ubiquitin

Ubiquitins (Ubs) are highly conserved proteins that are distributed widely in eukaryotic cells and that are linked to a vast range of proteins (Ciechanover 1998; Yamao 1999). Ubs are normally ligated to other proteins in the cell as a monomer or polymer to regulate their activity and/or entry into proteasomal and other degradation pathways (Welchman *et al*. 2005). Selective protein degradation is mainly carried out by the Ub system that plays important roles in many cellular functions, including immune regulation, cell cycle control, signal transduction, chromatin remodelling, membrane trafficking, transcriptional regulation, the nuclear transport process, and membrane receptor control by endocytosis (Pickart *et al*. 2004; Bai *et al*. 2007; Dantuma *et al*. 2006; Patterson, 2006; Mabb & Ehlers 2010). As has been reported for mammalian skeletal muscle where increased Ub expression occurs during muscle atrophy (Fang *et al*. 2000), they have also been observed in crustaceans during muscle

atrophy associated with the moulting event. Shean and Mykles (1995) reported that Ub expression levels were elevated in lobster premoult claw muscle but this phenomenon was absent in lobster and shrimp abdominal muscles where premoult atrophy is not significant (Koenders *et al*. 2002; Cesar & Yang 2007). Minor variation in total wet weight, total soluble proteins, and myofibrillar were also observed in cross sections of abdominal muscle during the premoult stages (Ceasar & Yang 2007). According to Mykles (1997), degradation of myofilaments does not occur in walking leg muscles and lobster abdominal muscle once muscles have fitted through their designated openings in the exoskeleton. Increased Ub mRNA in atrophic muscle in lobster and land crab large claws could suggest that additional Ub is required for muscle protein degradation (Koenders *et al*. 2002). Polyubiquitin mRNA levels in claw muscle were also elevated in premoult individuals compared with intermoult individuals and were significantly elevated relative to premoult abdominal muscle in lobster (Spees *et al*. 2003). Ubiquitination has also been shown to play a role in muscle-specific adult-stage abdominal atrophy in an insect, where selective repression of actin and MHC genes also occurs (Haas *et al*. 1995).

CANDIDATE GENES INDENTIFIED IN LIVESTOCK THAT ARE ALSO PRESENT IN CRUSTACEANS

Genetic improvement of livestock has been practiced for centuries to improve production efficiency and more recently genomic and proteomic approaches have been trialled. While new '-omic' technologies can be applied to a wide range of species, to date application has been confined to relatively few species because of limited available genome information. Crustacean species currently fall into this group. Therefore, many putative genes, hormones and enzymes that affect growth in livestock may have homologues in crustaceans and so may be important candidates for study. Recently a transcriptome dataset was generated for *M. rosenbergii* using 454 GS-FLX (Jung *et al*. 2011). After this a comparative genomic approach was applied to identifying putative candidate genes in *M. rosenbergii* based on genomic data from livestock. Selected putative genes identified in *M. rosenbergii* are summarized in Table 4. There are a number of noteworthy gene candidates in this list including calpain and TGF-β receptors. In addition to selected genes, physiologically and functionally important genes can be explored in more detail by applying a reliable comparative genomic framework. Some genes have been conserved through evolution from bacteria to humans but some have evolved differently and have unique biological functions.

In the case of crustaceans, genomic comparisons with insect genomic resources (e.g. tribolium, honeybee, pea aphid, and *drosophila*) are likely to be extremely useful for identifying candidate gene loci in crustaceans because a number of reviews have demonstrated close genetic relationships based on large-scale molecular and morphological datasets (Giribet & Edgecombe 2011; Von Reumont *et al*. 2012).

CONCLUSIONS

It is obvious that aquaculture must play a major role in the future in addressing increasing demand for aquatic protein given an uncertain future for wild fisheries. Genetic improvement of production species has already proven to be an important technology for addressing the global food crisis, but its role in aquaculture has to date been hampered by the limited genetic and genomic data available for many important production species. Recent technological advances in NGSs and bioinformatics have provided a major opportunity to understand the molecular basis of phenotypic variation in living species. With the growing analytical power of NSGs, the 1000 Genome Project (http://www.1000genomes.org/) has opened a new era of '-omics' for discovering target genes and functional mutations, and this will be further accelerated by the recent announcement of the 5000 Insect Genome Project (http://www.arthropodgenomes.org/wiki/i5k). These significant developments will also impact on crustacean biology, where progress has lagged behind that of many other groups (notably mammals).

While a small number of candidate genes have already been identified in crustacean species that affect growth, reproduction and immunity, information is still insufficient to apply it directly to MAS breeding programs. The question remains where we should search for candidate genes and which genes we should focus on to address MAS goals in crustacean species. Here we argue that one option is to identify candidate genes in crustaceans and other arthropod species that have been identified previously to have roles in growth and muscle development in other taxa. This can provide a platform for studying phenotypic variation because genes identified in related species via conservation through evolution often play similar or related roles in other target species. The second approach is to search for genes, hormones and enzymes that may be involved in unique physiological phenomena. In the case of crustaceans, moulting is a unique physiological event and is essential for normal growth, metamorphosis and reproduction. Identifying and understanding genes related to moult cycles and ones that affect individual growth responses will be a key development because it can

connect the physiological interactions of periodical growth and muscle development and degradation. The third approach is to search for homologous and/or orthologous sequences that have been identified as candidate genes affecting important growth phenotypes in public databases. A comparative genomic framework using data on insects and/or livestock species can provide a platform for rapidly targeting critical genes where little information is available currently for crustacean species. It will also be important to examine the roles of a number of different candidate genes because most growth related phenotypes are complex traits that commonly result from multi-gene, epistatic interactions.

The review work described here was undertaken employing these four principal search strategies. While many more putative genes involved in growth rate phenotypes are likely to be identified in the future, genes listed in this review including; 5-HT, Actin, α-AMY, CatL, CHH, Cyps, FABPs, GAPDH, HSPs, MF, MIH, MSTN, MYC, STAT, SPARC, and TRAX should be considered as primary targets for genes affecting growth phenotypes in crustacean species. While it would be impossible to cover all genes, hormones or enzymes that are likely to affect growth in crustacean species, the data presented in the current review can provide a foundation for new growth studies in crustaceans and NGS technological developments are likely to enhance the rate at which new genes are identified.

ACKNOWLEDGMENTS

Authors wish to thank to Dr. Woo-Jin Kim and two anonymous reviewers for useful comments. This research forms a part of Hyungtaek Jung's PhD project, and is supported by an International Postgraduate Research Scholarship (Australia) and a Queensland University of Technology Postgraduate Award (N7333798). Additional funding for this work was provided by Queensland University of Technology awarded to Peter B. Mather (VC NPSG award).

REFERENCES

Abdu U, Takac P, Laufer H, Sagi A (1998) Effect of methyl farnesoate on late larval development and metamorphosis in the prawn *Macrobrachium rosenbergii* (Decapoda, Palaemonidae): A juvenoid-like effect? *Biological Bulletin* **195**: 112–119.

Adam D, Cosgrove LJ, Booker GW, Wallace JC, Forbes BE (2005) Molecular interactions of the IGF system. *Cytokine and Growth Factor Review* **16**: 421–439.

Agaisse H, Perrimon N (2004) The roles of Jak/STAT signalling in *Drosophila* immune responses. *Immunological Reviews* **198**: 72–82.

Alexander LS, Qu A, Cutler SA, Mahajan A, Rothschild MF, Cai W *et al*. (2010) A calcitonin receptor (*CALCR*) single nucleotide polymorphism is associated with growth performance and bone integrity in response to dietary phosphorus deficiency. *Journal of Animal Science* **88**: 1009–1016.

Alipanah M, Kalashnikova L, Rodionov G (2007) Association of prolactin gene variants with milk production traits in Russian Red Pied cattle. *Iranian Journal of Biotechnology* **5**: 158–161.

Allais S, Journaux L, Levéziel H, Payet-Duprat N, Raynaud P, Hocquette JF *et al*. (2011) Effects of polymorphisms in the calpastatin and μ-calpain genes on meat tenderness in 3 French beef breeds. *Journal of Animal Science* **89**: 1–11.

Allayie SA, Ravichandran S, Bhat BA (2011) Hormonal regulatory role of eyestalk factors on growth of heart in mud crab, *Scylla serrata*. *Saudi Journal of Biological Sciences* **18**: 283–286.

Amin MA, Matsunaga S, Ma N, Takata H, Yokoyama M, Uchiyama S *et al*. (2007) Fibrillarin, a nucleolar protein, is required for normal nuclear morphology and cellular growth in HeLa cells. *Biochemical and Biophysical Research Communications* **360**: 320–326.

Aoki H, Ahsan MN, Watabe S (2004) Molecular and enzymatic properties of a cathepsin L-like proteinase with distinct substrate specificity from northern shrimp (*Pandalus borealis*). *Journal of Comparative Physiology* **174 B**: 59–69.

Arbouzova NI, Zeidler MP (2006) JakK/STAT signalling in *Drosophila*: insights into conserved regulatory and cellular functions. *Development* **133**: 2605–2616.

Arenal A, Pimentel R, Pimentel E, Martín L, Santiesteban D, Franco R *et al*. (2008) Growth enhancement of shrimp (*Litopenaeus schmitti*) after transfer of tilapia growth hormone gene. *Biotechnology Letters* **30**: 845–851.

Arlot-Bonnemains Y, Fouchereau-Peron M, Chesnais J, Taboulet J, Milhaud G, Moukhtar MS *et al*. (1991) Calcitonin gene-related peptide immunoreactivity in a crustacean *Nephrops norvegicus* and correlation with calcitonin. *General and Comparative Endocrinology* **83**: 1–6.

Asaro A, Del Valle JC, Mañanes AAL (2011) Amylase, maltase and sucrase activities in hepatopancreas of the euryhaline crab *Neohelice granulata* (Decapoda: Brachyura: Varunidae): partial characterization and response to low environmental salinity. *Scientia Marina* 75: 517–524.

Bai J, Wu Y, Li QW (2007) Protein degradation by ubiquitination pathway. *International Journal of Genetics* **30**: 114–118.

Banos G, Woolliams JA, Woodward BW, Forbes AB, Coffey MP (2008) Impact of single nucleotide polymorphisms in leptin, leptin receptor, growth hormone receptor, and diacylglycerol acyltransferase (DGAT1) gene loci on milk production, feed,

and body energy traits of UK dairy cows. *Journal of Dairy Science* **91**: 3190–3200.

Barillas-Mury C, Han YS, Seeley D, Kafatos FC (1999) *Anopheles gambiae* Ag-STAT, a new insect member of the STAT family, is activated in response to bacterial infection. *The EMBO Journal* **18**: 959–967.

Barker TH, Baneyx G, Cardo-Vila M, Workman GA, Weaver M, Menon PM *et al*. (2005) SPARC regulates extracellular matrix organization through its modulation of integrin-linked kinase activity. *Journal of Biological Chemistry* **280**: 36483–36493.

Bennett AK, Hester PY, Spurlock DE (2006) Polymorphisms in vitamin D receptor, osteoponin, insulin-like growth factor 1 and insulin, and their associations with bone, egg and growth traits in a layer-broiler cross in chickens. *Animal Genetics* **37**: 283–286.

Bennett AK, Hester PY, Spurlock DM (2007) Relationships of a transforming growth factor-β2 single nucleotide polymorphism and messenger ribonucleic acid abundance with bone and production traits in chickens. *Poultry Science* **86**: 829–834.

Berkowicz EW, Magee DA, Sikora KM, Berry DP, Howard DJ, Mullen MP *et al*. (2011) Single nucleotide polymorphisms at the imprinted bovine insulin-like growth factor 2 (IGF2) locus are associated with dairy performance in Irish Holstein-Friesian cattle. *Journal of Dairy Research* **78**: 1–8.

Berti PJ, Storer AC (1995) Alignment/phylogeny of the papain superfamily of cysteine proteinases. *Journal of Molecular Biology* **246**: 273–283.

Blott S, Kim JJ, Moisio S, Schmidt-Küntzel A, Cornet A, Berzi P *et al*. (2003) Molecular dissection of a quantitative trait locus: a phenylalanine-to-tyrosine substitution in the transmembrane domain of the bovine growth hormone receptor is associated with a major effect on milk yield and composition. *Genetics* **163**: 253–266.

Böcking D, Dirksen H, Keller R (2002) The crustacean neuropeptides of the CHH/MIH/GIH family: structures and biological activities. In: Wiese K, editor. *The crustacean nervous system*. pp. 84–97. Springer, Berlin.

Bolormaa S, Hayes BJ, Savin K, Hawken R, Barendse W, Arthur PF *et al*. (2011) Genome-wide association studies for feedlot and growth traits in cattle. *Journal of Animal Science* **89**: 1684–1697.

Bornstein P (2009) Matricellular proteins: an overview. *Journal of Cell Communication Signalling* **3**: 163–165.

Borst DW, Laufer H, Landau M, Chang ES, Hertz WA, Baker FC *et al*. (1987) Methyl farnesoate and its role in crustacean development. *Insect Biochemistry* **17**: 1123–1127.

Borst DW, Ogan JT, Tsukimura B, Claerhout T, Holford KC (2001) Regulation of the crustacean mandibular organ. *American Zoologist* **41**: 430–441.

Bradshaw AD, Sage EH (2001) SPARC, a matricellular protein that functions in cellular differentiation and tissue response to injury. *The Journal of Clinical Investigation* **107**: 1049–1054.

Bradshaw AD (2009) The role of SPARC in extracellular matrix assembly. *Journal of Cell Communication and Signalling* **3**: 239–246.

Bradshaw AD, Baicu CF, Rentz TJ, Van Laer AO, Boggs J, Lacy JM *et al*. (2009) Pressure overload-induced alterations in fibrillar collagen content and myocardial diastolic function: role of secreted protein acidic and rich in cysteine (SPARC) in post-synthetic procollagen processing. *Circulation* **119**: 269–280.

Brazin KN, Mallis RJ, Fulton DB, Andreotti AH (2002) Regulation of the tyrosine kinase Itk by the peptidyl-prolyl isomerase cyclophilin A. *Proceedings of the National Academy of Sciences of the United States of America* **99**: 1899–1904.

Brito R, Chimal ME, Gaxiola G, Rosas C (2000) Growth, metabolic rate, and digestive enzyme activity in the white shrimp *Litopenaeus setiferus* early postlarvae fed different diets. *Journal of Experimental Marine Biology and Ecology* **255**: 21–36.

Brown JW, Echeverria M, Qu LH (2003) Plant snoRNAs: Functional evolution and new modes of gene expression. *Trends in Plant Science* **8**: 42–49.

Buřič M, Kouba A, Kozak P (2010) Intra-sex dimorphism in crayfish females. *Zoology* **113**: 301–307.

Burtenshaw SM, Su PP, Zhang JR, Tobe SS, Dayton L, Bendena WG (2008) A putative farnesoic acid O-methyltransferase (FAMeT) orthologue in *Drosophila melanogaster* (CG10527): Relationship to juvenile hormone biosynthesis? *Peptides* **29**: 242–251.

Cadoret JP, Debon R, Cornudella L, Lardans V, Morvan A, Roch P (1999) Transient expression assays with the proximal promoter of a newly characterized actin gene from the oyster *Crassostrea gigas*. *FEBS Letters* **460**: 81–85.

Cameron JN, Thomas P (1992) Calcitonin-like immunoreactivity in the blue crab: tissue distribution, variations in the molt cycle, and partial characterization. *Journal of Experimental Zoology* **262**: 279–286.

Carter GW, Galas DJ, Galitski T (2009) Maximal extraction of biological information from genetic interaction data. *PLoS Computational Biology* **5**: e1000347.

Castellanos M, Jiménez-Vega F, Vargas-Albores F (2008) Single IB domain SIBD protein from *Litopenaeus vannamei* a novel member for the IGFBP family. *Comparative Biochemistry and Physiology* **3A**: 270–274.

Cesar JRO, Yang J (2007) Expression patterns of ubiquitin, heat shock protein70, α-actin and β-actin over the molt cycle in the abdominal muscle of marine shrimp *Litopenaeus vannamei*. *Molecular Reproduction and Development* **74**: 554-559.

Chang ES (2001) Crustacean hyperglycemic hormone family: old paradigms and new perspectives. *American Zoologist* **41**: 380–388.

Chang ES, Chang SA, Mulder EP (2001) Hormones in the lives of crustaceans: an overview. *American Zoologist* **41**: 1090–1097.

Chang ES, Mykles DL (2011) Regulation of crustacean molting: A review and our perspectives. *General and Comparative Endocrinology* **172**: 323–330.

Charmantier G, Charmantier-Daures M, Aiken DE (1989) Human somatotropin stimulates the growth of young American lobsters, *Homarus americanus* (Crustacea, Decapoda). *Comptes Rendus de l'Académie des Sciences-Series III* **308**: 21–26.

Chen DX, Jin QJ, Fang XT, Zhang CL, Sun JJ, Shi XY *et al*. (2010) Analysis of the polymorphisms in the caprine PRDM16, SHH and SF-1 genes and their association with production traits in goats. *Small Ruminant Research* **93**: 193–197.

Chen K, Hawken R, Flickinger GH, Rodriguez-Zas SL, Rund LA, Wheeler MB *et al*. (2012) Association of the porcine transforming growth factor beta type I receptor (TGFBR1) gene with growth and carcass traits. *Animal Biotechnology* **23**: 43–63.

Chen L, Mu C, Zhao J, Wang C (2011) Molecular cloning and characterization of two isoforms of cyclophilin A gene from *Venerupis philippinarum*. *Fish and Shellfish Immunology* **31**: 1218–1223.

Chen S, Huner JV, Malone RF (1993) Molting and mortality of red swamp and white river crawfish subjected to eyestalk ablation: a preliminary study for commercial soft-shell crawfish production. *Journal of the World Aquaculture Society* **24**: 48–57.

Chen S, Jingwei W, Huner JV, Malone RF (1995) Effects of temperature upon ablation-to-molt interval and mortality of red swamp crayfish (*Procambarus clarkii*) subjected to bilateral eyestalk ablation. *Aquaculture* **138**: 205–217.

Chen SH, Lin CY, Kuo CM (2005) *In silico* analysis of crustacean hyperglycaemic hormone family. *Marine Biotechnology* **7**: 193–206.

Chen WY, Ho KC, Leu JH, Liu KF, Wang HC, Kou GH *et al*. (2008) WSSV infection activates STAT in shrimp. *Developmental and Comparative Immunology* **32**: 1142–1150.

Chen YH, Jia XT, Huanga XD, Zhanga S, Li M, Xiea JF *et al*. (2011) Two *Litopenaeus vannamei* HMGB proteins interact with transcription factors *Lv*STAT and *Lv*Dorsal to activate the promoter of white spot syndrome virus immediate-early gene ie1. *Molecular Immunology* **48**: 793–799.

Cheng CH, Chen GD, Yeh MS, Chu CY, Hsu YL, Hwang PP *et al* (2010) Expression and characterization of the Jak kinase and STAT protein from brine shrimp, *Artemia franciscana*. *Fish and Shellfish Immunology* **28**: 774–782.

Chlenski A, Cohn SL (2010) Modulation of matrix remodeling by SPARC in neoplastic progression. *Semininars in Cell and Developmental Biology* **21**: 55–65.

Christians ES, Zhou Q, Renard J, Benjamin IJ (2003) Heat shock proteins in mammalian development. *Semininars in Cell and Developmental Biology* **14**: 283–290.

Chung JS, Zmora N, Katayama H, Tsutsui N (2010) Crustacean hyperglycemic hormone (CHH) neuropeptides family: Functions, titer, and binding to target tissues. *General and Comparative Endocrinology* **166**: 447–454.

Ciechanover A (1998) The ubiquitin-proteasome pathway: on protein death and cell life. *EMBO Journal* **17**: 7151–7160.

Ciobanu DC, Bastiaansen JWM, Magrin J, Rocha JL, Jiang DH, Yu N *et al*. (2010a) A major SNP resource for dissection of phenotypic and genetic variation in Pacific white shrimp (*Litopenaeus vannamei*). *Animal Genetics* **41**: 39–47.

Ciobanu DC, Lu L, Mozhui K, Wang X, Jagalur M, Morris JA *et al*. (2010b) Detection, validation, and downstream analysis of allelic variation in gene expression. *Genetics* **184**: 119–128.

Clark MC, Dever TE, Dever JJ, Xu P, Rehder V, Sosa MA *et al*. (2004) Arthropod 5-HT2 receptors: a neurohormonal receptor in decapods crustaceans that displays agonist independent activity resulting from an evolutionary alteration to the DRY motif. *Journal of Neuroscience* **24**: 3421–3435.

Clark MS, Thorne MAS, Toullec JY, Meng Y, Guan LL, Peck LS *et al*. (2011) Antarctic krill 454 pyrosequencing reveals chaperone and stress transcriptome. *PLoS One* **6**: e15919.

Colbourne JK, Pfrender ME, Gilbert D, Thomas WK, Tucker A, Oakley TH *et al*. (2011) The ecoreponsive genome of *Daphnia pulex*. *Science* **331**: 555–561.

Corsico B, Liou HL, Storch J (2004) The alphahelical domain of liver fatty acid binding proteinis responsible for the diffusion-mediated transfer of fatty acids to phospholipid membranes. *Biochemistry* **43**: 3600–3607.

Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS *et al*. (2010) The genetic landscape of a cell. *Science* **327**: 425–431.

Covi JA, Kim HW, Mykles DL (2008) Expression of alternatively spliced transcripts for a myostatin-like protein in the blackback land crab, *Gecarcinus lateralis*. *Comparative Biochemistry and Physiology* **150A**: 423–430.

Covi JA, Chang ES, Mykles DL (2009) Conserved role of cyclic nucleotides in the regulation of ecdysteroidogenesis by the crustacean molting gland. *Comparative Biochemistry and Physiology* **152A**: 470–477.

Crombach A, Hogeweg P (2008) Evolution of evolvability in gene regulatory networks. *PLoS Computational Biology* **4**: e1000112.

Curi RA, Oliveira HN, Silveira AC, Lopes CR (2005) Association between IGF-I, IGF-IR and GHRH gene polymorphisms and growth and carcass traits in beef cattle. *Livestock Production Science* **94**: 159–167.

Dahl SW, Halkier T, Lauritzen C, Dolenc I, Pedersen J, Turk V *et al*. (2001) Human recombinant pro-dipeptidyl peptidase I (cathepsin C) can be activated by cathepsins L and S but not by autocatalytic processing. *Biochemistry* **40**: 1671–1678.

Dantuma NP, Groothuis TA, Salomons FA, Neefjes J (2006) A dynamic ubiquitin equilibrium couples proteasomal activity to chromatin remodeling. *Journal of Cell Biology* **173**: 19–26.

Dearolf CR (1999) Jaks and STATs in invertebrate model organisms. *Cellular and Molecular Life Science* **55**: 1578–1584.

De Kleijn DPV, Van Herp F (1995) Molecular biology of neurohormone precursors in the eyestalk of crustacean. *Comparative Biochemistry and Physiology* **112B**: 573–579.

De Kleijn DPV, Van Herp F (1998) Involvement of the hyperglycemic neurohormone family in the control of reproduction in decapod crustaceans. *Invertebrate Reproduction and Development* **33**: 263–272.

Delgado I, Fuentes E, Escobar S, Navarro C, Corbeaux T, Reyes AE *et al*. (2008) Temporal and spatial expression pattern of the myostatin gene during larval and juvenile stages of the Chilean flounder (*Paralichthys adspersus*). *Comparative Biochemistry and Physiology* **151B**: 197–202.

De Loof A, Vanden Broeck J, Janssen I (1996) Hormones and the cytoskeleton of animals and plants. *International Review of Cytology* **166**: 1–58.

Denardi C, Ausoni S, Moretti P (1993) Type-2X-myosin heavy-chain is coded by a muscle fiber type-specific and developmentally-regulated gene. *Journal of Cell Biology* **123**: 823–835.

De Oliveira Cesar JR, Zhao B, Malecha S, Ako H, Yang J (2006) Morphological and biochemical changes in the muscle of the marine shrimp *Litopenaeus vannamei* during the molt cycle. *Aquaculture* **261**: 688–694.

Deryusheva S, Gall JG (2009) Small Cajal body-specific RNAs of *Drosophila* function in the absence of Cajal bodies. *Molecular Biology of the Cell* **20**: 5250–5259.

De Santis C, Jerry DR (2007) Candidate growth genes in finfish—where should we be looking? *Aquaculture* **272**: 22–38.

De Santis C, Evans BS, Smith-Keune C, Jerry DR (2008) Molecular characterisation, tissue expression and sequence variability of the barramundi (*Lates calcarifer*) myostatin gene. *BMC Genomics* **9**: 82–97.

De Santis C, Wade NM, Jerry DR, Preston NP, Glencross BD, Sellars MJ (2011) Growing backwards: an inverted role for the shrimp ortholog of vertebrate myostatin and GDF11. *Journal of Experimental Biology* **214**: 2671–2677.

Devaraj H, Natarajan A (2006) Molecular mechanisms regulating molting in a crustacean. *FEBS Journal* **273**: 839–846.

Dhar AK, Bowersb RM, Licona KS, Veazey G, Read B (2009) Validation of reference genes for quantitative measurement of immune gene expression in shrimp. *Molecular Immunology* **46**: 1688–1695.

Dominique JE, Gérard C (2006) Myostatin regulation of muscle development: Molecular basis, natural mutations, physiopathological aspects. *Experimental Cell Research* **312**: 2401–2414.

Du YC, Stillman B (2002) Yph1p, an ORC-interacting protein: potential links between cell proliferation control, DNA replication, and ribosome biogenesis. *Cell* **109**: 835–848.

Du ZQ, Ciobanu DC, Onteru SK, Gorbach D, Mileham AJ, Jaramillo G *et al*. (2010) A gene-based SNP linkage map for pacific white shrimp, *Litopenaeus vannamei*. *Animal Genetics* **41**: 286–294.

Duan C, Ren H, Gao S (2010) Insulin-like growth factors (IGFs), IGF receptors, and IGF-binding proteins: roles in skeletal muscle growth and differentiation. *General and Comparative Endocrinology* **167**: 344–351.

Ekblom R, Galindo J (2011) Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity* **107**: 1–15.

Escamilla-Chimal EG, Hiriart M, Sanchez-Soto MC, Fanjul-Moles ML (2002) Serotonin modulation of CHH secretion by isolated cells of the crayfish retina and optic lobe. *General and Comparative Endocrinology* **125**: 283–290.

Estany J, Tor M, Villalba D, Bosch L, Gallardo D, Jiménez N *et al*. (2007) Association of CA repeat polymorphism at intron 1 of insulin-like growth factor (IGF-I) gene with circulating IGF-I concentration, growth, and fatness in swine. *Physiological Genomics* **31**: 236–243.

Esteves A, Ehrlich R (2006) Invertebrate intracellular fatty acid binding proteins. *Comparative Biochemistry and Physiology* **2C**: 262–274.

Fang CH, Li BG, Sun XY, Hasselgren PO (2000) Insulin-like growth factor I reduces ubiquitin and ubiquitin-conjugating enzyme gene expression but does not inhibit muscle proteolysis in septic rats. *Endocrinology* **141**: 743–2751.

Fanjul-Moles ML (2006) Biochemical and functional aspects of crustacean hyperglycaemic hormone in decapods crustaceans: Review and update. *Comparative Biochemistry and Physiology* **142C**: 390–400.

FAO (2009) Fisheries statistical database, global aquaculture production (fisheriesglobal information system, online query).

Fatica A, Tollervey D (2002) Making ribosomes. *Current Opinion in Cell Biology* **14**: 313–318.

Flegel TW, Sritunyalucksana K (2011) Shrimp molecular responses to viral pathogens. *Marine Biotechnology* **13**: 587–607.

Fidantsef AL, Stout MJ, Thaler JS, Duffy SS, Bostock RM (1999) Signal interactions in pathogen and insect attack: expression of lipoxygenase, proteinase inhibitor II, and pathogenesis-related protein P4 in the tomato, *Lycopersicon esculentum*. *Physiological and Molecular Plant Pathology* **54**: 97–114.

Fortuna MA, Albaladejo RG, Fernández L, Aparicio A, Bascompte J (2009) Networks of spatial genetic variation across species. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 19044–19049.

Funk SE, Sage EH (1991) The Ca^{2+}-binding glycoprotein SPARC modulates cell cycle progression in bovine aortic endothelial cells. *Proceedings of the National Academy of Sciences of the United States of America* **88**: 2648–2652.

Gai Y, Wang L, Zhao J, Qiu L, Song L, Li L *et al*. (2009) The construction of a cDNA library enriched for immune genes and the analysis of 7535 ESTs from Chinese mitten crab *Eriocheir sinensis*. *Fish and Shellfish Immunology* **27**: 684–694.

Gai Y, Wang L, Song L, Zhao J, Qiu L, Xing K (2010) A putative endocrine factor SIBD (single insulin binding domain protein) involved in immune response of Chinese mitten crab *Eriocheir sinensis*. *Fish and Shellfish Immunology* **28**: 10–17.

Galat A (1999) Variations of sequences and amino acid compositions of proteins that sustain their biological functions: an analysis of the cyclophilin family of proteins. *Archives of Biochemistry Biophysics* **371**: 149–162.

Gamboa-Delgado J, Molina-Poveda C, Cahu C (2003) Digestive enzyme activity and food ingesta in juvenile shrimp *Litopenaeus vannamei* (Boone, 1931) as a function of body weight. *Aquaculture Research* **34**: 1403–1411.

Garofalo RS (2002) Genetic analysis of insulin signaling in *Drosophila*. *Trends in Endocrinology and Metabolism* **13**:156–162.

Garrett AJ, Rincon G, Medrano JF, Elzo MA, Silver GA, Thomas MG (2008) Promoter region of the bovine growth hormone receptor gene: single nucleotide polymorphism discovery in cattle and association with performance in Brangus bulls. *Journal of Animal Science* **86**: 3315–3323.

Garvin MR, Saitoh K, Gharrett AJ (2010) Application of single nucleotide polymorphisms to non-model species: a technical review. *Molecular Ecology Resources* **10**: 915–934.

Gauvry L, Mohan-Ram V, Ettelaie C, Ennion S, Goldspink G (1997) Molecular motors designed for different tasks and to operate at different temperatures. *Journal of Thermal Biology* **6**: 367–373.

Gerhardt CC, Van Heerikhuizen H (1997) Functional characteristics of heterologously expressed 5-HT receptors. *European Journal of Pharmacology* **334**: 1–23.

Gilad Y, Pritchard JK, Thornton K (2009) Characterizing natural variation using next-generation sequencing technologies. *Trends in Genetics* **25**: 463-471.

Gill JL, Bishop SC, McCorquodale C, Williams JL, Wiener P (2009) Associations between the 11-bp deletion in the myostatin gene and carcass quality in Angus-sired cattle. *Animal Genetics* **40**: 97–100

Giribet G, Edgecombe GD (2011) Reevaluating the Arthropod tree of life. *Annual Review of Entomology* **57**: 167–186.

Gitterle T, Rye M, Salte R, Cock J, Johansen H, Lozano C *et al*. (2005) Genetic (co)variation in harvest body weight and survival in *Penaeus* (*Litopenaeus*) *vannamei* under standard commercial conditions. *Aquaculture* **243**: 83–92.

Glatz JFC, Van der Vusse G (1996) Cellular fatty acid-binding proteins: their function and physiological significance. *Journal of Progress in Lipid Research* **3535**: 243–282.

Glenn KL, Grapes L, Suwanasopee T, Harris DL, Li Y, Wilson K *et al*. (2005) SNP analysis of AMY2 and CTSL genes in *Litopenaeus vannmei* and *Penaeus monodon* shrimp. *Animal Genetics* **36**: 235–236.

Goddard ME, Hayes BJ (2009) Mapping genes for complex traits in domestic animals and their use in breeding programs. *Nature Reviews Genetics* **10**: 381–391.

Gong YN, Li WW, Sun JL, Ren F, He L, Jiang H *et al*. (2010) Molecular cloning and tissue expression of the fatty acid-binding protein (Es-FABP) gene in female Chinese mitten crab (*Eriocheir sinensis*). *BMC Molecular Biology* **11**:71.

Gopal C, Gopikrishna G, Krishna G, Jahageerdar SS, Rye M, Hayes BJ *et al*. (2010) Weight and time of onset of female-superior sexual dimorphism in pond reared *Penaeus monodon*. *Aquaculture* **300**: 237–239.

Gornik SG, Westrop GD, Coombs GH, Neil DM (2010) Molecular cloning and localization of a calpain-like protease from the abdominal muscle of Norway lobster *Nephrops norvegicus*. *Molecular Biology Reports* **37**: 2009–2019.

Greene MW, Chen TT (1997) Temporal expression pattern of insulin-like growth factor mRNA during embryonic development in a teleost, rainbow trout (*Onchorynchus mykiss*). *Molecular Marine Biology and Biotechnology* **6**: 144–151.

Gu PL, Yu KL, Chan SM (2000) Molecular characterization of an additional shrimp hyperglycemic hormone: cDNA cloning, gene organization, expression and biological assay of recombinant proteins. *FEBS Letters* **472**: 122–128.

Gu PL, Tobe SS, Chow BKC, Chu KH, He JG, Chan SM (2002) Characterization of an additional molt-inhibiting hormone-like neuropeptide from the shrimp *Metapenaeus ensis*. *Peptides* **23**: 1875– 1883.

Gu ZL, Zhu DH, Li N, Li H, Deng XM, Wu CX (2004) The single nucleotide polymorphisms of the chicken myostatin gene are associated with skeletal muscle and adipose growth. *Science in China Series C Life Sciences* **47**: 25–30.

Guedes SD, Vitorino R, Domingues R, Tomer K, Correia AJF, Amado F (2005) Proteomics of immune-challenged *Drosophila melanogaster* larvae hemolymph. *Biochemical and Biophysical Research Communications* **328**: 106–115.

Gupta N, Pandey A, Malik G, Gupta SC (2009) Single nucleotide polymorphism (SNP) in growth hormone gene of Jakhrana, a prominent milk goat breed in India. *Small Ruminant Research* **81**: 35–41.

Gusev NB, Bukach OV, Marston SB (2005) Structure, properties, and probable physiological role of small heat shock protein with molecular mass 20 kD (Hsp20, HspB6). *Biochemistry Moscow* **70**: 629–637.

Gutiérrez A, Nieto J, Pozo F, Stern S, Schoofs L (2007) Effect of insulin/IGF-I like peptides on glucose metabolism in the white shrimp *Penaeus vannamei*. *General and Comparative Endocrinology* **153**: 170–175.

Haas AL, Baboshina O, Williams B, Schwartz LM (1995) Coordinated induction of the ubiquitin conjugation pathway accompanies the developmentally programmed death of insect skeletal muscle. *Journal of Biological Chemistry* **270**: 9407–9412.

Hamasaki K, Kitada S (2008) Potential of stock enhancement for decapods crustaceans. *Reviews in Fisheries Science* **16**: 164–174.

Hamilton GS, Steiner JP (1998) Immunophilins: beyond immunosuppression. *Journal of Medicinal Chemistry* **41**: 5119–5143.

Han R, Wei Y, Kang X, Chen H, Sun G, Li G *et al*. (2012) Novel SNPs in the PRDM16 gene and their associations with performance traits in chickens. *Molecular Biology Reports* **39**: 3153–3160.

Hao LL, Yu H, Zhang Y, Sun SC, Liu SC, Zeng YZ *et al*. (2011) Single nucleotide polymorphism analysis of exon 3 and 4 of IGF-1 gene in pigs. *Genetics and Molecular Research* **10**: 1689–1695.

Hara MR, Agrawal N, Kim SF, Cascio MB, Fujimuro M, Ozeki Y *et al*. (2005) S-nitrosylated GAPDH initiates apoptotic cell death by nuclear translocation following Siah1 binding. *Nature Cell Biology* **7**: 665–674.

Harrington WF, Rogers ME (1984) Myosin. *Annual Review of Biochemistry* **53**: 35–73.

Harris BS, Zhang Y, Card L, Rivera LB, Brekken RA, Bradshaw AD (2011) SPARC regulates collagen interaction with cardiac fibroblast cell surfaces. *American Journal of Physiology Heart and Circulatory Physiology* **301**: 841–847.

Hartnoll RG (2001) Growth in crustaceans-twenty years on. *Hydrobiologia* **449**: 111–122.

Harzsch S, Kreissl S (2010) Myogenesis in the thoracic limbs of the American lobster. *Arthropod Structure and Development* **39**: 423–435.

Hauser L, Carvalho GR (2008) Paradigm shifts in fisheries genetics: ugly hypotheses slain by beautiful facts. *Fish and Fishereis* **9**: 333–362.

Hauser L, Seeb JE (2008) Advances in molecular technology and their impact on fisheries genetics. *Fish and Fisheries* 9: 473–486.

Herman IM (1993) Actin isoforms. *Current Opinion in Cell Biology* **5**: 48–55.

Hess DC, Myers CL, Huttenhower C, Hibbs MA, Hayes AP, Paw J *et al*. (2009) Computationally driven, quantitative experiments discover genes required for mitochondrial biogenesis. *PLoS Genetics* **5**: e1000407.

Hickford JG, Forrest RH, Zhou H (2009) Association between a g.+6723G-A SNP in the myostatin gene (MSTN) and carcass traits in New Zealand Texel sheep. *Journal of Animal Science* **87**: 1856–1864.

Hightower RC, Meagher RB (1986) The molecular evolution of actin. *Genetics* **114**: 315–332.

Ho SY, Delgado L, Storch J (2002) Monoacylglycerol metabolism in human intestinal Caco-2 cells: evidence for metabolic compartmentation and hydrolysis. *Journal of Biological Chemistry* **277**: 1816–1823.

Holford KC, Edwards KA, Bendena WG, Tobe SS, Wang Z, Borst DW (2004) Purification and characterization of a mandibular organ protein from the American lobster, *Homarus americanus*: A putative farnesoic acid *O*-methyltransferase. *Insect Biochemistry and Molecular Biology* **34**: 785–798.

Hooper SL, Thuma JB (2005) Invertebrate muscles: muscle specific genes and proteins. *Physiological Reviews* **85**: 1001–1060.

Hooper SL, Hobbs KH, Thuma JB (2008) Invertebrate muscles: thin and thick filament structure; molecular basis of contraction and its regulation, catch and asynchronous muscle. *Progress in Neurobiology* **86**: 72–127.

Hopkins PM (2012) The eyes have it: A brief history of crustacean neuroendocrinology. *General and Comparative Endocrinology* **175**: 357–366.

Horn T, Sandmann T, Fischer B, Axelsson E, Huber W, Boutros M (2011) Mapping of signalling networks through synthetic genetic interaction analysis by RNAi. *Nature Methods* **8**: 341–346.

Hou SX, Zheng Z, Chen X, Perrimon N (2002) The Jak/STAT pathway in model organisms: emerging roles in cell movement. *Developmental Cell* **3**: 765–778.

Hou G, Wang D, Guan S, Zeng H, Huang X, Ma Y (2010) Associated analysis of single nucleotide polymorphisms of IGF2 gene's exon 8 with growth traits in Wuzhishan pig. *Molecular Biology Reports* **37**: 497–500.

Hou G, Huang M, Gao X, Li J, Gao H, Ren H *et al*. (2011) Association of Calpain 1 (CAPN1) and HRSP12 allelic variants in beef cattle with carcass traits. *African Journal of Biotechnology* **10**: 13714–13718.

Hu X, Guo H , He Y, Wang S, Zhang L, Wang S, *et al*. (2010) Molecular characterization of myostatin gene from Zhikong scallop *Chlamys farrei* (Jones et Preston 1904). *Genes and Genetics Systems* **85**: 207–218.

Hu KJ, Leung PC (2004) Shrimp cathepsin L encoded by an intronless gene has predominant expression in hepatopancreas, and occurs in the nucleus of oocyte. *Comparative Biochemistry and Physiology* **137 B**: 21–33.

Hu KJ, Leung PC (2007) Food digestion by cathepsin L and digestion-related rapid cell differentiation in shrimp hepatopancreas. *Comparative Biochemistry and Physiology* **146B**: 69–80.

Huang CF, Chuang NN (1998) Molecular cloning of *ras* cDNA from *Penaeus japonicus* (Crustacea, Decapoda): geranylgeranylation and guanine nucleotide binding. *Gene* **224**: 117–122.

Huang WJ, Leu JH, Tsau MT, Chen JC, Chen LL (2011a) Differential expression of LvHSP60 in shrimp in response to environmental stress. *Fish and Shellfish Immunology* **30**: 576–582.

Huang Z, Chen X, Chen (2011b) Myostatin: A novel insight into its role in metabolism, signal pathways, and expression regulation. *Cellular Signalling* **23**: 1441–1446.

Hui JHL, Tobe SS, Chan SM (2008) Characterization of the putative farnesoic acid *O*-methyltransferase (LvFAMeT) cDNA from white shrimp, *Litopenaeus vannamei*: Evidence for its role in molting. *Peptides* **29**: 252–260.

Hulata G (2001) Genetic manipulations in aquaculture: a review of stock improvement by classical and modern technologies. *Genetica* **111**: 155–173.

Hutchings JA, Fraser DJ (2008) The nature of fisheries-and farming-induced evolution. *Molecular Ecology* **17**: 294–313.

Hwa V, Oh Y, Rosenfeld RG (1999) The insulin-like growth factor-binding protein IGFBP superfamily. *Endocrine Review* **20**: 761–787.

Janecek S (1997) α-amylase family: molecular biology and evolution. *Progress in Biophysics and Molecular Biology* **67**: 67–97.

Jaendling A, McFarlane RJ (2010) Biological roles of translin and translin-associated factor-X: RNA metabolism comes to the fore. *Biochemical Journal* **429**: 225–234.

Jiang HL, Wang Y, Wu MZ, Gu ZL, Frank SJ, Torres-Diaz R (2007) Growth hormone stimulates hepatic expression of bovine growth hormone receptor messenger ribonucleic acid through signal transducer and activator of transcription 5 activation of a major growth hormone receptor gene promoter. *Endocrinology* **148**: 3307–3315.

Jiang H, Cai YM, Chen LQ, Zhang XW, Hu SN, Wang Q (2009a) Functional annotation and analysis of expressed sequence tags the hepatopancreas of mitten crab (*Eriocheir sinensis*). *Marine Biotechnology* **11**: 317–326.

Jiang H, Yin Y, Zhang X, Hu S, Wang Q (2009b) Chasing relationships between nutrition and reproduction: A comparative transcriptome analysis of hepatopancreas and testis from *Eriocheir sinensis*. *Compatative Biochemistry and Physiology* **4D**: 227–234.

Jirikowski G, Kreissl S, Richter S, Wolff C (2010) Muscle development in the marbled crayfish-insights from an emerging model organism (Crustacea, Malacostraca, Decapoda). *Development Genes and Evolution* **220**: 89–105.

Joulia D, Bernardi H, Garandel V, Rabenoelina F, Vernus B, Cabello G (2003) Mechanisms involved in the inhibition of myoblast proliferation and differentiation by myostatin. *Experimental Cell Research* **286**: 263–275.

Juhua Y, Xuefeng C, Jianlin L, Yongkai T, Hongxia L, Pao X *et al*. (2010) Isolation of IGF2 and association of IGF2 polymorphism with growth trait in genetically improved farmed tilapia, *Oreochromis niloticus* L. *Aquaculture Research* **41**: 743–750.

Jung HH, Lieber RL, Ryan AF (1998) Quantification of myosin heavy chain mRNA in somatic and branchial arch muscles using competitive PCR. *American Journal of Physiology Cell Physiology* **275**: 68–74.

Jung H, Lyons RE, Dinh H, Hurwood DA, McWilliam S, Mather PB (2011) Transcriptomics of a giant freshwater prawn (*Macrobrachium rosenbergii*): De novo assembly, annotation and marker discovery. *PLoS One* **6**: e27938.

Kamimura MT, Meier KM, Cavalli RO, Laurino J, Maggioni R, Marins LF (2008) Characterization of growth-related genes in the south-western Atlantic pink shrimp *Farfantepenaeus paulensis* (Pérez-Farfante 1967) through a modified DDRT-PCR protocol. *Aquaculture Research* **39**: 200–204.

Karrer KM, Peiffer SL, DiTomas ME (1993) Two distinct gene subfamilies within the family of cysteine proteinase genes. *Proceedings of the National Academy of Sciences of the United States of America* **90**: 3063–3067.

Kawahara-Miki R, Wada K, Azuma N, Chiba S (2011) Expression profiling without genome sequence information in a non-model species, Pandalid shrimp (*Pandalus latirostris*), by next-generation sequencing. *PLoS One* **6**: e26043.

Khandwala HM, McCutcheon IE, Flyvbjerg A, Friend KE (2000) The effects of insulin-like growth factors on tumorigenesis and neoplastic growth. *Endocrine Review* **21**: 215–244.

Kijas JW, McCulloch R, Edwards JEH, Oddy VH, Lee SH, Van der Werf J *et al*. (2007) Evidence for multiple alleles affecting muscling and fatness at the ovine GDF8 locus. *BMC Genetics* **8**: 80.

Kim BK, Kim KS, Oh CW, Mykles DL, Lee SG, Kim HJ *et al*. (2009a) Twelve actin-encoding cDNAs from the American lobster, *Homarus americanus*: Cloning and tissue expression of eight skeletal muscle, one heart, and three cytoplasmic isoforms. *Comparative Biochemistry and Physiology* **153B**: 178–184.

Kim H, Mykles DL, Goetz FW, Roberts SB (2004a) Characterization of a myostatin-like gene from the bay scallop *Argopecten irradians*. *Biochemica et Biophysica Acta* **1679**: 174–179.

Kim HW, Chang ES, Mykles DL (2005) Three calpains and ecdysone receptor in the land crab *Gecarcinus lateralis*: sequences, expression and effects of elevated ecdysteroid induced by eyestalk ablation. *Journal of Experimental Biology* **208**: 3177–3197.

Kim KS, Jeon JM, Kim HW (2009b) A myostatin-like gene expressed highly in the muscle tissue of Chinese mitten crab, *Eriocheir sinensis*. *Fisheries and Aquatic Sciences* **12**: 185–193.

Kim KS, Kim YJ, Jeon JM, Kang YS, Kang YS, Oh CW et al. (2010) Molecular characterization of myostatin-like genes expressed highly in the muscle tissue from Morotoge shrimp, *Pandalopsis japonica*. *Aquaculture Research* **41**: 862–871.

Kim KS, Lee JJ, Shin HY, Choi BH, Lee CK, Kim JJ *et al*. (2006) Association of melanocortin 4 receptor (MC4R) and high mobility group AT-hook1 (HMGA1) polymorphisms with pig growth and fat deposition traits. *Animal Genetics* **37**: 419–421.

Kim SH, Lessner SM, Sakurai Y, Galis ZS (2004b) Cyclophilin A as a novel biphasic modiator of endothelial activation and dysfunction. *American Journal of Pathology* **164**: 1567–1574.

Kim WJ, Jung HT, Gaffney PM (2011) Development of Type I genetic markers from expressed sequence tags in highly polymorphic species. *Marine Biotechnology* **13**: 127–132.

Kocour M, Kohlmann (2011) Growth hormone gene polymorphisms in trench, *Tinca tinca* L. *Aquaculture* **310**: 298–304.

Koenders A, Yu X, Chang ES, Mykles DL (2002) Ubiquitin and actin expression in claw muscles of land crab, *Geocarcinus lateralis*, and American lobster, *Homarus americanus*: Differential expression of ubiquitin in two slow muscle fiber types during molt-induced atrophy. *Journal of Experimental Zoology* **292**: 618–632.

Komali M, Kalarani V, Venkatrayulu CH, Chandra Sekhara Reddy D (2005) Hyperglycaemic effects of 5-Hydroxytryptamine and dopamine in the freshwater prawn, *Macrobrachium malcolmsonii*. *Journal of Experimental Zoology* **303A**: 448–455.

Kopchick JJ, Andry JM (2000) Growth hormone (GH), GH receptor, and signal transduction. *Molecular Genetics and Metabolism* **71**: 293–314.

Kregel KC (2002) Heat shock proteins: modifying factors in physiological stress responses and acquired thermotolerance. *Journal of Applied Physiology* **92**: 2177–2186.

Kreissl S, Uber A, Harzsch S (2008) Muscle precursor cells in the developing limbs of two isopods (Crustacea, Peracarida): an immunohistochemical study using a novel

monoclonal antibody against myosin heavy chain. *Development Genes and Evolution* **218**: 253–265.

Kuballa AV, Elizur A (2007) Novel molecular approach to study moulting in crustaceans. *Bulletin Fisheries Research Agency* **20**: 53–57.

Kuballa AV, Guyatt K, Dixon B, Thaggard H, Ashton AR, Paterson B *et al*. (2007) Isolation and expression analysis of multiple isoforms of putative farnesoic acid *O*-methyltransferase in several crustacean species. *General Comparative and Endocrinology* **150**: 48–58.

Kumagai H, Sakai H (1983) A porcine brain protein (35 K protein) which bundles microtubules and its identification as glyceraldehyde 3-phosphate dehydrogenase. *Journal of Biochemistry* **93**: 1259–1269.

Lachaise F, Le Roux A, Hubert M, Lafont R (1993) The molting gland of crustaceans: localization, activity, and endocrine control. *Journal of Crustacean Biology* **13**: 198–234.

L_AFramboise WA, Griffis B, Bonner P, Warren W, Ssclise D, Guthrie R *et al*. (2000) Muscle type–specific myosin isoforms in crustacean muscles. *Journal of Experimental Zoology* **183**: 185–201.

Lago-Lestón A, Ponce E, Muñoz ME (2007) Cloning and expression of hyperglycemic (CHH) and molt-inhibiting (MIH) hormones mRNAs from the eyestalk of shrimps of *Litopenaeus vannamei* grown in different temperature and salinity conditions. *Aquaculture* **270**: 343–357.

Lane TF, Sage EH (1990) Functional mapping of SPARC: peptides from two distinct $Ca^{+(+)}$-binding sites modulate cell shape. *Journal of Cell Biology* **111**: 3065–3076.

Laufer H, Borst D, Baker FC, Carraco C, Sinkus M, Reuter CC *et al*. (1987) Identification of a juvenile hormone-like compound in a crustacean. *Science* **235**: 202–205.

Laufer H, Ahl J, Rotllant G, Baclaski B (2002) Evidence that ecdysteroids and methyl farnesoate control allometric growth and differentiation in a crustacean. *Insect Biochemistry and Molecular Biology* **32**: 205–210.

Laycock MV, Hirama T, Hasnain S, Watson D, Storer AC (1989) Purification and characterization of a digestive cysteine proteinase from the American lobster (*Homarus americanus*). *Biochemical Journal* **263**: 439–444.

Le Boulay C, Sellos D, van Wormhoudt A (1998) Cathepsin L gene organization in crustaceans. *Gene* **218**: 77–84.

Le Boulay C, van Wormhoudt A, Sellos D (1996) Cloning and expression of cathepsin L-like proteinases in the hepatopancreas of the shrimp *Penaeus vannamei* during the intermolt cycle. *Journal of Comparative Physiology* **166B**: 310–318.

Lee SJ (2004) Regulation of muscle mass by myostatin. *Annual Review of Cell and Developmental Biology* **20**: 61–86.

Leekitcharoenphon P, Taweemuang U, Palittapongarnpim P, Kotewong R, Supasiri T, Sonthayanon B (2010) Predicted sub-populations in a marine shrimp proteome as revealed by combined EST and cDNA data from multiple Penaeus species. *BMC Research Notes* **3**: 295.

Leu J, Chen S, Wang Y, Chen Y, Su S, Lin C *et al*. (2011) A review of the major Penaeid shrimp EST studies and the construction of a shrimp transcriptome database based on the ESTs from four Penaeid shrimp. *Marine Biotechnology* **13**: 608–621.

Leverson JD, Ness SA (1998) Point mutations in v-Myb disrupt a cyclophilincatalyzed negative regulatory mechanism. *Molecular Cell* **1**: 203–211.

Li HF, Zhu WQ, Chen KW, Zhang TJ, Song WT (2009) Association of polymorphisms in the intron 1 of duck prolactin with egg performance. *Turkish Journal of Veterinry and Animal Sciences* **33**: 193–197.

Li P, Zha J, Huang H, Zhang Z, Sun H, Song D *et al*. (2009) Identification, mRNA expression and characterization of a novel ANK-like gene from Chinese mitten crab *Eriocheir japonica sinensis*. *Comparative Biochemistry and Physiology* **153B**: 332–339.

Li S, Friesen JA, Holford KC, Borst DW (2010) Methyl farnesoate synthesis in the lobster mandibular organ: The roles of HMG-CoA reductase and farnesoic acid *O*-methyltransferase. *Comparative Biochemistry and Physiology*, **155A**: 49–55.

Li T, Brouwer M (2009) Bioinformatic analysis of expressed sequence tags from grass shrimp *Palaemonetes pugio* exposed to environmental stressors. *Comparative Biochemistry and Physiology* **4D**: 187–195.

Li X, Bai J, Ye X, Hu Y, Li S, Yu L (2009) Polymorphisms in the 5' flanking region of the insulin-like growth factor I gene are associated with growth traits in largemouth bass *Micropterus salmoides*. *Fisheries Sciences* **75**: 351–358.

Li X, Bai J, Hu Y, Ye X, Li S, Yu L (2012) Genotypes, haplotypes and diplotypes of IGF-II SNPs and their associatin with growth traits in largemouth bass (*Micropterus salmoides*). *Molecular Biology Reports* **39**: 4359–4365.

Li XL, Xin JJ, Feng TT, Liu XL (2011b) Single nucleotide polymorphism discovery of molt inhibiting hormone gene 3 exons and its association with growth traits in white shrimp (*Litopenaeus vannamei*). *Journal of Animal and Veterinary Advances* **10**: 2856–2858.

Li WX (2008) Canonical and non-canonical Jak-STAT signaling. *Trends in Cell Biol* **18**: 545–551.

Li WW, Jin XK, He L, Jiang H, Gong YN, Xie YN *et al*. (2010) Molecular cloning, characterization, expression and activity analysis of cathepsin L in Chinese mitten crab, *Eriocheir sinensis*. *Fish and Shellfish Immunology* **29**: 1010–1018.

Li WW, Jin XK, He L, Gong YN, Jiang H, Wang Q (2011a) Molecular cloning and tissue expression of the fatty acid-binding protein (Es-FABP9) gene in the reproduction seasons of Chinese mitten crab, *Eriocheir sinensis*. *Molecular Biology Reports* **38**: 5169–5177.

Li Y, Dierens L, Byrne K, Miggiano E, Lehnert S, Preston N *et al*. (2006) QTL detection of production traits for the Kuruma prawn *Penaeus japonicus* (Bate) using AFLP markers. *Aquaculture* **258**: 198–210.

Li Z, Wu Y, Baraban JM (2008) The Translin/Trax RNA binding complex: clues to function in the nervous system. *Biochimica et Biophysica Acta* **1779**: 479–485.

Lin DT, Lechleiter JD (2001) Mitochondrial targeted cyclophilin D protects cells from cell death by peptidyl prolyl isomerization. *Journal of Biological Chemistry* **277**: 31134–31141.

Lin J, Arnold B, Della-Fera MA, Azain MJ, Hartzell DL, Baile CA (2002) Myostatin knockout in mice increases myogenesis and decreases adipogenesis. *Biochemical Biophysical Research Communications* **290**:701–706.

Liu HP, Chen RY, Zhang QX, Peng H, Wang KJ (2011) Differential gene expression profile from haematopoietic tissue stem cells of red claw crayfish, *Cherax quadricarinatus*, in response to WSSV infection. *Developmental and Comparative Immunology* **35**: 716–724.

Liu J, Shi GP, Zhang WQ, Zhang GR, Xu WH (2006) Cathepsin L function in insect moulting: molecular cloning and functional analysis in cotton bollworm, *Helicoverpa armigera*. *Insect Molecular Biology* **15**: 823–834.

Liu RZ, Li X, Godbout R (2008) A novel fatty acid-binding protein (FABP) gene resulting from tandem geneduplication in mammals: transcription in rat retina and testis. *Genomics* **92**: 436–445.

Loehlin DW, Werren JH (2012) Evolution of shape by multiple regulatory changes to a growth gene. *Science* **335**: 943–947.

Lorenzon S (2005) Hyperglycemic stress response in Crustacea. *Invertebrate Survival Journal* **2**: 132–141.

Lu J, Hou S, Huang W, Yu J, Wang W (2011) Polymorphisms in the myostatin gene and their association with growth and carcass traits in duck. *African Journal of Biotechnology* **10**: 11309–11312.

Lü A, Hu X, Chen H, Dong Y, Pang Y (2011) Single nucleotide polymorphisms of the prolactin receptor (PRLR) gene and its association with growth traits in chinese cattle. *Molecular Biology Reports* **38**: 261–266.

Lucy MC (2008) Functional differences in the growth hormone and insulin-like growth hormone and insulin-like growth factor axis in cattle and pigs: implications for post-partum nutrition and reproduction. *Reproduction in Domestic Animals* **43S**: 31–39.

Luo H, Dearolf CR (2001) The Jak/STAT pathway and *Drosophila* development. *Bioessays* **23**:1138–1147.

Lyons RE, Dierens LM, Tan SH, Preston NP, Li Y (2007) Characterization of AFLP markers associated with growth in the Kuruma prawn, *Marsupenaeus japonicus*, and identification of a candidate gene. *Marine Biotechnology* **9**: 712–721.

Mabb AM, Ehlers MD (2010) Ubiquitination in postsynaptic function and plasticity. *Annual Review of Cell Developmental Biology* **26**: 179–210.

Macario AJ, Conway de Macario E (2007) Molecular chaperones, multiple functions, pathologies, and potential applications. *Frontiers in Bioscience* **1**: 2588–2600.

Macias MT, Sastre L (1990) Molecular cloning and expression of four actin isoforms during *Artemia* development. *Nucleic Acids Research* **18**: 5219–5225.

MacLea KS, Covi JA, Kim H, Chao E, Medler S, Chang ES *et al*. (2010) Myostatin from the American lobster, *Homarus americanus*: Cloning and effects of molting on expression in skeletal muscles. *Comparative Biochemistry and Physiology*, **157A**: 328–337.

Magee DA, Sikora KM, Berkowicz EW, Berry DP, Howard DJ, Mullen MP *et al*. (2010) DNA sequence polymorphisms in a panel of eight candidate bovine imprinted genes and their association with performance traits in Irish Holstein-Friesian cattle. *BMC Genetics* **11**: 93.

Makgahlela ML, Fan B, Du ZQ, Rothschild MF (2009) Investigation of effects of three candidate genes on leg action and fat deposition traits in pigs. *South African Journal of Animal Science* **39**: 127–130.

Manor R, Weil S, Oren S, Glazer L, Aflalo ED, Ventura T *et al*. (2007) Insulin and gender: An insulin-like gene expressed exclusively in the androgenic gland of the male crayfish. *General and Comparative Endocrinology* **150**: 326–336.

Mareddy VR, Rosen O, Thaggard HB, Manor R, Kuballa AV, Aflalo ED *et al*. (2011) Isolation and characterization of the complete cDNA sequence encoding a putative insulin-like peptide from the androgenic gland of *Penaeus monodon*. *Aquaculture* **318**: 364–370.

Martin-Marti S, Onteru SK, Du ZQ, Rothschild MF (2010) SNP analyses of the 5HT1R and STAT genes in Pacific white shrimp, *Litopenaeus vannamei*. *Spanish Journal of Agricultural Research* **8**: 53–55.

Mayer MP, Bukau B (2005) Hsp70 chaperones, cellular functions and molecular mechanism. *Cellular and Molecular Life Sciences* **62**: 670–684.

McAndrew B, Napier J (2011) Application of genetics and genomics to aquaculture development: current and future directions. *Journal of Agricultural Science* **149**: 143–151.

McArthur MJ, Atshaves BP, Frolov A, Foxworth WD, Kier AB, Schroeder F (1999) Cellular uptake and intracellular trafficking of long chain fatty acids. *Journal of Lipid Research* **40**: 1371–1383.

McClure MC, Morsci NS, Schnabel RD, Kim JW, Yao P, Rolf MM *et al*. (2010) A genome scan for quantitative trait loci influencing carcass, post-natal growth and reproductive traits in commercial Angus Cattle. *Animal Genetics* **41**: 597–607.

McCurdy S, Baicu CF, Heymans S, Bradshaw AD (2010) Cardiac extracellular matrix remodeling: fibrillar collagens and secreted protein acidic and rich in cysteine (SPARC). *Journal of Molecular and Cellular Cardiology* **48**: 544–549.

McGrath ME (1999) The lysosomal cysteine proteinases. *Annual Review Biophysics and Biomolecular Structure* **28**: 181–204.

McPherron A, Lawler A, Lee S (1997) Regulation of skeletal muscle mass in mice by a new TGF-beta superfamily member. *Nature* **387**: 83–90.

McPherron A, Lawler A, Lee S (1999) Regulation of anterior posterior patterning of the axial skeleton by growth differentiation factor 11. *Nature Genetics* **22**: 260–264.

McPherron A, Lee S (1997) Double muscling in cattle due to mutations in the myostatin gene. *Proceedings of the National Academy of Sciences of the United States of America* **94**: 12457–12461.

Medeiros EF, Phelps MP, Fuentes FD, BradleyTM (2009) Overexpression of follistatin in trout stimulates increased muscling. *American Journal of Physiology* **297**: 235–242.

Medler S, Chang ES, Mykles DL (2007a) Muscle-specific calpain is localized in regions near motor endplates in differentiating lobster claw muscle. *Comparative Biochemistry and Physiology* **148A**: 591–598.

Medler S, Lilley TR, Riehl JH, Mulder EP, Chang ES, Mykles DL (2007b) Myofibrillar gene expression in differentiating lobster claw muscles. *Journal of Experimental Zoology* **307A**: 281–295.

Medler S, Mykles DL (2003) Analysis of myofibrillar proteins and transcripts in adult skeletal muscles of the American lobster *Homarus americanus*: variable expression of myosins, actin and troponins in fast, slow-twitch and slow-tonic fibres. *Journal of Experimental Zoology* **206**: 3557–3567.

Mermall V, Post PL, Mooseker MS (1998) Unconventional myosin in cell movement, membrane traffic, and signal transduction. *Science* **279**: 527–533.

Meyer-Siegler K, Mauro DJ, Seal G, Wurzer J, deRiel JK, Sirover MA (1991) A human nuclear uracil DNA glycosylase is the 37-kDa subunit of glyceraldehyde-3-phosphate dehydrogenase. *Proceedings of the National Academy of Sciences of the United States of America* **88**: 8460–8464.

Michael D, Oren M (2002) The p53 and Mdm2 families in cancer. *Current Opinion in Gentics and Development* **12**: 53–59.

Miller S (2010) Genetic improvement of beef cattle through opportunities in genomics. *Revista Brasileira de Zootecnia* **39S**: 247–255.

Montagné N, Desdevises Y, Soyez D, Toullec JY (2010) Molecular evolution of the crustacean hyperglycemic hormone family in ecdysozoans. *BMC Evolutionary Biology* **10**: 62.

Morrell PL, Buckler ES, Ross-Ibarra J (2012) Crop genomics: advances and applications. *Nature Review Genetics* **13**: 85–96.

Motamed K, Blake DJ, Angello JC, Allen BL, Rapraeger AC, Hauschka SD *et al*. (2003) Fibroblast growth factor receptor-1 mediates the inhibition of endothelial cell proliferation and the promotion of skeletal myoblast differentiation by SPARC: a role for protein kinase A. *Journal of Cellular Biochemistry* **90**: 408–423.

Mullen MP, Berry DP, Howard DJ, Diskin MG, Lynch CO, Giblin L *et al*. (2011) Single nucleotide polymorphisms in the insulin-like growth factor 1 (IGF-1) gene are associated with performance in Holstein-Friesian dairy cattle. *Frontiers in Genetics* **2**: 1–9.

Multhoff G (2006) Heat Shock Proteins in Immunity (ed.) *Handbook of Experimental Pharmacology*, pp. 279–304. Springer-Verlag, Berlin.

Murphy-Ullrich JE (2001) The de-adhesive activity of matricellular proteins: is intermediate of cell adhesion an adaptive state? *Journal of Clinical Investigation* **107**: 785–90.

Mykles DL (1997) Crustacean muscle plasticity: molecular mechanisms determining mass and contractile properties. *Compartive and Biochemistry and Physiology* **117B**: 367–378.

Mykles DL (1988) Histochemical and biochemical characterization of two slow fiber types in decapod crustacean muscles. *Journal of Experimental Zoology* **245**: 232–243.

Nadjar-Boger E, Funkenstein B (2011) Myostatin-2 gene structure and polymorphism of the promoter and first intron in the marine fish *Sparus aurata*: evidence for DNA duplications and/or translocations. *BMC Genetics* **12**: 22.

Nagaraju GPC (2007) Is methyl farnesoate a crustacean hormone? *Aquaculture* **272**: 39–54.

Nakagawa T, Roth W, Wong P, Nelson A, Farr A, Deussing J *et al*. (1998) Cathepsin L: critical role in Ii degradation and CD4 T cell selection in the thymus. *Science* **280**: 450–453.

Nakatsuji T, Lee C, Watson RD (2009) Crustacean molt-inhibiting hormone: structure, function, and cellular mode of action. *Comparative Biochemistry and Physiology* **152A**: 139–148.

Neil DM, Fowler WS, Tobasnik G (1993) Myofibrillar protein composition correlates with histochemistry in fibers of the abdominal flexor muscles of the Norway lobster. *Journal of Experimental Biology* **183**: 185-201.

Ochs RL, Lischwe MA, Spohn WH, Busch H (1985) Fibrillarin: a new protein of the nucleolus identified by autoimmune sera. *Biology of the Cell* **54**: 123–133.

Okumura T, Katsumi A (2001) Effects of bilateral eyestalk ablation on molting and ovarian development in the giant freshwater prawn, *Macrobrachium rosenbergii*. *Journal of Fisheries Science* **67**: 11–25.

Okuno A, Hasegawa Y, Nishiyama M, Ohira T, Ko R, Kurihara M *et al*. (2002) Preparation of an active recombinant peptide of crustacean androgenic gland hormone. *Peptides* **23**: 567–572.

Ola SI, Ai JS, Liu JH, Wang Q, Wang ZB, Chen DY *et al*. (2008) Effects of gonadotrophins, growth gormone, and activin A on enzymatically isolated follicle growth, oocyte chromatin organization, and steroid secretion. *Molecular Reproduction and Development* **75**:89–96.

Ongvarrasopone C, Roshorm Y, Somyong S, Pothiratana C, Petchdee S, Tangkhabuanbutra J *et al*. (2006) Molecular cloning and functional expression of the *Penaeus monodon* 5-HT receptor. *Biochimica et Biophysica Acta* **1759**: 328–339.

Opiyo SO, Moriyama EN (2009) Mining the Arabidopsis and rice genomes for cyclophilin protein families. *International Journal of Bioinformatics Research and Applications* **5**: 295–309.

Ou J, Eng Q, Li Y, Xiu Y, Du J, Gu W *et al*. (2012) Identification and comparative analysis of the *Eriocheir sinensis* microRNA transcriptome response to *Spiroplasma*

eriocheiris infection using a deep sequencing approach. *Fish and Shellfish Immunology* **32**: 345–352.

Palacios R, Gazave E, Goñi J, Piedrafita G, Fernando O, Navarro A *et al*. (2009) Allele-specific gene expression is widespread across the genome and biological processes. *PLoS One* **4**: e4150.

Pan J, Wang W, Song W, Chen J, Li C, Zhao Q (2007) Molecular cloning and expression pattern of myostatin gene in yellow catfish (*Pelteobragus fulvidraco*). *DNA Sequence* **18**: 279–287.

Patterson C (2006) Search and destroy: the role of protein quality control in maintaining cardiac function. *Journal of Molecular and Cellular Cardiology* **40**: 438–441.

Pausch H, Flisikowski K, Jung S, Emmerling R, Edel C, Götz KU *et al*. (2011) Genome-wide association study identifies two major loci affecting calving ease and growth-related traits in cattle. *Genetics* **187**: 289–297.

Pavasovic A, Richardson NA, Mather PB, Anderson AJ (2006) Influence of insoluble dietary cellulose on digestive enzyme activity, feed digestibility and survival in the red claw crayfish, *Cherax quadricarinatus* (von Martens). *Aquaculture Research* **37**: 25–32.

Pavasovic A, Anderson AJ, Mather PB, Richardson NA (2007) Influence of dietary protein on digestive enzyme activity, growth and tail muscle composition in redclaw crayfish, *Cherax quadricarinatus* (von Martens). *Aquaculture Research* **38**: 644–652.

Peng M, Chen X, Zeng D, Ma N, Li Y (2008) SNPs detection of alpha-amylase gene in Pacific white leg shrimp *Litopenaeus vannamei*. *Fisheries Sciecne* **27**: 363–365.

Perry MJ, Tait J, Hu J, White SC, Medler S (2009) Skeletal muscle fiber types in the ghost crab, *Ocypode quadrata*: implications for running performance. *Journal of Experimental Biology* **212**: 673–683.

Prapavorarat A, Pongsomboon S, Tassanakajon A (2010) Identification of genes expressed in response to yellow head virus infection in the black tiger shrimp, *Penaeus monodon*, by suppression subtractive hybridization. *Developmental and Comparative Immunology* **34**: 611–617.

Pisuttharachai D, Yasuike M, Aono H, Murakami K, Kondo H, Aoki T *et al*. (2009) Expressed sequence tag analysis of phyllosomas and hemocytes of Japanese spiny lobster *Panulirus japonicas*. *Fisheries Science* **75**: 195–206.

Pickart CM, Eddins MJ (2004) Ubiquitin: structures, functions, mechanisms. *Biochimica et Biophysica Acta* **1695**: 55–72.

Pollak MN, Schernhammer ES, Hankinson SE (2004) Insulin-like growth factors and neoplasia. *Nature Reviews Cancer* **4**: 505–518.

Pollard TD, Cooper JA (1986) Actin and actin binding proteins: A critical evaluation of mechanisms and functions. *Annual Review Biochemistry* **55**: 987–1035.

Qiu L, Jiang S, Huang J, Wang W, Zhu C, Su T (2009) Molecular cloning and mRNA expression of cyclophilin A gene in black tiger shrimp (*Penaeus monodon*). *Fish and Shellfish Immunology* **26**: 115–121.

Ranford JC, Coates ARM, Henderson B (2000) Chaperonins are cell-signalling proteins: the unfolding biology of molecular chaperones. *Expert Reviews in Molecular Medicine* **201**: 1–17.

Rebbapragada A, Benchabane H, Wrana JL, Celeste AJ, Attisano L (2003) Myostatin signals through a transforming growth factor β-like signaling pathway to block adipogenesis. *Molecular and Cellular Biology* **23**: 7230–7242.

Reinecke M, Björnsson B, Dickhoff WW, McCormick SD, Navarro I, Power DM *et al*. (2005) Growth hormone and insulin-like growth factors in fish: where we are and where to go. *General and Comparative Endocrinology* **142**: 20–24.

Ren Q, Du ZQ, Zhao XF, Wang JX (2009) An acyl-CoA-binding protein (FcACBP) and a fatty acid binding protein (FcFABP) respond to microbial infection in Chinese white shrimp, *Fenneropenaeus chinensis*. *Fish and Shellfish Immunology* **27**: 739–747.

Rinehart JP, Li A, Yocum GD, Robich RM, Hayward SA, Denlinger DL (2007) Up-regulation of heat shock proteins is essential for cold survival during insect diapause. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 11130–11137.

Rodgers BD, Garikipati DK (2008) Clinical, agricultural, and evolutionary biology of myostatin: a comparative review. *Endocrine Reviews* **29**: 513–34.

Rosen O, Manor R, Weil S, Gafni O, Linial A, Aflalo ED *et al*. (2010) A sexual shift induced by silencing of a single insulin-like gene in crayfish: ovarian upregulation and testicular degeneration. *PLoS One* **5**: e15281.

Rössler H, Rieck C, Delong T, Hoja U, Schweizer E (2003) Functional differentiation and selective inactivation of multiple Saccharomyces cerevisiae genes involved in very-long-chain fatty acid synthesis. *Molecular Genetics and Genomics* **269**: 290–298.

Rotllant J, Liu D, Yan YL, Postlethwait JH, Westerfield M, Du SJ (2008) Sparc (Osteonectin) functions in morphogenesis of the pharyngeal skeleton and inner ear. *Matrix Biology* **27**: 561–572.

Rungrassamee W, Leelatanawit R, Jiravanichpaisal P, Klinbunga S, Karoonuthaisiri N (2010) Expression and distribution of three heat shock protein genes under heat shock stress and under exposure to *Vibrio harveyi* in *Penaeus monodon*. *Developmental and Comparative Immunology* **34**: 1082–1089.

Sage H, Bornstein P (1991) Extracellular protein that modulate cell-matrix interaction. SPARC, tenascin, and thrombospondin. *Journal of Biological Chemistry* **266**: 14831–14834.

Sage H, Vernon RB, Funk SE, Everitt EA, Angello J (1989) SPARC, a secreted protein associated with cellular proliferation, inhibits cell spreading in vitro and exhibits Ca^{+2}-dependent binding to the extracellular matrix. *Journal of Cell Biology* **109**: 341–356.

Santiesteban D, Martín L, Arenal A, Franco R, Sotolongo J (2010) Tilapia growth hormone binds to a receptor in brush border membrane vesicles from the hepatopancreas of shrimp *Litopenaeus vannamei*. *Aquaculture* **306**: 338–342.

Schiaffino S, Reggiani C (1996) Molecular diversity of myofibrillar proteins: gene regulation and functional significance. *Physiological Reviews* **76**: 371–423.

Schmid FX, Mayr LM, Mucke M, Schonbrunner ER (1993) Prolyl isomerases-role in protein-folding. *Advances in Protein Chemistry* **44**: 25–66.

Schwartz MW, Woods SC, Porte D Jr, Seeley RJ, Baskin DG (2000) Central nervous system control of food intake. *Nature* **404**: 661–671.

Seeb JE, Carvalho G, Hauser L, Naish K, Roberts S, Seeb LW (2011) Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. *Molecular Ecology Research* **11S**: 1–8.

Sellers JR (2000) Myosins: a diverse superfamily. *Biochimica et Biophysica Acta* **1496**: 3–22.

Sellick GS, Pitchford WS, Morris CA, Cullen NG, Crawford AM *et al*. (2007) Effect of myostatin F94L on carcass yield in cattle. *Animal Genetics* **38**: 440–446.

Serrano L, Blanvillain G, Soyez D, Chairmantier G, Grousset E, Aujoulat F *et al*. (2003) Putative involvement of crustacean hyperglycemic hormone isoforms in the neuroendocrine mediation of osmoregulation in the crayfish *Astacus leptodatylus*. *Journal of Experimental Biology* **206**: 979–988.

Shean BS, Mykles DL (1995) Polyubiquitin in crustacean striated muscle: Increased expression and conjugation during molt-induced claw muscle atrophy. *Biochimica et Biophysica Acta* **1264**: 312–322.

Shendure J, Ji H (2008) Next-generation DNA sequencing. *Nature Biotechnology* **26**: 1135–1145.

Sherman EL, Nkrumah JD, Murdoch BM, Li C, Wang Z, Fu A *et al*. (2008) Polymorphisms and haplotypes in the bovine neuropeptide Y, growth hormone receptor, ghrelin, insulin-like growth factor 2, and uncoupling proteins 2 and 3 genes and their associations with measures of growth, performance, feed efficiency, and carcass merit in beef cattle. *Journal of Animal Science* **86**: 1–16.

Shi XZ, Li XC, Wang S, Zhao XF, Wang JX (2010) Transcriptome analysis of hemocytes and hepatopancreas in red swamp crayfish, *Procambarus clarkia*, challenged with white spot syndrome virus. *Invertebrate Survival Journal* **7**: 119–131.

Shibata M, Ohshima K, Kojima T, Muramoto T, Matsumoto K, Komatsu M *et al*. (2003) Nucleotide sequence of myostatin gene and its developmental expression in skeletal muscles of Japanese black cattle. *Journal of Animal Science* **74**: 383–90.

Silva Gunawardene YIN, Bendena WG, Tobe SS, Chan SM (2003) Comparative immunohistochemistry and cellular distribution of farnesoic acid *O*-methyltransferase in the shrimp and the crayfish. *Peptides* **24**: 1591–1597.

Silverman H, Costello WJ, Mykles DL (1987) Morphological fiber type correlates of physiological and biochemical properties in crustacean muscle. *American Zoologist* **27**: 1011–1019.

Simon CJ, Jeffs A (2011) The effect of dietary carbohydrates on the growth response, digestive gland glycogen and digestive enzyme activities of early spiny lobster juveniles, *Jasus edwardsii*. *Aquaculture Nutrition* **17**: 613–626.

Singh R, Green MR (1993) Sequence-specific binding of transfer RNA by glyceraldehyde-3-phosphate dehydrogenase. *Science* **259**: 365–368.

Sirover MA (2011) On the functional diversity of glyceraldehyde-3-phosphate dehydrogenase: Biochemical mechanisms and regulatory control. *Biochimica et Biophysica Acta* **1810**: 741–751.

Skinner DM (1985) Molting and regeneration. In: Bliss, D.E., Mantel, L.H. (ed) *The Biology of Crustacea*, pp. 43–146. Academic Press, New York.

Söderhäll I, Tangprasittipap A, Liu HP, Sritunyalucksana K, Prasertsan P, Jiravanichpaisal P *et al*. (2006) Characterization of a hemocyte intracellular fatty acid-binding protein from crayfish (*Pacifastacus leniusculus*) and shrimp (*Penaeus monodon*). *FEBS Journal* **273**: 2902–2912.

Sosa MA, Spitzer N, Edwards DH, Baro DJ (2004) A crustacean serotonin receptor: cloning and distribution in the thoracic ganglia of crayfish and freshwater prawn. *Journal of Comparative Neurology* **473**: 526–537.

Spees JL, Chang SA, Mykles DL, Snyder MJ, Chang ES (2003) Molt cycle-dependent molecular chaperone and polyubiquitin gene expression in lobster. *Cell Stress and Chaperones* **3**: 258–264.

Staněk D, Neugebauer KM (2006) The Cajal body: a meeting place for spliceosomal snRNPs in the nuclear maze. *Chromosoma* **115**: 343–354.

Stillman JH, Colbourne JK, Lee CE, Patel NH, Phillips MR, Towle DW *et al*. (2008) Recent advances in crustacean genomics. *Integrative and Comparative Biology* **48**: 852–868.

Stinckens A, Georges M, Buys N (2010) Mutations in the myostatin gene leading to hypermuscularity in mammals: indications for a similar mechanism in fish? *Animal Genetics* **42**: 229–234.

Storch J, Veerkamp JH, Hsu KT (2002) Similar mechanisms of fatty acid transfer from human and roden fatty acid-binding proteins to membranes: liver, intestine, heart muscle and adipose tissue FABPs. *Molecular and Cellular Biochemistry* **239**: 25–33.

Sun C, Shao HL, Zhang XW, Zhao XF, Wang JX (2011a) Molecular cloning and expression analysis of signal transducer and activator of transcription (STAT) from the Chinese white shrimp *Fenneropenaeus chinensis*. *Molecular Biology Reports* **38**: 5313–5319.

Sun PS, Venzon NC Jr, Calderon FRO, Esaki DM (2005) Evaluation of methods for DNA delivery into shrimp zygotes of *Penaeus* (*Litopenaeus*) *vannamei*. *Aquaculture* **243**: 19–26.

Sun PS, Soderlund M, Venzon N Jr, Ye D, Lu Y (2007) Isolation and characterization of two actins of the Pacific white shrimp, *Litopenaeus vannamei*. *Marine Biology* **151**: 2145–2151.

Sun L, Dong X, Fan B, Liu B (2011b) The association of ANKRD2 with loin depth and muscle firmness in pigs. *Journal of Animal Veterinary Advances* **10**: 1462–1468.

Sutherland, JD, Witke W (1999) Molecular genetic approaches to understanding the actin cytoskeleton. *Current Opinion in Cell Biology* **11**: 142–151.

Sweedler JV, Li L, Rubakhin SS, Alexeeva V, Dembrow NC, Dowling O *et al*. (2002) Identification and characterization of the feeding circuit-activating peptides, a novel neuropeptide family of aplysia. *Journal of Neuroscience* **22**: 7797–7808.

Szappanos B, Kovács K, Szamecz B, Honti F, Costanzo M, Baryshnikova A *et al*. (2011) An integrated approach to characterize genetic interaction networks in yeast metabolism. *Nature Genetics* **43**: 656–662.

Tagmount A, Wang M, Lindquist E, Tanaka Y, Teranishi KS, Sunagawa S *et al*. (2010) The porcelain crab transcriptome and PCAD, the porcelain crab microarray and sequence database. *PLoS One* **5**: e9327.

Tahmoorespur M, Taheri A, Gholami H, Ansary M (2011) PCR-SSCP variation of GH and STAT5A genes and their association with estimated breeding values of growth traits in Baluchi sheep. *Animal Biotechnology* **22**: 37–43.

Tamone SL, Prestwich GD, Chang ES (1997) Identification and characterization of methyl farnesoate binding proteins from the crab, *Cancer magister*. *General Comparative Endocrinology* **105**: 168–175.

Tanaka S, Nambu F, Nambu Z (2001) Isolation of a cDNA encoding a putative SPARC from the brine shrimp, *Artemia franciscana*. *Gene* **268**: 53–58.

Tangprasittipap A, Tiensuwan M, Withyachumnarnkul B (2010) Characterization of candidate genes involved in growth of black tiger shrimp *Penaeus monodon*. *Aquaculture* **307**: 150–156.

Tao WJ, Boulding EG (2003) Associations between single nucleotide polymorphisms in candidate genes and growth rate in Arctic charr (*Salvelinus alpinus* L.). *Heredity* **91**: 60–69.

Termine JD, Kleinman HK, Whitson SW, Conn KM, McGarvey ML, Martin GR *et al*. (1981) Osteonectin, a bone-specific protein linking mineral to collagen. *Cell* **26**: 99–105.

Thanh NM, Barnes AC, Mather PB, Li Y, Lyons RE (2010) Single nucleotide polymorphisms in the actin and crustacean hyperglycemic hormone genes and their correlation with individual growth performance in giant freshwater prawn *Macrobrachium rosenbergii*. *Aquaculture* **301**: 7–15.

Tierney AJ (2001) Structure and function of invertebrate 5-HT receptors: a review. *Comparative Biochemistry and Physiology* **128A**: 791–804.

Tisdale EJ (2001) Glyceraldehyde-3-phosphate dehydrogenase is required for vesicular transport in the early secretory pathway. *Journal of Biological Chemistry* **276**: 2480–2486.

Tiu SHK, He JG, Chan SM (2005) Organization and expression study of the shrimp (*Metapenaeus ensis*) putative 5-HT receptor: Up-regulation in the brain by 5-HT. *Gene* **353**: 41–52.

Tiu SHK, Hult EF, Yagi KJ, Tobe SS (2012) Farnesoic acid and methyl farnesoate production during lobster reproduction: possible functional correlation with retinoid X receptor expression. *General Comparative Endocrinology* **175**: 259–269.

Toullec JY, Van Wormhoudt A (1987) Quantitative variations during the intermoult cycle of peptides related to human growth hormone in *Palaemon serratus* (Crustacea; Decapoda). *Comptes Rendus de l'Académie des Sciences-Series III* **305**: 265–269.

Tristan C, Shahani N, Sedlak TW, Sawa A (2011) The diverse functions of GAPDH: Views from different subcellular compartments. *Cellular Signalling* **23**: 317–323.

Tsuchida K, Sunada Y, Noji S, Murakami T, Uezumi A, Nakatani M (2006) Inhibitors of the TGF-beta superfamily and their clinical applications. *Mini-Reviews in Medical Chemistry* **6**: 1255–1261.

Turk B, Turk D, Turk V (2000) Lysosomal cysteine proteinases: more than scavengers. *Biochimica et Biophysica Acta* **1477**: 98–111.

Van der Maarel MJEC, van der Veen B, Uitdehaag JCM, Leemhuis H, Dijkhuizen L (2002) Properties and applications of starch converting enzymes of the α-amylase family. *Journal of Biotechnology* **94**: 137–155.

Van Laere AS, Nguyen M, Braunschweig M, Nezer C, Collette C, Moreau L (2003) A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. *Nature* **425**: 832–836.

Van Wormhoudt A, Sellos DY (1996) Cloning and sequencing analysis of three amylase cDNAs in the shrimp *Pennaeus vannamei* (Crustacae decapoda): evolutionary aspects. *Journal of Molecular Evolution* **42**: 543–551.

Varadaraj K, Kumari SS, Skinner DM (1996) Actin-encoding cDNAs and gene expression during the intermolt cycle of the Bermuda land crab *Gecarcinus lateralis*. *Gene* **171**: 177–184.

Vázquez-Acevedo N, Reyes-Colón D, Ruíz-Rodríguez EA, Rivera NM, Rosenthal J, Kohn AB (2009) Cloning and immunoreactivity of the 5-HT1Mac and 5-HT2Mac receptors in the central nervous system of the freshwater prawn *Macrobrachium rosenbergii*. *Journal of Comparative Neurology* **513**: 399–416.

Venema J, Tollervey D (1999) Ribosome synthesis in *Saccharomyces cerevisiae*. *Annual Review of Genetics* **33**: 261–311.

Ventura T, Manor R, Aflalo ED, Weil S, Raviv S, Glazer L (2009) Temporal silencing of an androgenic gland-specific insulin-like gene affecting phenotypical gender differences and spermatogenesis. *Endocrinology* **150**:1278–1286.

Ventura T, Rosen O, Sagi A (2011) From the discovery of the crustacean androgenic gland to the insulin-like hormone in six decades. *General Comparative Endocrinology* **173**: 381–388.

Verri T, Mandal A, Zilli L, Bossa D, Mandal PK, Ingrosso L *et al*. (2001) D-Glucose transport in decapod crustacean hepatopancreas. *Comparative Biochemistry and Physiology* **130A**: 585–606.

Venkitraman PR, Jayalakshmy KV, Balasubramanian T, Maheswari N, Nair KKC (2004) Effects of eyestalk ablation on molting and growth of penaeid prawn *Metapenaeus dobsoni* (de Man). *Indian Journal of Experimental Biology* **42**: 403–412.

Viitala S, Szyda J, Blott S, Schulman N, Lidauer M, Maki-Tanila A *et al*. (2006) The role of the bovine growth hormone receptor and prolactin receptor genes in milk, fat and protein production in Finnish Ayrshire dairy cattle. *Genetics* **173**: 2151–2164.

Von Reumont BM, Jenner RA, Wills MA, Dell'Ampio E, Pass G, Ebersberger I *et al*. (2012) Pancrustacean phylogeny in the light of new phylogenomic data: support for remipedia as the possible sister group of hexapoda. *Molecular Biology and Evolution* **29**: 1031–1045.

Wade NM, Anderson M, Sellars MJ, Tume RK, Preston NP, Glencross BD (2012) Mechanisms of colour adaptation in the prawn *Penaeus monodon*. *Journal of Experimental Biology* **215**: 343–350.

Wagner GP, Zhang J (2011) The pleiotropic structure of the genotype-phenotype map: the evolvability of complex organisms. *Nature Reviews Genetics* **12**: 204–213.

Wagner KR, Liu X, Chang X, Allen RE (2005) Muscle regeneration in the prolonged absence of myostatin. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 2519–2524.

Wang J, Li ZJ, Lan XY, Hua LS, Huai YT, Huang YZ *et al*. (2010) Two novel SNPs in the coding region of the bovine PRDM16 gene and its associations with growth traits. *Molecular Biology Reports* **37**:571–577.

Wang Q, Li H, Li N, Leng L, Wang Y, Tang Z (2006) Identification of single nucleotide polymorphism of adipocyte fatty acid-binding protein gene and its association with fatness traits in the chicken. *Poultry Science* **85**: 429–434.

Wang X, Meng X, Song B, Qiu X, Liu H (2010) SNPs in the myostatin gene of the mollusk *Chlamys farreri*: Associated with growth traits. *Comparative Biochemistry and Physiology* **155B**: 327–330.

Wang Y, Price SE, Jiang H (2003) Cloning and characterization of the bovine class 1 and class 2 insulin-like growth factor-I mRNAs. *Domestics Animal Endocrinology* **25**: 315–328.

Wang Y, Shu D, Li L, Qu H, Yang C, Zhu Q (2007) Identification of single nucleotide polymorphism of H-FABP gene and its association with fatness traits in chickens. *Asian-Australian Journal of Animal Sciences* **20**: 1812–1819.

Wang YB (2007) Effect of probiotics on growth performance and digestive enzyme activity of the shrimp *Penaeus vannamei*. *Aquaculture* **269**: 259–264.

Wang YH, Huang ML (2010) Organogenesis and tumorigenesis: Insight from the Jak/STAT pathway in the *Drosophila* eye. *Developmental Dynamics* **239**: 2522–2533.

Waters SM, McCabe MS, Howard DJ, Giblin L, Magee DA, MacHugh DE *et al*. (2010) Associations between newly discovered polymorphisms in the *Bos taurus* growth hormone receptor gene and performance traits in Holstein-Friesian dairy cattle. *Animal Genetics* **42**: 39–49.

Watson RD, Lee KJ, Qiu S, Luo M, Umphrey HR, Roer RD *et al*. (2001) Molecular cloning, expression, and tissue distribution of crustacean molt inhibiting hormone. *American Zoologist* **41**: 407–417.

Webster SG, Keller R, Dircksen H (2012) The CHH-superfamily of multifunctional peptide hormones controlling crustacean metabolism, osmoregulation, moulting, and reproduction. *General Comparative Endocrinology* **175**: 217–233.

Welchman RL, Gordon C, Mayer RJ (2005) Ubiquitin and ubiquitinlike proteins as multifunctional signals. *Nature Reviews Molecular Cell Biology* **6**: 599–609.

Westerberg R, Mansson JE, Golozoubova V, Shabalina IG, Backland EC, Tvrdik P *et al*. (2006) ELOVL3 is an important component for early onset of lipid recruitment in brown adipose tissue. *Journal of Biological Chemistry* **281**: 4958–4968.

Wheeler MB, Monaco E, Bionaz M, Tanaka T (2010) The role of existing and emerging biotechnologies for livestock production: toward holism. *Acta Scientiae Veterinariae* **38S2**: 463–484.

Williams KC (2007) Nutritional requirements and feeds development for post-larval spiny lobster: a review. *Aquaculture* **263**: 1–14.

Williamson M, Lenz C, Winther AM, Nassel DR, Grimmelikhuijzen CJ (2001) Molecular cloning, genomic organization, and expression of a B-type (cricket-type) allatostatin preprohormone from *Drosophila melanogaster*. *Biochemical and Biophysical Research Communications* **281**: 544–550.

Windish WW, Mhatre NS (1965) Microbial amylases. In; Wayne WU, editor.Advances in *Applied Microbiology* **7**: 273–304.

Wongsurawat T, Leelatanawit R, Thamniemdee N, Uawisetawthana U, Karoonuthaisiri N, Menasveta P *et al*. (2010) Identification of testis-relevant genes using in silico anaylsis from testis ESTs and cDNA microarrary in the black tiger shrimp (*Penaeus monodon*). *BMC Molecular Biology* **11**: 35.

Worm B, Barbier EB, Beaumont N, Duffy JE, Folke C, Halpern BS *et al*. (2006) Impacts of biodiversity loss on ocean ecosystem services. *Science* **314**: 787–790.

Wu HH, Ivkovic S, Murray RC, Jaramillo S, Lyons KM, Johnson JE *et al*. (2003) Autoregulation of neurogenesis by GDF11. *Neuron* **37**: 197–207.

Wu P, Qi D, Chen L, Zhang H, Zhang X, Qin JG *et al*. (2009) Gene discovery from an ovary cDNA library of oriental river prawn *Macrobrachium nipponense* by ESTs annotation. *Comparative Biochemistry and Physiology* **4D**: 111–120.

Wu Q, Brown MR (2006) Signaling and function of insulin-like peptides in insects. *Annual Review Entomology* **51**: 1–24.

Xiang J, Wang B, Li F (2008) Generation and analysis of 10,443 ESTs from cephalothorax of *Fenneropenaeus chinensis*. *Institue of Electrical and Electronics Engineers* 74–80.

Xu B, Zhang PJ, Mai YL, Miao HZ (2000) Studies of the effects of recombinant fish growth hormone on survival and growth enhancement of Chinese prawn *Penaeus chinensis*. *Marine Sciences* **24**: 54–56.

Xu C, Wu G, Zohar Y, Du SJ (2003) Analysis of myostatin gene structure, expression and function in zebrafish. *Journal of Experimental Biology* **206**: 4067–4079.

Xu L, Jiang H, Chen H, Gu Z (2011) Genetic architecture of growth traits revealed by global epistatic interactions. *Genome Biology and Evolution* **3**: 909–914.

Xu Q, Liu Y, Liu R (2010) Expressed sequence tags from cDNA library prepared from gills of the swimming crab, *Portunus tribuberculatus*. *Journal of Experimental Biology and Ecology* **394**: 105–115.

Yamada H, Nishimura F, Naruishi K, Chou HH, Takashiba S, Albright GM *et al*. (2000) Phenytoin and cyclosporin A suppress the expression of MMP-1, TIMP-1and cathepsin, but not cathepsin B in cultured gingival fibroblasts. *Journal of Periodontology* **71**: 955–960.

Yamada T, Sasaki S, Sukegawa S, takahagi Y, Morita M, Murakami H *et al*. (2011) Possible association of single nucleotide polymorphism in Titin gene with growth-related trait in Japanese black beef cattle. *Journal of Animal Veterinary Advances* **10**: 2603–2606.

Yamao F (1999) Ubiquitin system: selectivity and timing of protein destruction. *Journal of Chemistry* **125**: 223–229.

Yazawa R, Watanabe K, Koyama T, Ruangapan L, Tassanakajon A, Hirono I *et al*. (2005) Development of gene transfer technology for black tiger shrimp, *Penaeus monodon*. *Journal of Experimental Zoology* **303A**: 1104–1109.

Ye HQ, Chen SL, Sha ZX, Liu Y (2007) Molecular cloning and expression analysis of the myostatin gene in sea perch (*Lateolabrax japonicus*). *Marine Biotechnology* **9**: 262–272.

Yeh HY, Klesius PH (2008) Channel catfish, *Ictalurus punctatus*, cyclophilin A and B cDNA characterization and expression analysis. *Veterinary and Immunology Immunopathology* **121**: 370–377.

Yodmuang S, Udomkit A, Treerattrkool S, Panyim S (2004) Molecular and biological characterization of molt-inhibiting hormone of *Penaeus monodon*. *Journal of Experimental Marine Biology and Ecology* **312**: 101–114.

You XY, Liu YP, Zhu Q, Yang ZQ (2007) Study on SNP of the H-FABP gene and its association with slaughter performance in chicken. *Journal of Hereditas* **29**: 230–234.

Yu M, Cheng Y, Rothschild MF (2006) SNP analysis of molting related genes in *Penaeus monodon* and *Litopenaeus vannamei* shrimp. *Archiv Tierzucht Dummerstorf* **49**: 411–412.

Yu SL, Chung HJ, Sang BC, Bhuiyan MSA, Yoon D, Kim KS *et al*. (2007) Molecular cloning and characterization of bovine HMGA1 gene. *Asian-Australian Journal of Animal Sciences* **20**: 1662–1669.

Zhang C, Liu Y, Huang K, Zeng W, Xu D, Wen Q *et al*. (2011a) The association of two single nucleotide polymorphisms (SNPs) in growth hormone (GH) gene with litter size and superovulation response in goat-breeds. *Genetics and Molecular Biology* **34**: 49–55.

Zhang G, Ding F, Wang J, Dai G, Xie K , Zhang L, Wang W *et al*. (2011b) Polymorphism in exons of the myostatin gene and its relationship with body weight traits in the Bian chicken. *Biochemical Genetics* **49**: 9–19.

Zhang W, Wan H, Jiang H, Zhao Y, Zhang X, Hu S *et al*. (2011c) A transcriptome analysis of mitten crab (*Eriocheir sinensis*). *Genetics and Molecular Biology* **34**: 136–141.

Zhang Y, Benton JL, Beltz BS (2011d) 5-HT receptors mediate lineage-dependent effects of serotonin on adult neurogenesis in *Procambarus clarkia*. *Neural Development* **6**: 2.

Zhang ZR, Liu YP, Yao YG, Jiang XS, Du HR, Zhu Q (2009) Identification and association of the single nucleotide polymorphisms in calpain3 (*CAPN3*) gene with carcass traits in chickens. *BMC Genetics* **10**: 10.

Zhao D, Song S, Wang Q, Zhang X, Hu S, Chen L (2009) Discovery of immune-related genes in Chinese mitten crab (*Eriocheir sinensis*) by expressed sequence tag analysis of haemocytes. *Aquaculture* **287**: 297–303.

Zhao WM, Zhao RX, Qiao N, Xu Q, Huang ZY, Li X (2011) Association of GH polymorphism with growth traints in goose. *Journal of Animal and Veterinary Advances* **10**: 692–697.

Zhou H, Mitchell AD, McMurtry JP, Ashwell CM, Lamont SJ (2005) Insulin-like growth factor-I gene polymorphism associations with growth, body composition, skeleton integrity, and metabolic traits in chickens. *Poultry Science* **84**: 212–219.

Zhou JP, Zhu XP, Zhang W, Qin F, Zhang SW, Jia ZH (2011) A novel single-nucleotide polymorphism in the 5' upstream region of the prolactin receptor gene is associated with fiber traits in Liaoning cashmere goats. *Genetics and Molecular Research* **10**: 2511–2516.

Zhou J, Wang WN, He WY, Zheng Y, Wang L, Xin Y *et al*. (2010) Expression of HSP60 and HSP70 in white shrimp, *Litopenaeus vannamei* in response to bacterial challenge *Journal of Invertebrate Pathology* **103**: 170–178.

Zhu XJ, Dai ZM, Liu J, Yang WJ (2005) Actin gene in prawn, *Macrobrachium rosenbergii*: Characteristics and differential tissue expression during embryonic development. *Comparative Biochemistry and Physiology* **140B**: 599–605.

Zimmerman AW, Veerkamp JH (2002) New insights into the structure and function of fatty acid-binding proteins. *Cellular and Molecular Life Sciences* **11**: 1096–1116.

Zmora N, Trant J, Zohar Y, Chung JS (2009a) Molt-inhibiting hormone stimulates vitellogenesis at advanced ovarian developmental stages in the female blue crab, *Callinectes sapidus* 1: an ovarian stage dependent involvement. *Saline Systems* **5**: 7.

Zmora N, Sagi A, Zohar Y, Chung JS (2009b) Molt-inhibiting hormone stimulates vitellogenesis at advanced ovarian developmental stages in the female blue crab, *Callinectes sapidus* 2: novel specific binding sites in hepatopancreas and cAMP as a second messenger. *Saline Systems* **5**: 6.

Zou Z, Zhang Z, Wang Y, Han K, Fu M, Lin P (2011) EST analysis on the gonad development related organs and microarray screen for differentially expressed genes in mature ovary testis of *Scylla paramamosain*. *Comparative Biochemistry and Physiology* **6D**: 150–157.

TABLES and FIGURES

Table 1. Representation of crustacean species in the public databases (at 31/07/12) over 2000 ESTs.

Genus name	Number of ESTs	Number of sequence read archive (SRA)	Number of whole genome sequences
Daphnia	174,436	0	2
Eriocheir	17,080	0	0
Euphausia	6,956	1	1*
Fenneropenaeus	11,278	0	1*
Homarus	51,910	0	0
Litopenaeus	163,387	2	1*
Macrobrachium	12,935	2	0
Marsupenaeus	3,843	0	0
Panulirus	2,744	0	0
Penaeus	52,234	0	0
Petrolisthes	97,806	0	1*
Portunus	14,506	0	1*
Scylla	3,892	0	0

*Unpublished data deposited in GenBank.

Table 2. Genomics resources available since 2008 for commercially important crustaceans.

Order	Species (Common name)	Tissue library	Targeting	Study approaches[a]	Genomic data types[b]	Total # of ESTs	Reference
Cladocera	*Daphnia pulex* (Water flea)	Whole genome	Ecoresponsive genome	T, N	E, G	1,554,564	Colbourne *et al*. 2011
Decapoda	*Eriocheir sinensis* (Chinese mitten crab)	Hepatopancreas	Digestive & immune	T	E	3,297	Jiang *et al*. 2009a
	E. sinensis	Hepatopancreas & Testis	Nutrition & reproduction	T	E	6,287	Jiang *et al*. 2009b
	E. sinensis	Haemocyte	Immune	T	E	7,535	Gai *et al*. 2009
	E. sinensis	Haemocyte	Immune	T	E	3,041	Zhao *et al*. 2009
	E. sinensis	Testis	Reproduction	T	E	2,990	Zhang *et al*. 2011c
	E. sinensis	Haemocyte	Immune	N	E, M	28,348,633	Ou *et al*. 2012
	Portunus trituberculatus (Swimming crab)	Gill	Salinity adaptation	T	E	4,433	Xu *et al*. 2010
	Scylla paramamosain (Mud crab)	Testis and ovary	Reproduction	T	E	5,169	Zou *et al*. 2011
	Petrolisthes cinctipes (Porcelain Crab)	Heart, gill, nerve, muscle, whole crab	Thermal adaptation & climate change	T	E	122,495	Tagmount *et al*. 2010

	Procambarus clarkia (Red swamp crayfish)	Haemocyte	Immune	T	E	9,115	Shi *et al*. 2010
	Cherax quadricarinatus (Redclaw crayfish)	Haematopoietic tissue	Immune	T	E	2,123	Liu *et al*. 2011
	Panulirus japonicas (Japanese spiny lobster)	Phyllosoma & haemocyte	Immune	T	E	2,673	Pisuttharachai *et al*. 2009
	Fenneropenaeus chinensis (Fleshy prawn)	Cephalothorax	Growth	T	E	10,446	Xiang *et al*. 2008
	Macrobrachium nipponense (Oriental rive prawn)	Ovary	Reproduction	T	E	3,256	Wu *et al*. 2009
	M. rosenbergii (Giant freshwater prawn)	Muscle, ovary, testis	Growth & reproduction	N	E	787,731	Jung *et al*. 2011
	Penaeus monodon (Black tiger shrimp)	Testis	Reproduction	T	E	4,803	Wongsurawat *et al*. 2010
	Penaeid shrimps	Multi libraries	Recombined available ESTs & cDNAs	T	E	270,000	Leekitcharoenphon *et al*. 2010
	Penaeid shrimps	Multi libraries	Recombined available ESTs & cDNAs	T	E	200,000	Leu *et al*. 2011
	Pandalus latirostris	Ovary	Reproduction	N	E	181,947,878	Kawahara-Miki *et*

	(Hokkai shrimp)						*al*. 201
Euphausiacea	*Eupaausia superb* (Antarctic krill)	Whole krill	Stress	N	E	699,248	Clark *et al*. 2011

[a]For study approaches: T, traditional approaches including suppression subtractive hybridization (SSH) and cloning; N, next generation sequencing (Illumina, IonTorrent, Roche, SOLiD)

[b]For genomic data types: E, EST libraries including cDNAs and transcriptome; G, nuclear genome; M, microRNA.

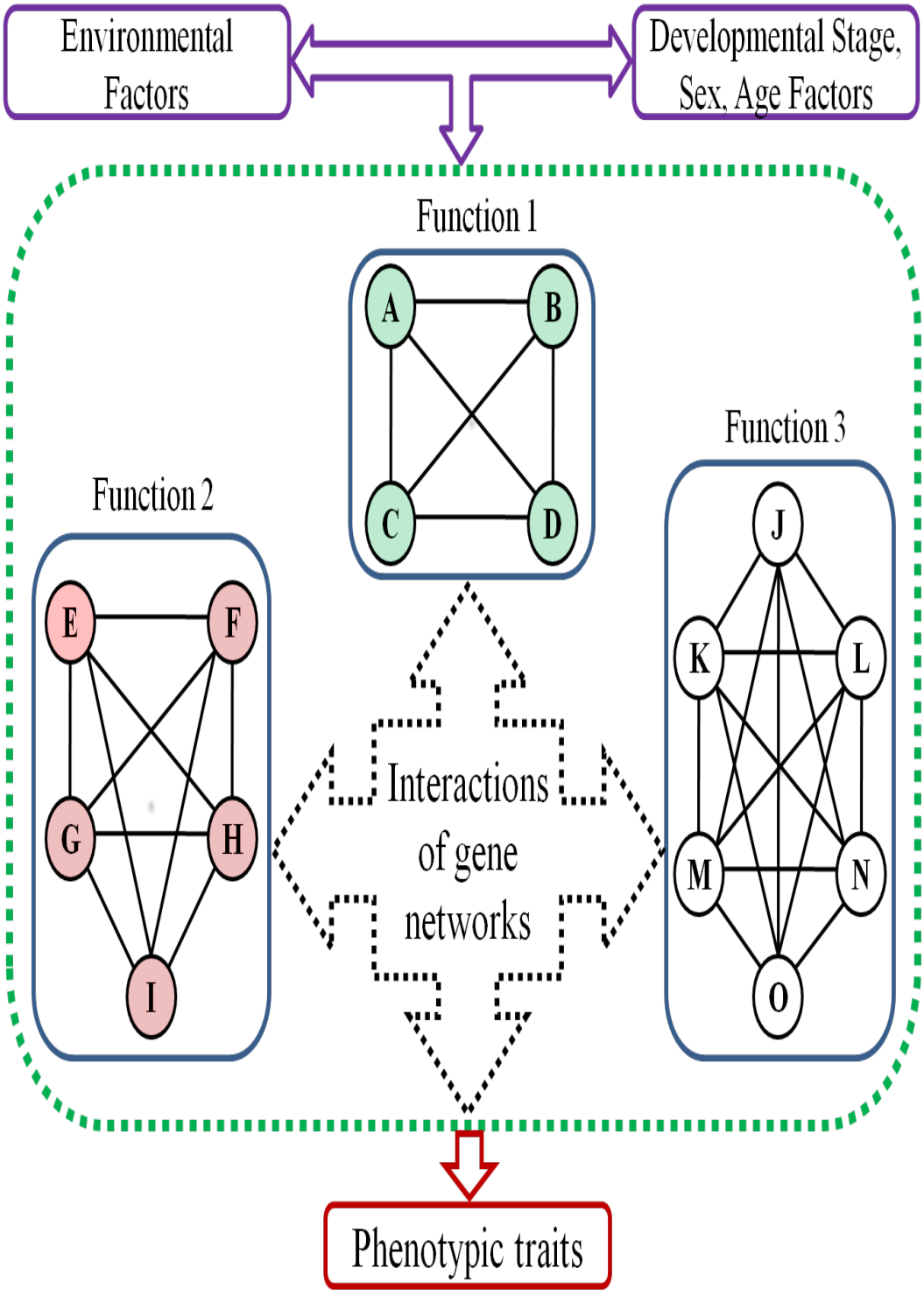


Figure 1. Simplified diagram describing gene networks. Environmental factors can be from inner or outer cells. Green dot box represents cell boundary. Small circles indicate individual gene and solid black lines in each blue box indicate epistatic gene-gene interactions, direct and/or indirect for each function. Black dot box indicates direct and/or indirect interactions of gene networks including pleiotropic interactions.

Table 3. DNA variants and their relationships in candidate loci involved in regulation of growth in crustaceans.

Gene name (symbol)	Species	DNA variant	Effect on trait	Reference
5-Hydroxytryptamine receptor (5HT1R)	*L. vannamei*	Exon2_A88G (Syn) Exon2_C109T (Syn) (Trend) Exon2_C395G (Syn) (Trend) Exon2_C398G (Syn)	Tended to be associated with increased body weight	Marti *et al*. 2010.
Actin	*M. rosenbergii*	All synonymous	All synonymous so no further statistic analysis made.	Thanh *et al*. 2010.
Alpha-amylase (AMY2)	*L. vannamei* *P. monodon*	Intron2_G351A	Suggested no significant association within studied populations	Glenn *et al*. 2005.
Cathepsin-L (CTSL)	*L. vannamei* *P. monodon*	Intron3_G178C Intron3_C681G (Trend)	Tended to be associated with higher body weight in one population.	Glenn *et al*. 2005.
Crustaean hyperglycemic hormone (CHH)	*M. rosenbergii*	IntronCHH3_G2402T (All three traits) IntronCHH3_G2561A (All three traits) IntronCHH3_G2407A (Body weight) IntronCHH3_G2409A (Body weight)	Four intronic SNPs were associated significantly with growth traits (body weight, carapace length, standard length). Four SNPs were in LD, and haplotype TGAA had significant associations with high growth.	Thanh *et al*. 2010.
	L. vannamei	CHH Intron1_A153G CHH Exon1_G59A (Non-Syn)	Suggested no significant effect on body weight within studied populations.	Yu *et al*. 2006.
Molt-inhibiting hormone (MIH)	*L. vannamei*	Exon1_A171G (Syn) Exon_T480G (Syn)	Significant association with growth traits (body weight) in MIH Exon1 (A171G) SNP but not in Exon2 (T480G) SNP.	Li *et al*. 2011b.

Signal transducer and activator of transcription (STAT)	*L. vannamei*	Intron12_A165G	All homozygous so no further association studies conducted.	Marti *et al*. 2010.
Unnamed genes	*L. vannamei*	†SYG48, 239, 273, 329, 379, 486 ‡SYG416, 510	Significant and suggested association with weekly gain, grow-out survival, and pathogen resistance.	Ciobanu *et al*. 2010a

†EST-SNP IDs from single marker association analysis. No further annotation is available due to unknown genes.

‡EST-SNP IDs from haplotype association analysis. No further annotation is available due to unknown genes.

Table 4. Summary of association studies between growth and candidate gene polymorphisms in livestock.

Gene name (symbol)	Common name (reference)	General functions	Polymorphisms and growth association	Putative genes (crustacean species)	Crustacean species reference
Ankyrin repeat domain 2 (ANKRD2)	Pig (Sun *et al*. 2011b)	ANK, a family of adapter molecules mediating linkages between integral membrane and cytoskeletal proteins Muscle ankyrin repeat proteins (MARPs) family, essential for cardiogenesis and cardiac hypertrophy (CARP), skeletal muscular stretch and development (ANKRD2/ARPP) and during recovery following starvation (DARP).	ANKRD2 SNP showed significant association with growth and carcass traits (porcine loin depth and meat firmness).	ANK repeat and mynd domain-containing protein2, ANKRD11, ANKRD-containing protein12, 17 (*M. rosenbergii*)	Jung *et al*. 2011
				ANK-like gene (*Eriocheir japonica sinensis*)	Li *et al*. 2009
Calcitonin receptor (CALCR)	Pig (Alexander *et al*. 2010)	Involved in regulating calcium homeostasis and osteoclast-mediated bone resorption.	CALCR SNP showed bone integrity and its response to dietary P restriction.	CALCR, Calcitonin gene-related peptide type 1 receptor (*M. rosenbergii*)	Jung *et al*. 2011
	Cattle (Magee *et al*. 2010)		CALCR SNPs were associated with milk traits and body condition scores. One SNP in CALCR was negatively associated with calving interval.	Calcitonin-like (*Callinectes sapidus*)	Cameron & Thomas 1992
				Calcitonin gene-related peptide (*Nephrops norvegicus*)	Arlot-Bonnemains *et al*. 1991
Calpain (CAPN)	Beef (Allais *et al*. 2011)	An important protease that hydrolyzes proteins in myofibrils	μ-Calpain (CAPN1) SNP had a significant effect on shear force and meat tenderness.	Muscle-specific calpain, CAPNb, c, t, 5, 7 (*M. rosenbergii*)	Jung *et al*. 2011
	Beef (Hou *et al*. 2011)	Involved in meat tenderness.	CAPN1 SNPs were significantly related with marbling and meat tenderness.	Calpain-like protease (*N. norvegicus*)	Gornik *et al*. 2010
	Chicken (Zhang		Calpain3 (CAPN3) SNPs showed	Muscle-specific calpain	Medler *et al*.

	et al. 2009)	Involved in muscle growth and development as proenzymes that are regulated by $Ca2^{+}$ binding and autoproteolytic modification	significant associations with body weight, carcass weight, breast weight, and leg muscle weight.	(*Homarus americnus*)	2007
				CAPNb, m, t (*Gecarcinus lateralis*)	Kim *et al*. 2005
High mobility group (HMG)	Pig (Makgahlela *et al*. 2009)	Involved in changing of DNA structure and chromatic organization by DNA activities such as transcription, replication, and recombination.	Associations between high mobility group AT-look1 (HMGA1) genotypes with fat deposition traits were mostly suggested.	HMG, HMG2-like, HMG non-histone (*M. rosenbergii*)	Jung *et al*. 2011
	Cattle (Yu *et al*. 2007)	Involved in growth, fatness, and lean meat in pigs.	The bovine map location showed that the HMGA1 gene and SNP was closely located with a previously identified meat quality QTL region indicating this gene is the most likely positional candidate for meat quality.	High mobility group box (*Litopenaeus vannamei*)	Chen *et al*. 2011
	Pig (Kim *et al*. 2006)		HMGA1 gene polymorphism was related to growth, fatness and lean meat content.		
PR domain containing 16 (PRDM16)	Chicken (Han *et al*. 2012)	A member of PRDM family known to participate in the development of mammalian nervous system, control cell proliferation/differentiation, and control a robust, bi-directional switch in cell fate between skeletal myoblasts and brown fat cells.	PRDM16 SNPs were significantly associated with growth, fatness and meat quality traits.	PRDM5, 14 (*M. rosenbergii*)	Jung *et al*. 2011
	Goat (Chen *et al*. 2010)		PRDM16 SNPs had remarkable effects on growth traits, body length, chest circumference, body weight, body height, and thicker cannon circumference.		
	Cattle (Wang *et al*. 2010)	Displayed an abnormal morphology, reduced thermogenic gene expression and elevated expression of muscle-specific genes in PRDM16-deficient brown fat.	PRDM16 SNPs were significantly associated with growth traits (body weight and average daily gain).		
Prolactin	Duck (Li *et al*.	Essential role for the initiation and	SNP had significant effect on egg	Prolactin regulatory element-	Jung *et al*.

(PRL)	2009)	maintenance of lactation, osmoregulation and egg production as a polypeptide hormone secreted by the anterior pituitary gland.	weight and double-yolk percentage.	binding (*M. rosenbergii*)	2011
	Cattle (Alipanah *et al*. 2007)		PRL SNP was significantly associated with milk yield and milk fat yield.		
Prolactin receptor (PRLR)	Cattle (Lü *et al*. 2011)	Involved in reproduction as a specific receptor for PRL.	SNP was significantly associated with growth traits (hucklebone width, body weight, and average daily gain).	Prolactin regulatory element-binding (*M. rosenbergii*)	Jung *et al*. 2011
	Goat (Zhou *et al*. 2011)	Involved in all action of PRL and growth hormone (GH) as receptors.	PRLR SNP was significantly associated with cashmere fibre weight and diameter.		
RAS protein-specific guanine nucleotide-releasing factor 1 gene (RASGRF1)	Cattle (Magee *et al*. 2010)	Signal transduction, and cellular proliferation and differentiation. RAS, role in growth and development.	RASGRF1 SNP was associated with a reduction in milk protein percentage, an increase in somatic cell score.	RAS associated domain-containing protein2, RAS-like protein2, RAS-related protein rab-2a, 7a, 11b, RAS-related protein rap-1b precursor (*M. rosenbergii*)	Jung *et al*. 2011
				RAS-like protein (*Penaeus japonicas*)	Huang & Chuang 1998
Tetraspanin 32 (TSPAN32)	Cattle (Magee *et al*. 2010)	Possible tumour-suppressor Functions. Role in immunological signalling, growth regulation, cell motility, viral infections and membrane architecture.	TSPAN32 SNP was associated with culled cow carcass weight.	Tetraspanin 26a, 96f (*M. rosenbergii*)	Jung *et al*. 2011
				Tetrspanin D107 (*Penaeus monodon*)	Prapavorarat *et al*. 2010
				Tetraspanin family protein (*Panulirus japonicas*)	Pisuttharachai *et al*. 2009
				Tetraspanin 11 (*Cherax quadricarinatus*)	Liu *et al*. 2011
Titin (TTN)	Cattle (Yamada *et al*. 2011)	Involved in myofibrillogenesis. Constitute an elastic matrix in striated	TTN SNP was marginally associated with rib eye area suggesting possible effect on the growth-related trait.	Titin-like protein (*M. rosenbergii*)	Jung *et al*. 2011
				Titin (*Palaemonetes*	Li & Brouwer

		muscle sacromeres, smooth muscle, and some non-muscle cells as a family of giant structural proteins.		*pugio*)	2009
Transforming growth factor beta type 1 receptor (TGFBR1)	Pig (Chen *et al*. 2012)	Involved in TGF-β mediated cell growth and differentiation. Associated with growth traits and reproduction traits	TGFBR1 SNPs were significantly associated with growth rates (daily gains) and carcass traits (loin-eye-area, lean percentage, muscle colour).	Transforming growth beta, Transforming growth factor beta regulator 1 (*M. rosenbergii*)	Jung *et al*. 2011
Transforming growth factor beta type 2 receptor (TGFBR2)	Chicken (Bennett *et al*. 2007)	Associated with growth, development and bone strength. Mutations in the TGFBR1, 2 genes involved in cardiovascular, craniofacial, neurocognitive, and skeletal development.	TGFBR1 SNP was significantly associated with bone mineral density and content, body weight, and egg production.	Transforming growth beta, Transforming growth factor beta regulator 1 (*M. rosenbergii*)	Jung *et al*. 2011

CAHPTER 3: Transcriptomics of a Giant Freshwater Prawn (*Macrobrachium rosenbergii*): *de novo* assembly, annotation and marker discovery

- Presented at the 2012 World Aquaculture Society 'Global aquaculture securing our future' 1-5 September 2012, Prague, Czech Republic.

- Published in PLoS One (2011). PLoS One (6) pages e27938.

Preface to Chapter 3

This study was initiated to characterise its transcriptome and identify genes important for growth using 454 pyrosequencing because little is known about *M. rosenbergii* genome. A collection of 787,731 sequence reads (244.37 Mb) obtained from 454 pyrosequencing analysis of cDNA prepared from muscle, ovary and testis tissues taken from 18 adult prawns was assembled into 123,534 expressed sequence tags (ESTs). The majority of sequences matched the crustacne and insect sequences. Amongst the transcripts, several growth-related candidate genes, domains and KEGG pathways were derived including 834 single nucleotide polymorphisms and 658 simple sequence repeats motifs. The *M. rosenbergii* transcriptome data reported here should provide an invaluable resource for improving our understanding of this species' genome structure and biology. The data will also instruct future functional studies to manipulate or select for genes influencing growth that should find practical applications in aquaculture breeding programs.

Transcriptomics of a Giant Freshwater Prawn (*Macrobrachium rosenbergii*): *de novo* assembly, annotation and marker discovery.

Jung H[1*], Lyons RE[2], Dinh H[1], Hurwood DA[1], McWilliam S[2], Mather PB[1]

[1]Biogeosciences, Queensland University of Technology, Brisbane, Queensland 4001, Australia

[2]Commonwealth Scientific and Industrial Research Organisation Livestock Industries, St Lucia, Queensland 4067, Australia

ABSTRACT

Background: Giant freshwater prawn (*Macrobrachium rosenbergii* or GFP), is the most economically important freshwater crustacean species. However, as little is known about its genome, 454 pyrosequencing of cDNA was undertaken to characterise its transcriptome and identify genes important for growth.

Methodology and Principal Findings: A collection of 787,731 sequence reads (244.37 Mb) obtained from 454 pyrosequencing analysis of cDNA prepared from muscle, ovary and testis tissues taken from 18 adult prawns was assembled into 123,534 expressed sequence tags (ESTs). Of these, 46% of the 8,411 contigs and 19% of 115,123 singletons possessed high similarity to sequences in the GenBank non-redundant database, with most significant (E value < $1e^{-5}$) contig (80%) and singleton (84%) matches occurring with crustacean and insect sequences. KEGG analysis of the contig open reading frames identified putative members of several biological pathways potentially important for growth. The top InterProScan domains detected included RNA recognition motifs, serine/threonine-protein kinase-like domains, actin-like families, and zinc finger domains. Transcripts derived from genes such as actin, myosin heavy and light chain, tropomyosin and troponin with fundamental roles in muscle development and construction were abundant. Amongst the contigs, 834 single nucleotide polymorphisms, 1198 indels and 658 simple sequence repeats motifs were also identified.

Conclusions: The *M. rosenbergii* transcriptome data reported here should provide an invaluable resource for improving our understanding of this species' genome structure and biology. The data will also instruct future functional studies to manipulate or select for genes influencing growth that should find practical applications in aquaculture breeding programs.

INTRODUCTION

Of the 200 or so aquaculture species, decapod crustaceans including prawns, lobsters and crabs contribute substantially to the US$60 billion global industry [1]. Amongst farmed crustaceans, the giant freshwater prawn (*Macrobrachium rosenbergii*) has increasingly become an aquaculture species of major commercial value, with revenue in Asia alone currently worth >US$1 billion annually (FAO 2009; Schwantes *et al*. 2009; Thanh *et al*. 2010a; Thanh *et al*. 2010b). Due to its high value, research is now focusing on improving the growth performance of farmed *M. rosenbergii* (Nhan *et al*. 2009; Schwantes *et al*. 2009; Thanh *et al*. 2009; Thanh *et al*. 2010a; Thanh *et al*. 2010b). However, little is known about this species' basic biology and genome make-up so that they can be exploited to improve farm productivity of this species.

Genomics approaches are now being applied widely to elucidate genetic factors conferring economically significant traits and/or phenotypes and to manage genetic diversity in cultured crustacean species (Staelens *et al*. 2008; Robalino *et al*. 2009; Wu *et al*. 2009; Du *et al*. 2010). Whilst their application to cultured fish species has produced significant production gains, such gains are only beginning to be realized in penaeid species (Tassanakajon *et al*. 2006; Lyons *et al*. 2007; Hamasaki & Kitada 2008; Leekitcharoenphon *et al*. 2010; Leu *et al*. 2011), and no detailed genetic analyses have yet been reported for *M. rosenbergii*. Such basic information is essential to better understand a species' biology and to devise strategies to improve productivity in culture. DNA microsatellites (Chand *et al*. 2005; Divu *et al*. 2008; See *et al*. 2009) and mitochondrial DNA sequence comparisons (Miller *et al*. 2005) have been used to examine the phylogeography of *M. rosenbergii* (de Bruyn *et al*. 2004; de Bruyn *et al*. 2007) sampled from Asia and northern Australia and genes potentially associated with pathogen defence responses (Baruah *et al*. 2009; Cam *et al*. 2009; Sung *et al*. 2009) and sexual maturation traits (Ngernsoungnern *et al*. 2009) have also been identified. However, more genome-wide or transcriptome-wide datasets have yet to be generated as a basis for functional genomics approaches (Vera *et al*. 2008; Wheat 2008; Bai *et al*. 2011a; Bai *et al*. 2011b) aimed at improving the aquaculture performance of this species.

Roche 454 Genome Sequencing FLX technology is particularly useful as a shotgun method for generating data broadly across novel genomes, and it is relatively cheap (Emrich *et al*. 2007; Vera *et al*. 2008; Parchman *et al*. 2010) and exceptionally accurate (Margulies *et*

al. 2005; Emrich *et al*. 2007; Vera *et al*. 2008; Wheat 2008; Bai *et al*. 2011a; Bai *et al*. 2011b). Here it was used to characterize the transcriptome of *M. rosenbergii* using cDNA prepared from mRNA isolated from muscle, ovary and testis tissues. Expressed sequence tag (EST) sequences generated were assembled and annotated with putative functions where possible, and database searches were performed to identify candidate protein domains, genes and gene families potentially involved with growth. A variety of markers potentially useful for genomic population studies including simple sequence repeats (SSRs) located within coding regions and single nucleotide polymorphisms (SNPs) detected amongst deep coverage sequence regions reads are also reported.

MATERIALS AND METHODS

Tissue samples

M. rosenbergii with variable growth phenotypes were sampled from cohorts that were reared in a GFP stock improvement program in Vietnam (Dinh *et al*. 2012). Muscle and ovary tissue was sampled from adult females from high and low growth performance families and tissues preserved in 95% ethanol (454 sequencing run #1). Muscle was not sampled from males as their growth performance is confounded by social factors (Thanh *et al*. 2009). Muscle and ovary tissue from adult females and testis and eye-stalk tissue from adult males preserved in RNAlater (Ambion) were also analysed (454 sequencing run #2).

RNA extraction

In 454 sequencing run #1, TRIzol® reagent (Invitrogen) (Chromczynski & Mackey 1995) was used to extract total RNA from either muscle tissue or ovary tissue pooled from the three heaviest females from the high growth performance cohort and from the three lightest females from the low growth performance cohort. In 454 sequencing run #2, total RNA was extracted similarly from muscle/ovary (female) and testis/eye-stalk (male) from groups of three prawns in the same growth categories as used in 454 sequencing run #1. Total RNA was purified further using a RNA Easy Kit (QIAGEN). RNA yields and quality were checked using both a Bioanalyzer nanochip (Agilent) and a Nanodrop spectrophotometer (Thermo). Equal amounts of total RNA purified from each tissue type were pooled and mRNA was

isolated using the MicroPoly(A) Purist^{TM} Kit (Ambion) according to the manufacturer's protocol.

Library construction and 454 pyrosequencing

mRNA purified from pooled muscle, ovary, testis and eye-stalk total RNA from males and females of high and low growth performance were sent to the Australian Genome Research Facility (AGRF), Brisbane, Australia, for cDNA synthesis using a cDNA Rapid Library Preparation Kit (Roche) and subjected to 454 GS-FLX sequence analysis. Due to issues with poor RNA and cDNA quality and low yields from eyestalk tissue, this tissue was excluded from the cDNA library. The cDNA library sequenced thus comprised a pool of cDNAs prepared from muscle tissue from the three heaviest females, ovary tissue from the three heaviest and three lightest females and testis tissue from the three heaviest males. Each cDNA was normalized prior to pooling to reduce sequence coverage of high copy number mRNAs and samples tagged for downstream identification. cDNA yields were quantified using a Quant-iT RiboGreen fluorometer (Invitrogen) and average lengths were determined by analysis of an aliquot (1 µl) on the Bioanalyzer (Agilent) using a LapChip 7500. Oligonucleotide adapters A and B (Roche) were ligated to cDNA 5' and 3' ends and cDNA was amplified by PCR using the same primers and a proof reading polymerase. Emulsion PCR (emPCR) set up, breaking, enrichment and pico-titer plate (PTP) loading steps were performed according to Roche protocols (Margulies *et al*. 2005; Emrich *et al*. 2007). Each of the two sequencing runs employed half of a PTP and was sequenced twice using Roche 454 GS FLX chemistry (Roche) according to the manufacturer's protocol.

Sequence cleaning and assembly

All sequence reads taken directly from the 454 GS-FLX sequencer were run through the sff file program (Roche) to remove sequencing adapters A and B, poor sequence data and barcodes. Contigs and singletons were renamed in a format 'A (M, O, T)_000001' where prefix 'A' was used for all assembled contigs derived from M, O, T cDNA libraries, with M (Muscle), O (Ovary), and T (Testis) standing for an individual library and assembly, and 000001 standing for the first arbitrary contig assignment number. In the case of singletons, the same prefix codes (A, M, O, T) for cDNA library origin(s) were added in front of each

read name (e.g. A_G1OH9PT01AF0I7). Sequences containing homopolymers of a single nucleotide comprising >60% of the read and that were >100 nucleotides in length were discarded. Trimmed sequences were assembled *de novo* using the default parameters of Newbler 2.5.3 (Roche). Each dataset of mRNA sequences from muscle, ovary and testis tissue was considered separately as being representative of the transcriptome of that tissue type at the time of sampling. On the assumption that some transcripts would be replicated across tissue-type datasets, these were merged in the combined dataset. After initial quality filtering, AGRF provided assembled contig and singleton datasets for analysis.

Annotation of mRNAs

BLASTx searches (Altschul *et al*. 1997) of the GenBank non-redundant (nr) database hosted by the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/) were performed on all contigs and singletons to identify putative mRNA functions (E-value threshold $<1e^{-5}$) as well as new ESTs. Numbers of ESTs that were either unique or shared among the libraries were visualized using a 3-way Venn diagram constructed using Venny (Oliveros 2007). Total EST numbers in the Venn diagram quadrants excluding abundant ESTs for ribosomal proteins counted redundant ESTs only once. The Blast2GO software suite (Gonesa *et al*. 2005; Götz *et al*. 2008) was used to predict functions of individual ESTs, assign Gene Ontology terms (The Gene Ontology Consortium 2000 & 2008), and to predict metabolic pathways using Kyoto Encyclopaedia of Genes and Genome (KEGG) (Kanehisa *et al*. 2006; Kanehisa *et al*. 2008). To identify protein domains, all translated sequences were interrogated against the InterPro databases (http://www.ebi.ac.uk/Tools/pfa/iprscan/) using the InterProScan tool (Hunter *et al*. 2009). The numbers of contigs annotated with each GO term for each library were quantified using WEGO (Ye *et al*. 2006).

Identification of EST-SSR motifs and EST-SNPs

All EST sequences were searched for SSR motifs using the QDD program (Meglécz *et al*. 2010). Default settings were employed to detect perfect di-, tri-, tetra-, penta-, and hexa-nucleotide motifs (including compound motifs). To be assigned, dinucleotide SSRs required a minimum of 6 repeats, and all other SSR types a minimum of 5 repeats. The maximum

interruption between 2 neighbouring SSRs to consider it being a compound SSR was set at 100 nucleotides. Perl script modules linked to the primer modelling software Primer3 (Rozen & Skaletsky 2000) were used to design PCR primers flanking for each unique SSR region identified.

Multiple nucleotide sequence alignments of contigs identified among the EST libraries derived from individual *M. rosenbergii* with divergent growth phenotypes were undertaken to identify putative SNPs. Alignments employed methods developed previously for plants and other species of agricultural importance (Bekal *et al*. 2008; Gorbach *et al*. 2010; Parchman *et al*. 2010) and included assessments of raw data alignments used in the initial assembly of contigs. Since no reference sequences were available, SNPs were identified as superimposed nucleotide peaks where 2 or more reads contained polymorphisms at the variant allele. SNPs were identified using default parameters in gsMapper (Roche) to align contigs from the individual and merged tissue type and prawn phenotype datasets and SNPs were predicted with high confidence when (i) the difference existed in at least three non-duplicated reads, (ii) the difference occurred in both the forward and reverse sequence reads unless present in at least seven same direction reads with quality scores over 20 (or 30 if the difference involves a 5-mer or more) and (iii) the difference comprised a single-base overcall or undercall forming a consensus differing from the each contig reference. Indels were segregated into simple types containing an insertion or deletion of at least one nucleotide compared with the reference sequence or complex types also containing nucleotides substitutions.

For the merged EST dataset, loose or strict criteria to maximize the discovery of rare alleles or to minimize the possibility of false-positive identifications were not considered (Vera *et al*. 2008; Gorbach *et al*. 2010). In addition, only an overall transition vs transversion (Ts/Tv) ratio was calculated across the dataset.

Data deposition

All *M. rosenbergii* EST sequences obtained were submitted to NCBI Sequence Read Archive under Accession no. SRP007672.

RESULTS AND DISCUSSION

Roche 454 GS-FLX sequencing and contig assembly

cDNA prepared to mRNA purified from muscle, ovary and testis tissues from *M. rosenbergii* were sequenced using the 454 GS-FLX platform. Sequences that passed basic quality standards were clustered and assembled *de novo*. In 454 sequencing run #1, a total of 121,214 EST sequences (total = 36.45 Mb) were assembled from mRNA isolated from either muscle or ovary tissue sampled from 6 adult females and preserved in ethanol prior to analysis. Average EST length was 295 nucleotides (nt). Assembly of high quality ESTs generated 1983 contigs averaging 673 nt in length. Due to technical issues with the first 454 GS-FLX run, the expected amount of data (200 Mb) was not retrieved. Therefore a second 454 sequencing run was conducted to increase genomic data, including the addition of testis-derived RNA. In 454 sequencing run #2, a total of 666,517 EST sequences were assembled from mRNA isolated from muscle and ovary from 9 adult females and 3 adult male testis tissues and preserved in RNAlater solution (Ambion) prior to analysis. Eyestalk-derived RNA was also extracted, but ultimately excluded from sequencing run #2 as quality control indicators suggested it contained PCR and proteinase inhibitors leading to failure of cDNA fragmentation, as detected in the bioanalyzer traces (samples were not fragmented). For the remaining three tissue types, the average EST length was 311 nt in 454 sequencing run #2. After removing adaptor sequences, the combined run #1 and #2 dataset contained 244.37 Mb of sequence comprising 787,731 reads averaging 310 nt in length, and the average coverage depth was 29.85 sequences per nucleotide position (Table 1). This average EST read length is longer and the sequencing coverage depth is substantially higher than has been reported in similar 454 sequencing analyses in non-model species including Glanville fritillary (197 nt at 2.3 x coverage; Vera *et al*. 2008), flooded or rose gum (245 nt; Novaes *et al*. 2008) or shore pine (306 nt at 3.6 x coverage; Parchman *et al*. 2010). As shown in Figure 1, assembly of high quality *M. rosenbergii* EST sequences generated 8,411 contigs varying in length from 40 nt to 7,531 nt (average 845 nt; total 212,142,540 nt), with 5,724 (68%) being >500 nt in length. The long individual read lengths combined with the 29.85-fold average coverage contributed to this high proportion of long contig sequences. Singletons ranged from 50 nt to 773 nt in length (average 279 nt, total 32,228,442 nt) (Figure 1). To our knowledge, this is the first comprehensive study of the transcriptome of *M. rosenbergii*.

Comparative analyses of ESTs

From BLASTx searches of *M. rosenbergii* EST coding sequences, 3,757 of the 8,411 (46%) contigs and 21,965 of the 115,123 (19%) singletons possessed significant similarity (E value $<1e^{-5}$) with proteins in the GenBank non-redundant (nr) database (Table S1). As might be expected, coding sequences in the majority of contigs (80%) and singletons (84%) matched well to crustacean and other arthropod proteins (Figure 2) which are in agreement with previous prawn studies (Leekitcharoenphon *et al*. 2010; Leu *et al*. 2011). After redundant and ribosomal protein sequences were excluded, 2,448 contig and 10,627 singleton sequences were identified as putative genes based on BLASTx matches.

Species most represented in the BLASTx searches included some penaeid shrimps, crabs and freshwater and marine crayfish species including giant tiger shrimp (*Penaeus monodon*), green mud crab (*Scylla paramamosain*), fleshy shrimp (*Fenneropenaeus chinensis*), Kuruma shrimp (*Marsupenaeus japonicas*), white leg shrimp (*Litopenaeus vannamei*), red swamp crayfish (*Procambarus clarkia*), and American lobster (*Homarus americanus*). Similarities in EST coding sequences are indicative of close evolutionary relationship of *M. rosenbergii* with other crustaceans. Only a few contig (1.8%) or singleton (3.9%) coding sequences matched protein sequences reported for *M. rosenbergii*, and again this was expected due to the limited number of *M. rosenbergii* EST (2365) and protein sequences (373) currently available in the NCBI databases. The *M. rosenbergii* EST sequences generated here thus will vastly expand the number of genes identified in this species.

More putative gene ESTs were detected in mRNA isolated from ovary tissue than from muscle or testis tissue (Figure 3). Only around 4% of the 3,757 contigs and 14% of the 21,965 singletons significantly matched either predicted or hypothetical genes (E value $<1e^{-5}$) due to the limited genomic information available for prawn species in the public database (Table S1). A significant number of *M. rosenbergii* ESTs did not possess coding sequences matching any sequences in the GenBank nr database which is not surprising for prawn EST studies (Leekitcharoenphon *et al*. 2010; Leu *et al*. 2011). Whilst most of these likely represent ESTs spanning only untranslated mRNA regions, chimeric EST sequences derived from assembly errors or ESTs containing non-conserved protein regions, as reported in other

transcriptome analyses (Wang *et al*. 2004; Liang *et al*. 2008; Mittapalli *et al*. 2010), it is also possible that some may constitute novel genes unique to this species.

Amongst ESTs derived from muscle tissue, coding sequences with homology to arginine kinase, ATP synthase, eukaryotic translation initiation factor, myosin heavy and light chain, sarcoplasmic calcium-binding protein, tropomyosin, and troponin were most abundant. Amongst ESTs derived from ovary tissue, coding sequences with homology to aldehyde dehydrogenase, ATP binding, cd63 antigen, cell division cycle, Chk1 checkpoint-like protein, e3 ubiquitin, eukaryotic translation initiation factor, ovary development-related protein, serine threonine-protein kinase, transmembrane protein, and WD repeat-containing protein were most abundant. Amongst ESTs derived from testis tissue, coding sequences with homology to eukaryotic translation initiation factor, kazal-type proteinase inhibitor, male reproductive-related protein, serine proteinase inhibitor, and viral A-type inclusion protein were most abundant. ESTs detected commonly across the 3 tissues included actins, elongation factors, eukaryotic translation initiation factor, heat shock protein, NADH dehydrogenase, reverse transcriptase, RNA-binding protein, senescence-associated protein, tubulin, ubiquitin and zinc finger protein (Figure 3, Table S1). Although this work was mainly focused on finding putative genes related with muscle development and growth, several putative functional transcripts identified here will lay the foundation for future studies aimed at investigating the role of sex determination, reproduction-related and xenobiotic genes which have been studied successfully in other species (Vera *et al*. 2008; Hale *et al*. 2009; Hale *et al*. 2010). These findings could be the best source for deciphering the putative function of novel genes in each tissue but further studies need to be conducted to understand the molecular functions of specific reported genes.

Gene Ontology assignments

Gene Ontology (GO) terms could be assigned to 8411 *M. rosenbergii* contigs based on BLAST matches to proteins with known functions (Figure 4, Table S2). EST coding sequences were assigned to cellular components (4,550 sequences, Figure 4A), molecular function (6,055 sequences, Figure 4B) and biological processes (8,806 sequences, Figure 4C). Amongst ESTs assigned molecular functions, many were assigned binding (45.9%) or catalytic functions (32.3%), predominantly actin and zinc ion proteins (Table S2). Recent studies of crustaceans have highlighted the importance of actin in constructing muscle tissues

and that it shows variable expression in different muscle types (Hooper & Thuma 2005; Zhu *et al*. 2005; Hooper *et al*. 2008; Kim *et al*. 2009). The cellular component assignments showed many EST coding sequences to likely possess cell (22.8%) and cell part (22.5%) functions, whilst those assigned biological functions were mostly predicted to be involved in cellular (17.6%) or metabolic processes (16.5%) including proteolysis, carbohydrate metabolism or oxidation-reductive functions. Analyses of the transcriptomes of other crustaceans have identified ESTs possessing similar arrays of potential metabolic functions (Tassanakajon *et al*. 2006; Leelatanawit *et al*. 2009; Leekitcharoenphon *et al*. 2010; Leu *et al*. 2011).

KEGG analysis

Many of the coding sequences present in the *M. rosenbergii* EST contig dataset were identified to occur in KEGG pathways; metabolic pathways (*n* = 320), biosynthesis of secondary metabolites (*n* = 135), oxidative phosphorylation (*n* = 66), biosynthesis of phenylpropanoids (*n* = 59), and biosynthesis of alkaloids derived from histidine and purine (*n* = 51) (Table S3). Metabolic pathways, implicated in the kinetic impairment of muscle glutamine homeostasis in adult and old glucocorticoid-treated rats (Minet-Quinard *et al*. 2004), showed the highest number of transcripts here. A skeletal muscle structure in rat intrauterine growth restriction indicated that changes in metabolic pathways were involved in obesity (Huber *et al*. 2009). A total of 66 transcripts were involved in oxidative phosphorylation. The integrity of the inner membrane and the associated complexes is essential to oxidative phosphorylation to generate ATP to supply readily-available free energy for the body (Lesser 2006). However, malfunction of oxidative phosphorylation could accentuate ATP depletion with the basic energy conservation system due to anoxic conditions in the tissues which could lead to metabolic failure (Gnaiger *et al*. 2000).

Interestingly, we recovered a high number of transcripts that were mapped to the phenylpropanoids biosynthesis pathway (59). Phenylpropanoids not only play an important role in contributing to all aspects of plant responses towards biotic and abiotic stimuli (Vogt 2010) but also have a potential dietary importance from plant derived compounds (Ferrer *et al*. 2008). A total of 51 transcripts also were predicted to the alkaloid biosynthesis pathway from histidine and purine in the *M. rosenbergii* EST contig dataset. Alkaloids, regarded as basic plant derived metabolites, are important components of plant defence, growth and

development systems (Hagel *et al*. 2008; Ziegler & Facchini 2008). In a study of sponges and ascidians, an abundance of alkaloids was reported that displayed biological activities such as metabolites (Kashman *et al*. 2010). Considering the omnivorous dietary habit of *M. rosenbergii*, finding these pathways was not surprising. Although not all of the major genes reported in putative KEGG pathways were found in the current study, this information provides insight into the specific responses and functions involved in molecular processes in *M. rosenbergii* metabolism and muscle contraction against biotic and abiotic stimuli.

Protein domains

InterProScan searches identified 19,036 protein domains among the 8,411 *M. rosenbergii* contigs (Table S4). Consistent with similar analyses in insects and other crustaceans (Leekitcharoenphon *et al*. 2010; Leu *et al*. 2011: Bai *et al*. 2011a), domains that dominated occur in RNA-binding proteins, protein kinases and transcription factors (zinc finger domains) (Table 2) that are essential for cellular processing functions including signal transduction and transcription regulation, regulation of RNA stability and translation control (RNA recognition motifs), innate immunity, cell division, proliferation, apoptosis and cell differentiation (McNeil *et al*. 2001; Sutherland *et al*. 2005).

The most common DNA-binding motifs present in eukaryotic and prokaryotic transcription factors (Bouhouche *et al*. 2002) were prevalent in the *M. rosenbergii* sequences, with 179 C2H2-type and 102 C2H2-like zinc finger (Znf) domains identified. Transcription factors usually contain several Znf domains capable of making multiple contacts with DNA (Wolfe *et al*. 2000), and can also bind to RNA and protein targets (Brayer & Segal 2008). A total of 112 nucleotide-binding α-β plait domains found in RNA-binding domains from various ribonucleoproteins or in viral DNA-binding domains (Bochkarev *et al*. 1995; Kielkopf *et al*. 2004) were predicted to exist among the *M. rosenbergii* EST coding sequences. In addition, 108 Armadillo-type fold and 84 Armadillo-like helical domains which form structural domains consisting of a multi-helical fold comprised of 2 curved layers of α-helices (Kraemer *et al*. 1999), were predicted.

Among *M. rosenbergii* EST coding sequences, 104 domains containing WD40/YVTN repeat-like sequences, 90 domains containing WD40-repeat sequences and 88 domains containing WD40 repeat-like sequences were predicted. These domains are involved in a variety of functions ranging from signal transduction and transcription regulation to cell cycle

control and apoptosis (Smith 1999; Li & Roberts 2001). A total of 86 Ran GTPase families which are involved in regulating GTP hydrolases (Bourne *et al*. 1990), contain GTP-binding domains (Bourne *et al*. 1991) and regulate receptor-mediated transport between the nucleus and the cytoplasm (Scheffzek *et al*. 1995; Rush *et al*. 1996) were also predicted, as were 84 immunoglobulin (Ig)-like fold domains. Ig-like fold domains are involved in a variety of functions including cell-cell recognition, cell-surface receptors, muscle structure and the immune system (Teichmann & Chothia 2000), and are often involved with protein-protein interactions mediated by their β-sheets as in other Ig-like domains (Teichmann & Chothia 2000; Potapov *et al*. 2004). Other domains identified abundantly included Serpin (*ser*ine *p*roteinase *in*hibitor) domains (n = 79) and NAD(P)-binding domains (n = 72). Interestingly, few PAZ (n = 3) or PIWI (n = 8) domains believed to be important components of the dsRNA-induced silencing complex were identified. The relative absence of ESTs with such domains is perplexing based on the detection of genes encoding Dicer and Argonaut type proteins in penaeid shrimp (Dechklar *et al*. 2008; Su *et al*. 2008; Wu *et al*. 2010) and the clear demonstration of effective RNAi-mediated knockdown of gene expression in shrimp (Dechklar *et al*. 2008). Similar transcriptome analyses of other tissues including haemocytes from the lymphoid organs for example that are primary mediators of pathogen defence responses (Leekitcharoenphon *et al*. 2010; Soonthornchai *et al*. 2010; Leu *et al*. 2011) might be useful for indentifying if expression of ESTs encoding putative RNAi-related domains are more cell specific than domains required broadly for cell functioning. Although an original aim of this study was to identify candidate genes, gene families or gene domains potentially involved with growth phenotypes and/or other production traits important for *M. rosenbergii* aquaculture, none were differentiated from cell function or pathogen defence type activities. The identification of such ESTs has been confounded in most studies of shrimp to date focussing on the identification and characterisation of pathogen defence-related genes (Leekitcharoenphon *et al*. 2010; Soonthornchai *et al*. 2010; Leu *et al*. 2011). Thus genes mediating growth performance and potentially of value in selective breeding programs await discovery.

Putative genes affecting muscle development and/or function

The *M. rosenbergii* EST sequence database was mined for coding sequences with domains involved potentially with muscle development and function (Table 3). Despite recent

advances in sequencing technologies, few genes with such functions have been characterised from any crustaceans, and only 2365 ESTs assigned to *M. rosenbergii* and 5536 ESTs assigned to *Macrobrachium* were available in NCBI databases before this study. However, the 123,534 ESTs from the *M. rosenbergii* individuals selected from high and low growth performance cohorts should contain genes potentially expressed differentially and with functional characteristics suggestive of roles in muscle mass accumulation and other growth-related functions.

In the current study, both actin and myosin proteins including tropomyosin and troponin showed a high number of transcripts. It has been reported that actins are expressed in abundance as they are critical to formation of muscle filaments (Kabsch & Vandekerckhove 1992; Dominguez & Holmes 2011). Different actin isoforms have been identified in various crustaceans (Zhu *et al*. 2005), and are likely to be involved in playing important roles in cytoskeletal structure, cell division and mobility, and muscle contraction (Zhu *et al*. 2005; Hooper *et al*. 2008; Kim *et al*. 2009). The large super-family of myosin proteins interact with actin filaments by hydrolysing adenosine triphosphate that combine to form thick muscle filaments (Hayashida *et al*. 1991). Myosin heavy chain (MHC) isoforms differ in their shortening velocity compared with other isoforms due to the enhanced ability of the myosin head to hydrolyse ATP (Schiaffino & Reggiani 1996). Multiple MHC isoforms are expressed ubiquitously in all eukaryotic cells and they are the most abundant contractile protein present in skeletal muscle (DeNardi *et al*. 1993; Jung *et al*. 1998). If growth rates of *M. rosenbergii* are dictated primarily by the efficiency at which feed is converted into muscle mass, it is likely that myosin gene expression levels could provide a good molecular marker of individual growth potential, as found in the Atlantic pink shrimp *Farfantepenaeus paulensis* (Kamimura *et al*. 2008). In studies of other crustaceans, high expression levels of genes encoding fast and slow myosin isoforms have been found to be accompanied by elevated expression of other genes encoding for example, actin, myofibrillar protein, tropomyosin, troponin I, and troponin T (Medler *et al*. 2004; Medler *et al*. 2005; Abdel Rahman *et al*. 2010). According to Perry *et al*. (2009), differences in expression levels of myofibrillar protein isoforms correlate well with individual body size in crabs, with changes in expression spanning several orders of magnitude occurring at different life stages. Tropomyosins comprise a family of closely related proteins present both in muscle and non-muscle cells (MacLeod 1987). In striated muscle, tropomyosin mediates interactions between the troponin complex and actin to mediate muscle contraction (Wolska & Wieczorek 2003).

A high number of actin and myosin protein transcripts observed here may regulate muscle development and function in *M. rosenbergii*. However, further studies are needed to confirm these observations.

High occurrence of calponin and transgelin was also observed in the transcriptome of *M. rosenbergii*. Calponin is a smooth muscle-specific protein capable of binding actin, tropomyosin and calmodulin and is also involved in mediating muscle contraction (Strasser *et al*. 1993) as its interaction with actin inhibits actomyosin Mg-ATPase activity. In previous studies of invertebrates and vertebrates, caldesmon and calponin were shown to interact with actin, tropomyosin, and Ca^{2+}-calmodulin (Meyer-Rochow & Royuela 2002; Hooper & Thuma 2005; Hooper *et al*. 2008). In addition, transgelin is a calponin which is expressed exclusively in smooth muscle-containing tissues in adult animals and is one of the earliest markers of differentiated smooth muscle cells (Prinjha *et al*. 1994; Solway *et al*. 1995).

The current study reports a number of putative genes, transcription factors, and early regulators that are potentially involved in muscle development and function in *M. rosenbergii*. Further studies need to be performed, however, to learn the molecular functions of these reported genes which were observed to be expressed more abundantly in adult female and male prawns compared with earlier developmental stages or slow growth performance individuals.

Genes of interest related to growth

The transcriptome of *M. rosenbergii* was examined primarily to identify genes associated functionally with individual growth. For this reason, an EST dataset was compiled from tissues of individuals from high and low growth performance families (Table 3). Amongst these, a putative cyclophilin was identified. Although cyclophilins possess diverse functions and have been linked to innate immunity (Belfiore *et al*. 2004; Towers 2007) and testicular development (Vogt 2010), expression levels of cyclophilin-like proteins have also been found to be highly correlated with body-weight in the shrimp *P. monodon* (Tangprasittipap *et al*. 2010).

Intracellular fatty acid-binding proteins (FABPs), identified in the current transcriptomic study, are members of a lipid-binding protein super-family that occur in both invertebrates and vertebrates, and together with acyl-CoA-binding protein (ACBP) are

involved in lipid metabolism (Zimmerman & Veerkamp 2002). Few FABPs have been identified in invertebrates (Zimmerman & Veerkamp 2002; Esteves & Ehrlich 2006), and their physiological roles remain largely unknown. However, in the locust *Schistocerca gregaria*, FABP expression has been reported to be strictly adult specific and is controlled by fatty acids in adult muscle (Haunerland *et al*. 1993). Locust flight muscle employs fatty acids exclusively as the energy source for sustained flight and it is likely that FABP is involved in intracellular fatty acid transport (Van der Horst 1990).

In the current study, we found high occurrence of LIM domain proteins, which play important biological roles in cytoskeleton organisation, cell fate determination and organ development (Bach 2002). Previously, one LIM domain gene (ISL1) has been identified as a positional candidate for obesity and for controlling leptin levels, and is suggested to be involved in body weight regulation and glucose homeostasis (Barat-Houari *et al*. 2002). In a study of the red crab *Gecarcoidea natalis*, two genes encoding LIM proteins, a paxillin-like transcript (pax) and a muscle LIM protein (mlp), were up-regulated in muscle of crabs in the wet season (Postel *et al*. 2010). These proteins could play a fundamental role in muscle development and reconstruction, and their comparative up-regulation is consistent with a remodelling of leg muscle needed for migration during the wet season (Postel *et al*. 2010).

Physiologically, O-methyltransferase (OMT) plays an important regulatory role in plant and animal growth, development, reproduction and immune response (Ibrahim *et al*. 1998; Kuballa *et al*. 2007). OMT transcripts observed in the current study could represent a potential candidate gene for developing novel traits in prawns. Methyl farnesoate (MF), the sesquiterpenoid precursor of insect juvenile hormone III (JH III), is produced and released by mandibular organs in decapod crustaceans (Borst *et al*. 1987; Laufer *et al*. 1987a; Tamone *et al*. 1997). The physiological function of MF, however, is not well understood in crustaceans, but by analogy with established functions of JH III in insects, MF has been suggested to play an important role in regulation of growth and reproduction in crustaceans (Laufer *et al*. 1987a, b). In some crustaceans, circulating titer and biosynthesis of MF appear to be correlated positively with maturation of the ovary (Laufer *et al*. 1987b; Sagi *et al*. 1991). MF has also been suggested to play a role in delaying onset of molting in larval crustaceans (Borst *et al*. 1987; Sagi *et al*. 1991). This evidence implicates MF in both crustacean growth and reproduction. Farnesoic acid O-methyltransferase (FAMeT; also known as S-adenosyl-methionine: farnesoic acid O-methyltransferase) is the enzyme that catalyses the final step in the MF biosynthetic pathway in crustaceans (Feyereisen *et al*. 1981; Wang *et al*. 1994).

Studies of crustacean FAMeT indicate that it may directly or indirectly (through MF) modulate reproduction and growth in crustaceans (Gunawardene *et al*. 2003; Ruddell *et al*. 2003; Holford *et al*. 2004; Hui *et al*. 2008) by interacting with eyestalk neuropeptides as a consequence of its presence in neurosecretory cells in the X-organ-sinus gland. It is also believed that MF is the crustacean homolog for insect juvenile hormone, a molecule that may also regulate growth and reproduction in crustaceans (Hui *et al*. 2008). If growth rates of *M. rosenbergii* are dictated primarily by the efficiency at which feed is converted into muscle mass, it is likely that FABP, LIM domain and FAMeT gene expression levels could provide candidate molecular markers of individual growth potential.

Another interesting finding in the current study is the expression of profilin, a small actin-binding protein found in eukaryotic cells that are critical for cytoskeletal dynamics (Critchley *et al*. 1999; Pollard 2008). Profilins are potent regulators of actin filament dynamics and promote exchange of ADP to ATP on actin and by affinity to profilin–actin complexes for actin filament ends (Buss *et al*. 1992). Profilins have diverse roles in cellular processes, including membrane trafficking, small-GTPase signalling and nuclear activities, neurological diseases, and tumor formation (Witke *et al*. 1998; Rawe *et al*. 2006; Birbach 2008). Genetic studies have shown the importance of profilins for cell proliferation and differentiation. Profilin gene disruption leads to grossly impaired growth, motility and cytokinesis, and embryonic lethality in multicellular organisms, for example in insects and mice (Magdolen *et al*. 1988; Haugwitz *et al*. 1994; Verheyen & Cooley 1994).

The current study identified a number of putative genes that are potentially involved with growth in *M. rosenbergii*. However, further studies are needed to understand the molecular functions of these putative genes with growth performance and development in *M. rosenbergii*.

Putative Molecular Markers

SNPs in *M. rosenbergii* EST contigs were identified from alignments of multiple sequences used for contig assembly. Of the 834 SNPs detected, 555 were putative transitions (Ts) and 279 were putative transversions (Tv), giving a mean Ts : Tv ratio of 1.99 : 1.00 across the transcriptome (Figure 5, Table S5). The SNP types A↔ G and C ↔ T were most common and SNP densities varied among genes, possibly due in part, to the effects of strong historical

selection and the relative functional importance of individual genes. The Ts : Tv ratio can help identify genes affected by selection (Morton *et al*. 2006). Although alignments also identified a total of 1198 indels across the transcriptome (Figure 5, Table S5), this must be treated with caution because of technical problems associated with 454 pyrosequencing (Margulies *et al*. 2005; Hale *et al*. 2009).

Moreover, a total 658 simple sequence repeats (SSRs) or microsatellites comprising 61.85% dinucleotide repeats, 35.87% trinucleotide repeats and 2.28% tetra/penta/hexa-nucleotide repeats were detected (Figure 6, Table S6) in the contigs as well as singletons. PCR primers could be designed for almost all predicted polymorphic SSRs (Table S6) but these have yet to be validated as markers useful for examining *M. rosenbergii* adaptation and ecology as has been done with other non-model species (Morin *et al*. 2004; Barbzuk *et al*. 2007; Kim *et al*. 2011). In addition, SNPs and SSRs detected here are likely to be highly transferable to other closely related species as has been found for other crustacean species (Ellis & Burke 2007; Gorbach *et al*. 2010; Kim *et al*. 2011). It is envisaged that the potential markers identified here within the ESTs will provide an invaluable resource for studying the evolution and molecular ecology of *Macrobrachium* species and for genome mapping and quantitative trait loci (QTL) analysis. However, many of the putative *M. rosenbergii* SNPs identified could simply represent allelic variants and future studies are planned to validate which are real. As ESTs were generated from 3 different tissue types, differential expression of different tissue-specific alleles is possible. However, this is rare as it requires somatic mutation or chimerisms between tissues.

CONCLUSION

Here we report the first comprehensive EST dataset covering the transcriptome of the giant freshwater prawn *M. rosenbergii*, a non-model prawn species for which little molecular knowledge currently exists. The 123,534 putative ESTs (115,123 singletons and 8,411 contigs) identified and assembled will enable gene discovery in *M. rosenbergii*, assist in evolutionary studies and with the significant number of putative growth-related genes identified should facilitate genomics approaches to improving the growth performance of domesticated GFP stocks used for aquaculture. In addition, the large number of SNPs and SSRs detected provide targets for identifying polymorphisms across *M. rosenbergii* populations useful for parentage assignment and for managing inbreeding in cultured

populations. Moreover, the EST sequences reported should prove invaluable for gene mining and annotation and phylogenetic analyses as well as provide a resource that can be exploited as molecular markers and in gene expression studies in this commercially important aquaculture species.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the help provided by the Australian Genome Research Facility (Matthew Tinning, Tim Bruxner, Rachel Kliese, and Adam Skarshewski) in regard to pyrosequencing and initial analysis of the raw sequence data. The authors thank Markus Dwyer for assisting with computer service. We also thank Leanne Dierens and Barney Hines for assisting with RNA isolations. We gratefully acknowledge Jeffrey Cowley, Nick Hudson, Woo-Jin Kim, Peter Prentis and two anonymous reviewers' contributions for providing constructive comments on the manuscript.

SUPPORTING INFORMATION

Table S1 Summary of BLASTx results for contigs and singletons of *M. rosenbergii*.

(XLSX)

Table S2 Gene Ontology of *M. rosenbergii* contig sequences.

(XLSX)

Table S3 KEGG summary of *M. rosenbergii* contig sequences.

(XLSX)

Table S4 InterProScan domain search of *M. rosenbergii* contig sequences.

(XLSX)

Table S5 Putative SNPs and Indels in *M. rosenbergii* contig sequences.

(XLSX)

Table S6 Putative microsatellite loci in *M. rosenbergii* contig and singleton sequences.

(XLSX)

REFERENCES

Abdel Rahman AM, Kamath S, Lopata AL, Helleur RJ (2010) Analysis of the allergenic proteins in black prawn (*Penaeus monodon*) and characterization of the major allergen tropomyopsin using mass spectrometry. *Rapid Communication in Mass Spectrometry* **24**: 2462–2470.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z *et al*. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**: 3389–3402.

Bach I (2002) The LIM domain: regulation by association. *Mechanisms of Development* **91**: 5–17.

Bai X, Mamidala, Rajarapu SP, Jones SC, Mittapalli O (2011a) Transcriptomics of the bed bug (*Climex lectularius*). *PLoS One* **6**: e16336.

Bai X, Rivera-Vega L, Mamidala P, Bonello P, Herms DA *et al*. (2011b) Transcriptomic signature of ash (*Fraxinus* spp.) phloem. *PLoS One* **6**: e16368.

Barat-Houari M, Clément K, Vatin V, Dina C, Bonhomme G *et al*. (2002) Positional candidate gene analysis of Lim domain homeobox gene (Isl-1) on chromosome 5q11–q13 in a French morbidly obese population suggests indication for association with type 2 diabetes. *Diabetes* **51**: 1640–1643.

Barbazuk WB, Emrich SJ, Chen HD, Li L, Schnable PS (2007) SNP discovery via 454 transcriptome sequencing. *Plant Journal* **51**: 910–918.

Baruah K, Cam DTV, Dierckens K, Wille M, Defoirdt T *et al*. (2009) In vivo effects single or combined N-acyl homoserine lactone quorum sensing signals on the performance of *Marcrobrachium rosenbergii* larvae. *Aquaculture* **288**: 233–238.

Belfiore M, Pugnale P, Saudan Z, Puoti A (2004) Roles of the *C. elegans* cyclophilin like protein MOG-6 in MEP-1 binding and germline fates. *Development* **131**: 2935–2945.

Bekal S, Craig JP, Hudson ME, Niblack TL, Domier LL *et al*. (2008) Genomic DNA sequence comparison between two inbred soybean cyst nematode biotypes facilitated by massively parallel 454 micro-bead sequencing. *Molecular Genetics and Genomics* **279**: 535–543.

Birbach A (2008) Profilin, a multi-modal regulator of neuronal plasticity. *BioEssays* **30**: 994–1002.

Bochkarev A, Barwell JA, Pfuetzner RA, Furey W Jr, Edwards AM *et al*. (1995) Crystal structure of the DNA-binding domain of the Epstein-Barr virus origin-binding protein EBNA 1. *Cell* **83**: 39–46.

Borst DW, Laufer M, Landau ES, Chang WA, Hertz FC *et al*. (1987) Methyl farnesoate (MF) and its role in crustacean reproduction and development. *Insect Biochemistry* **17**: 1123–1127.

Bouhouche N, Syvanen M, Kado CI (2002) The origin of prokaryotic C2H2 zinc finger regulators. *Trends in Microbiology* **8**: 77–81.

Bourne HR, Sanders DA, McCormick F (1990) The GTPase superfamily: a conserved switch for diverse cell functions. *Nature* **348**: 125–132.

Bourne HR, Sander DA, McCormick F (1991) The GTPase superfamily: conserved structure and molecular mechanism. *Nature* **349**: 117–127.

Brayer KJ, Segal DJ (2008) Keep your fingers off my DNA: protein-protein interactions mediated by C2H2 zinc finger domains. *Cell Biochemistry and Biophysics* **50**: 111–131.

Buss F, Temm-Grove C, Henning S, Jockusch BM (1992) Distribution of profiling in fibroblasts correlates with the presence of highly dynamic actin filaments. *Cell Motility and the Cytoskeleton* **22**: 51–61.

Cam DTV, Nhan DT, Ceuppens S, Hao NV, Dierckens K *et al*. (2009) Effect of N-acyl homoserine lactone-degrading enrichment cultures on *Marcrobrachium rosenbergii* larviculture. *Aquaculture* **294**: 5–13.

Chand V, de Bruyn M, Mather PM (2005) Microsatellite loci in the eastern form of the giant freshwater prawn (*Marcrobrachium rosenbergii*). *Molecular Ecology Notes* **5**: 308–310.

Chromczynski P, Mackey K (1995) Short technical report. Modification of TRIZOL reagent procedure for isolation of RNA from Polysaccaride-and proteoglycan-rich sources. *Biotechniques* **19**: 942–945.

Critchley DR, Hold MR, Barry ST, Priddle H, Hemmings L *et al*. (1999). Integrin-mediated cell adhesion: the cytoskeletal connection. *Biochemical Society Symposia* **65**: 79–99.

de Bruyn M, Wilson JC, Mather PB (2004) Huxley's line demarcates extensive genetic divergence between eastern and western forms of the giant freshwater prawn, *Marcrobrachium rosenbergii*. *Molecular Phylogenetics and Evolution* **30**: 251–257.

de Bruyn M, Mather PB (2007) Molecular signatures of Pleistocene sea-level changes that affected connectivity among freshwater shrimp in Indo-Australian waters. *Molecular Ecology* **16**: 4295–4307.

Dechklar M, Udomkit A, Panyim S (2008) Characterization of argonaute cDNA from *Penaeus monodon* and implication of its role in RNA interference. *Biochemical and Biophysical Research Communications* **367**: 768–774.

DeNardi C, Ausoni S, Moretti P, Gorza L, Velleca M *et al*. (1993) Type 2X-myopsin heavy chain is coded by a muscle fiber type-specific and developmentally regulated gene. *Journal of Cell Biology* **123**: 823–835.

Dinh H, Coman G, Hurwood D, Mather P (2012) Experimental assessment of the utility of VIE tags in a stock improvement program for giant freshwater prawn (*Macrobrachium rosenbergii*) in Vietnam. *Aquaculture Research* **43**: 1471–1479.

Divu D, Khushiramani R, Malathi S, Karunasagar I, Karunasagar I (2008) Isolation, characterization and evaluation of microsatellite DNA markers in the giant freshwater prawn, *Marcrobrachium rosenbergii*, from South India. *Aquaculture* **284**: 481–284.

Dominguez R, Holmes K (2011) Actin structure and function. *Annual Review of Biophysics* **40**: 169–186.

Du ZQ, Ciobanu DC, Onteru SK, Gorbach D, Mileham AJ *et al*. (2010) A gene-based SNP linkage map for pacific white shrimp, *Litopenaeus vannamei*. *Animal Genetics* **41**: 286–294.

Ellis JR, Burke JM (2007) EST-SSRs as a resource for population genetic analyses. *Heredity* **99**: 125–132.

Emrich SJ, Barbazuk WB, Li L, Schnable PS (2007) Gene discovery and annotation using LCM-454 transcriptome sequencing. *Genome Research* **17**: 69–73.

Esteves A, Ehrlich R (2006) Invertebrate intracellular fatty acid binding proteins. *Comparative Biochemistry and Physiology Part C* **2**: 262–274.

FAO (2009) Fisheries Statistical Database, Global Aquaculture Production (Fisheries Global Information System, online query).

Feyereisen R, Friedel T, Tobe SS (1981) Farnesoic acid stimulation of C16 juvenile hormone biosynthesis by corpora allata of adult female *Diploptera punctata*. *Insect Biochemistry* **11**: 401–409.

Ferrer JL, Austin MB, Stewart Jr C, Noel JP (2008) Structure and function of enzyme involved in the biosynthesis of phenylpropanoids. *Plant Physiology and Biochemistry* **46**: 356–370.

Gnaiger E, Méndez G, Hand SC (2000) High phosphorylation efficiency and depression of uncoupled respiration in mitochondria under hypoxia. *Proceedings of the National Academy of Sciences of the USA* **97**: 11080–11085.

Gonesa A, Götz S, Garcia-Gomez JM, Terol J, Talon M *et al*. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**: 3674–3676.

Gorbach DM, Hu ZL, Du ZQ, Rothschild MF (2010) Mining ESTs to determine the usefulness of SNPs across shrimp species. *Animal Biotechnology* **21**: 100–103.

Götz S, Garcia-Gomez JM, Terol J, William TD, Gagaraj SH *et al*. (2008) High-throughput functional annotation and data mining with Blast2GO suite. *Nucleic Acids Research* **36**: 3420–3435.

Gunawardene YI, Bendena WG, Tobe SS, Chana SM (2003) Comparative immunohistochemistry and cellular distribution of farnesoic acid O-methyltransferase in the shrimp and the crayfish. *Peptides* **24**: 1591–1597.

Hagel JM, Weljie AM, Vogel HJ, Facchini PJ (2008) Quantitative ^{1}H nuclear magnetic resonance metabolite profiling as a functional genomics platform to investigate alkaloid biosynthesis in Opium Poppy. *Plant Physiology* **147**: 1805–1821.

Hale MC, McCormick CR, Jackson JR, DeWoody JA (2009) Next-generation pyrosequencing of gonad transcriptomes in the polyploidy lake sturgeon (*Acipenser fulvescens*): the relative merits of normalization and rarefaction in gene discovery. *BMC Genomics* **10**: 203.

Hale MC, Jackson JR, DeWoody JA (2010) Discovery and evaluation of candidate sex-determining genes and xenobiotics in the gonads of lake sturgeon (*Acipenser fulvescens*). *Genetica* **138**: 745–456.

Hamasaki K, Kitada S (2008) Potential of stock enhancement for decapods crustaceans. *Reviews in Fisheries Science* **16**: 164–174.

Haugwitz M, Noegel AA, Karakesisoglou J, Schleicher M (1994) *Dictyostelium amoebae* that lack G-actin-sequestering profilins show defects in F-actin content, cytokinesis, and development. *Cell* **79**: 303–314.

Haunerland NH, Xinmei Ch, Andolfatto P, Chisholm JM,Wang Z (1993) Developmental changes of FABP concentration, expression and intracellular distribution in locust flight muscle. *Molecular and Cellular Biochemistry* **123**: 153–158.

Hayashida M, Maita T, Matsuda G (1991) The primary structure of skeletal muscle myosin heavy chain: I. Sequences of amino-terminal 23 kDa fragment. *Journal of Biochemistry* **110**: 54–59.

Holford KC, Edwards KA, Bendena WG, Tobe SS, Wang Z *et al*. (2004) Purification and characterization of a mandibular organ protein from the American lobster, *Homarus americanus*: a putative farnesoic acid O-methyltransferase. *Insect Biochemistry and Molecular Biology* **34**: 785–798.

Hooper SL, Thuma JB (2005) Invertebrate muscles: muscle specific genes and proteins. *Physiological Reviews* **85**: 1001–1060.

Hooper SL, Hobbs KH, Thuma JB (2008) Invertebrate muscles: thin and thick filament structure; molecular basis of contraction and its regulation, catch and asynchronous muscle. *Progress in Neurobiology* **86**: 72–127.

Huber K, Miles JL, Norman AM, Thompson NM, Davison M *et al*. (2009) Prenatally induced changes in muscle structure and metabolic function facilitate exercise-induced obesity prevention. *Endorinology* **150**: 4135–4144.

Hui JH, Tobe SS, Chan SM (2008) Characterization of the putative farnesoic acid O-methyltransferase (LvFAMeT) cDNA from white shrimp, *Litopenaeus vannamei*: Evidence for its role in molting. *Peptides* **29**: 252–260.

Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, et al. (2009) InterPro: the integrative protein signature database. *Nucleic Acids Research* **37**: D211–215.

Ibrahim RK, Bruneau A, Bantignies B (1998) Plant o-methyltransferase: molecular analysis, common signature and classification. *Plant Molecular Biology* **36**: 1–10.

Jung HH, Lieber RL, Ryan AF (1998) Quantification of myosin heavy chain mRNA in somatic and brachial arch muscles using competitive PCR. *American Journal of Physiology* **275**: 68–74.

Kabsch W, Vandekerckhove J (1992) Structure and function of actin. *Annual Review of Biophysics and Biomolecular Structure* **21**: 49–76.

Kamimura MT, Meier KM, Cavalli RO, Laurino J, Maggioni R *et al*. (2008) Characterization of growth-related genes in the south-western Atlantic pink shrimp *Farfantepenaeus paulensis* (Pérez-Farfante 1967) through a modified DDRT-PCR protocol. *Aquaculture Research* **39**: 200–204.

Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M *et al*. (2006) From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Research* **34**: D354–357.

Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M *et al*. (2008) KEGG for linking genomes to life and the environment. *Nucleic Acids Research* **35**: D480–484.

Kashman Y, Bishara A, Aknin M (2010) Recent *N*-atom containing compounds from Indo-Pacific invertebrates. *Marine Drugs* **8**: 2810–2836.

Kielkopf CL, Lucke S, Green MR (2004) U2AF homology motifs: protein recognition in the RRM world. *Genes and Devopment* **18**: 1513–1526.

Kim BK, Kim KS, Oh C, Mykles DL, Lee SG *et al*. (2009) Twelve actin-encoding cDNAs from the American lobster, *Homarus americanus*: Cloning and tissue expression of eight skeletal muscle, one heart, and three cytoplasmic isoforms. *Comparative Biochemistry and Physiology Part B* **153**: 178–184.

Kim WJ, Jung HT, Gaffney PM (2011) Development of Type I genetic markers from expressed sequence tags in highly polymorphic species. *Marine Biotechnology* **13**: 127–132.

Kraemer B, Crittenden S, Gallegos M, Moulder G, Barstead R *et al*. (1999) NANOS-3 and FBF proteins physically interact to control the sperm-oocyte switch in *Caenorhabditis elegans*. *Current Biology* **9**: 1009–1018.

Kuballa A, Guyatt K, Dixon B, Thaggard H, Ashton AR, et al. (2007) Isolation and expression analysis of multiple isoforms of putative farnesoic acid *O*-methyltransferase in several crustacean species. *General and Comparative Endocrinology* **150**: 48–58.

Laufer H, Borst DW, Baker FC, Carrasco C, Sinkus M *et al*. (1987a) Identification of a juvenile hormone-like compound in a crustacean. *Science* **235**: 202–205.

Laufer H, Landau M, Homola E, Borst W (1987b) Methyl farnesoate: its site of synthesis and regulation of secretion in a juvenile crustacean. *Insect Biochemistry* **17**: 1129–1131.

Leekitcharoenphon P, Taweemuang U, Palittapongarnpim P, Kotewong R, Supasiri T *et al*. (2010) Predicted sub-populations in a marine shrimp proteome as revealed by

combined EST and cDNA data from multiple *Penaeus* species. *BMC Research Notes* **3**: 295.

Leelatanawit R, Sittikankeaw K, Yocawibun P, Klinbunga S *et al*. (2009) Identification, characterization and expression of sex-related genes in testes of the giant tiger shrimp *Penaeus monodon*. *Comparative Biochemistry and Physiology Part A* **152**: 66–76.

Leu J, Chen S, Wang Y, Chen Y, Su S *et al*. (2011) A review of the major Penaeid shrimp EST studies and the construction of a shrimp transcriptome database based on the ESTs from four *Penaeid* shrimp. *Marine Biotechnology* **13**: 608–621.

Lesser MP (2006) Oxidative stress in marine environments: Biochemistry and physiological ecology. *Annual Reviews of Physiology* **68**: 253–278.

Li D, Roberts R (2001) WD-repeat proteins: structure characteristics, biological function, and their involvement in human diseases. *Cellular and Molecular Life Science* **58**: 2085–2097.

Liang H, Carlson JE, Leebens-Mack JH, Wall PK, Mueller LA *et al*. (2008) An EST database for *Liriodendron tulipifera* L. floral buds: the first EST resource for functional and comparative genomics in *Liriodendrom*. *Tree Genetics and Genomes* **4**: 419–433.

Lyons RE, Dierens LM, Tan SH, Preston NP, Li Y (2007) Characterization of AFLP markers associated with growth in the Kuruma prawn, *Marsupenaeus japonicus*, and identification of a candidate gene. *Marine Biotechnology* **9**: 712–721.

MacLeod AR (1987) Genetic origin of diversity of human cytoskeletal tropomyosins. *Bioessays* **6**: 208–212.

Magdolen V, Oechsner U, Muller G, Bandlow W (1988) The intron-containing gene for yeast profiling (PFY) encodes a vital function. *Molecular and Cellular Biology* **8**: 5108–5115.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS *et al*. (2005) Genome sequencing in microfabricated high-density picolitre reactions. *Nature* **437**: 376–380.

McNeil GP, Schroeder AJ, Roberts MA, Jackson FR (2001) Genetic analysis of functional domains within the *Drosophila* LARK RNA-binding protein. *Genetics* **159**: 229–240.

Medler S, Lilley T, Mykles DL (2004) Fiber polymorphism in skeletal muscles of the American lobster, *Homarus americanus*: continuum between slow-twitch (S1) and slow-tonic (S2) fibers. *Journal of Experimental Biology* **207**: 2755–2767.

Medler S, Brown KJ, Chang ES, Mykles DL (2005) Myosin heavy chain gene expression in adult lobster skeletal muscles. *Biological Bulletin* **208**: 127–137.

Meyer-Rochow VB, Royuela M (2002) Calponin, caldesmon, and chromatophores: The smooth muscle connection. *Microscopy Research and Technique* **58**: 504–513.

Miller AD, Murphy NP, Burridge CP, Austin CM (2005) Complete mitochondrial DNA sequences of the Decapod Crustaceans *Pseudocarcinus gigas* (Menippidae) and *Marcrobrachium rosenbergii* (Palaemonidae). *Marine Biotechnology* **7**: 339–349.

Minet-Quinard R, Moinard C, Villie F, Vasson MP, Cynober L (2004) Metabolic pathways implicated in the kinetic impairment of muscle glutamine homeostasis in adult and old glucocorticoid-treated rats. *American Journal of Physiology Endocrinology and Metabolism* **287**: E671–676.

Mittapalli O, Bai X, Mamidala P, Rajarapu SP, Bonello P, et al. (2010) Tissue-specific transcriptomics of the exotic invasive insect pest emerald ash borer. *PLoS One* **5**: e13708.

Meglécz E, Constedoat C, Dubut V, Gilles A, Malausa T *et al*. (2010) QDD: a user-friendly program to select microsatellite markers and design primers from large sequencing projects. *Bioinfomatics Applications Note* **26**: 403–404.

Morin PA, Luikart G, Wayne RK, the SNP workshop group (2004) SNPs in ecology, evolution and conservation. *Trends in Ecology and Evolution* **19**: 208–216.

Morton BR, Bi IV, McMullen MD, Gaut BS (2006) Variation in mutation dynamics across the maize genome as a function of regional and flanking base composition. *Genetics* **172**: 569–577.

Nhan DT, Wille M, Hung LT, Sorgeloos P (2009) Comparison of reproductive performance and offspring quality of giant freshwater prawn (*Macrobrachium rosenbergii*) broodstock from different regions. *Aquaculture* **298**: 36–42.

Ngernsoungnern P, Ngernsoungnern A, Weerachatyanukul W, Meeratana P, Hanna PJ et al. (2009) Abalone egg-laying hormone induces rapid ovarian maturation and early spawning of giant freshwater prawn, *Macrobrachium rosenbergii*. *Aquaculture* **296**: 143–149.

Novaes E, Drost DR, Farmerie WG, Pappas Jr GJ, Grattapaglia D et al. (2008) High-throughput gene and SNP discovery in *Eucalyptus grandis*, and uncharacterized genome. *BMC Genomics* **9**: 312.

Oliveros JC (2007) VENNY: an interactive tool for comparing list with Venn Diagram. VENNY Website. Available: http://bioinfogp.cnb.csic.es/tools/venny/index.html. Accessed 2011 Nov 01.

Parchman TL, Geist KS, Grahnen JA, Benkman CW, Buerkle CA (2010) Transcriptome sequencing in an ecologically important tree species: assembly, annotation, and marker discovery. *BMC Genomics* **11**: 180.

Perry MJ, Tait J, Hu J, White SC, Medler S (2009) Skeletal muscle fiber types in the ghost crab, *Ocypode quadrata*: implications for running performance. *Journal of Experimental Biology* **212**: 673–683.

Pollard TD (2008) Progress towards understanding the mechanism of cytokinesis in fission yeast. *Biochemical Society Transactions* **36**: 425–430.

Postel U, Thompson F, Barker G, Viney Mm Morris S (2010) Migration-related changes in gene expression in leg muscle of the Christmas Island red crab *Gecarcoidea natalis*: seasonal preparation for long-distance walking. *Journal of Experimental Biology* **213**: 1740–1750.

Potapov V, Sobolev V, Edelman M, Kister A, Gelfand I (2004) Protein-protein recognition: juxtaposition of domain and interface cores in immunoglobulins and other sandwich-like proteins. *Journal of Molecular Biology* **342**: 665–679.

Prinjha RK, Shapland CE, Hsuan JJ, Totty NF, Mason JJ et al. (1994) Cloning and sequencing of cDNAs encoding the actin cross-linking protein transgelin defines a new family of actin-associated proteins. *Cell Motility and the Cytoskeleton* **2**8: 243–255.

Rawe VY, Payne C, Schatten G (2006) Profilin and actin-related proteins regulate microfilament dynamics during early mammalian embryogenesis. *Human Reproduction* **21**: 1143–1153.

Robalino J, Carnegie RB, O'Leary N, Ouvry-Patat SA, de la Vega E, et al. (2009) Contributions of functional genomics and proteomics to the study of immune responses in the Pacific white leg shrimp *Litopenaeus vannamei*. *Veterinary Immunology and Immunopathology* **128**: 110–118.

Rozen S, Skaletsky H (2000) Primer3 in the WWW for general users and for biologist programmers. *Methods in Molecular Biology* **132**: 365–386.

Ruddell CJ, Wainwright G, Geffen A, White MR, Webster SG, et al. (2003) Cloning, characterization and developmental expression of a putative farnesoic acid O-methyltransferase in the female edible crab *Cancer pagurus*. *Biological Bulletin* **205**: 308–318.

Rush MG, Drivas G, D'Eustaschi P (1996) The small nuclear GTPase Ran: how much does it run? *Bioessays* **18**: 103–112.

Sagi A, Homola E, Laufer H (1991) Methyl farnesoate in the prawn *Macrobrachium rosenbergii*: synthesis by the mandibular organ in vitro, and titers in the hemolymph. *Comparative Biochemistry and Physiology Part B* **99**: 879–882.

Scheffzek K, Klebe C, Fritz-Wolf K, Kabsch W, Wittinghofer A (1995) Crystal structure of the nuclear Ras-related protein Ran in its GDP-bond form. *Nature* **374**: 378–381.

Schiaffino S, Reggiani C (1996) Molecular diversity of myofibrillar proteins: gene regulation and functional significance. *Physiological Reviews* **76**: 371–423.

Schwantes VS, Diana JS, Yi Y (2009) Social, economic, and production characteristics of giant freshwater prawn *Macrobrachium rosenbergii* culture in Thailand. *Aquaculture* **287**: 120–127.

See LM, Tan SG, Hassan R, Siraj SS, Bhassu S (2009) Development of microsatellite markers from an enriched genomic library for the genetic analysis of the Malaysian giant freshwater prawn, *Macrobrachium rosenbergii*. *Biochemical Genetics* **47**: 722–726.

Smith TF, Gaitatzes C, Saxena K, Neer EJ (1999) The WD repeat: a common architecture for diverse functions. *Trends in Biochemical Sciences* **24**: 181–1885.

Solway J, Seltzer J, Samaha FF, Kim S, Alger LE *et al*. (1995) Structure and expression of a smooth cell-specific gene, SM22 alpha. *Journal of Biological Chemistry* **270**: 13460–13469.

Soonthornchai W, Rungrassamee W, Karoonuthaisiri N, Jarayabhand P, Klingunga S *et al*. (2010) Expression of immune-related genes in the digestive organ of shrimp, *Penaeus monodon*, after an oral infection by *Vibrio harveyi*. *Developmental and Comparative Immunology* **34**: 19–28.

Staelens J, Rombaut D, Vercauteren I, Argue B, Benzie J *et al*. (2008) High-density linkage maps and sex-linked markers for the black tiger shrimp (*Penaeus monodon*). *Genetics* **179**: 917–925.

Strasser P, Gimona M, Moessler H, Herzog M, Small JV (1993) Mammalian calponin: Identification and expression of genetic variants. *LEBS Letters* **330**: 13–18.

Su J, Oanh DTH, Lyons RE, Leeton L, van Hulten MCW *et al*. (2008) A key gene of RNA interference pathway in the black tiger shrimp, *Penaeus monodo*: Identification and fuctional characterization of Dicer-1. *Fish and Shellfish Immunology* **24**: 223–233.

Sung HH, Yang CW, Lin YH, Chang PT (2009) The effect of two CpG oligodeoxynucleotides with different sequences on haemocytic immune responses of giant freshwater prawn, *Macrobrachium rosenbergii*. *Fish and Shellfish Immunology* **26**: 256–263.

Sutherland LC, Rintala-Maki ND, White RD, Morin CD (2005) RNA binding motif (RBM) proteins: a novel family of apoptosis modulators? *Journal of Cellular Biochemistry* **94**: 5–24.

Tamone SL, Prestwich GD, Chang ES (1997) Identification and characterization of methyl farnesoate binding proteins from the crab, *Cancer magister*. *General and Comparative Endocrinology* **105**: 168–175.

Tangprasittipap A, Tiensuwan M, Withyachumnarnkul B (2010) Characterization of candidate genes involved in growth of black tiger shrimp. *Aquaculture* **307**: 150–156.

Tassanakajon A, Klinbunga S, Paunglarp N, Rimphanitchayakit V, Udomkit A *et al*. (2006) *Penaeus monodon* gene discovery project: The generation of an EST collection and establishment of a database. *Gene* **38**4: 104–112.

The Gene Ontology Consortium (2000) Gene Ontology: Tool for the unification of biology. *Nature Genetics* **25**: 25–29.

The Gene Ontology Consortium (2008) The Gene Ontology project in 2008. *Nucleic Acids Research* **36**: D440–444.

Teichmann SA, Chothia C (2000) Immunoglobulin superfamily proteins in *Caenorhabditis elegans*. *Journal of Molecular Biology* **296**: 1367–1383.

Thanh NM, Nguyen NH, Ponzoni RW, Vu NT, Barnes AC *et al*. (2010a) Estimates of strain additive and non-additive genetic effects for growth traits in a diallel cross of three strains of giant freshwater prawn (*Macrobrachium rosenbergii*) in Vietnam. *Aquaculture* **299**: 30–36.

Thanh NM, Barnes AC, Mather PB, Yutao L, Lyons RE (2010b) Single nucleotide polymorphisms in the actin and crustacean hyperglycemic hormone genes and their correlation with individual growth performance in giant freshwater prawn *Macrobrachium rosenbergii*. *Aquaculture* **301**: 7–15.

Thanh NM, Ponzoni RW, Nguyen NH, Vu NT, Barnes A *et al*. (2009) Evaluation of growth performance in a diallel cross of three strains of giant freshwater prawn (*Macrobrachium rosenbergii*) in Vietnam. *Aquaculture* **287**: 75–83.

Towers GJ (2007) The control of viral infection by tripartite motif proteins and cyclophilin A. *Retrovirology* **4**: 40–50.

Van der Horst DJ (1990) Lipid transport functions of lipoproteins in flying insects. *Biochimica et Biophysica Acta* **1047**: 195–211.

Vera JC, Wheat CW, Fescemyer HW, Frilander MJ, Crawford DL *et al*. (2008) Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. *Molecular Ecology* **17**: 1636–1647.

Verheyen EM, Cooley L (1994) Profilin mutations disrupt multiple actin-dependent processes during *Drosophila* development. *Development* **120**: 717–728.

Vogt T (2010) Pnenylpropanoid biosynthesis. *Molecular Plant* **3**: 2–20.

Wang J-PZ, Lindsay BG, Leebens-Mack J, Cui L, Wall K *et al*. (2004) EST clustering error evaluation and correction. *Bioinformatics* **20**: 2973–2984.

Wang Z, Ding Q, Yagi KJ, Tobe SS (1994) Terminal stages in juvenile hormone biosynthesis in corpora allata of *Diploptera punctata*: development changes in enzyme activity and regulation by allatostatins. *Journal of Insect Physiology* **40**: 217–223.

Witke W, Podtelejnikov AV, Di Nardo A, Sutherland JD, Gurniak CB *et al*. (1998) In mouse brain profilin I and profilin II associate with regulators of the endocytic pathway and actin assembly. *EMBO Journal* **17**: 967–976.

Wheat CW (2008) Rapidly developing functional genomics in ecological model systems via 454 transcriptome sequencing. *Genetica* **138**: 433–451.

Wolfe Sa, Nekludova L, Pabo CO (2000) DNA recognition by Cys2His2 zinc finger proteins. Annu Rev Biophys Biomol Struct 29: 183–212.

Wolska BM, Wieczorek DM (2003) The role of tropomosin in the regulation of myocardial contraction and relaxation. *Pflügers Arch* **446**: 1–8.

Wu P, Qi D, Chen L, Zhang H, Zhang X *et al*. (2009) Gene discovery from an ovary cDNA library of oriental river prawn *Macrobrachium nipponense* by ESTs annotation. *Comparative Biochemistry and Physiology Part D* **4**: 111–120.

Wu Q, Luo Y, Lu R, Lau N, Lai EC *et al*. (2010) Virus discovery by deep sequencing and assembly of virus-derived small silencing RNAs. *Proceedings of the National Academy of Sciences of the USA* **107**: 1606–1611.

Ye J, Zheng H, Zhang Y, Chen J, Zhang Z *et al*. (2006) WEGO: a web tool for plotting GO annotations. *Nucleic Acids Research* **34**: W293–297.

Zhu X, Dai Z, Liu J, Yang W (2005) Actin gene in prawn, *Macrobrachium rosenbergii*: Characteristics and differential tissue expression during embryonic development. *Comparative Biochemistry and Physiology Part B* **140**: 599–605.

Ziegler J, Facchini PJ (2008) Alkaloid biosythensis: metabolism and trafficking. *Annual Review of Plant Biology* **59**: 735–769.

Zimmerman AW, Veerkamp JH (2002) New insights into the structure and function of fatty acid-binding proteins. *Cellular and Molecular Life Sciences* **59**: 1096–1116.

TABLES and FIGURES

Table 1. Summary of 454 pyrosequencing, assembly and analysis of *M. rosenbergii* transcriptomic sequences.

Dataset name		**All**	**Muscle**	**Ovary**	**Testis**
Total number of bases (Mp)		244.37	114.38	86.06	43.94
Average read length (bp)		310	311	308	311
No. of reads	Total	787,731	367,379	279,393	140,959
	Assembled	645,837	323,044	189,771	112,271
	Singleton	115,123	33,622	77,455	24,995
	Repeat	276	136	197	77
No. of contigs	Total contigs	8,411	1,723	5,346	1004
	Average contig read length (bp)	845	1,027	796	848
	Largest contig (bp)	7,531	7,304	6,955	7,530
	No. of large Contigs > 500bp	5,724	1,171	3,559	683
Average coverage (x)		29.85	59.56	14.92	43.09

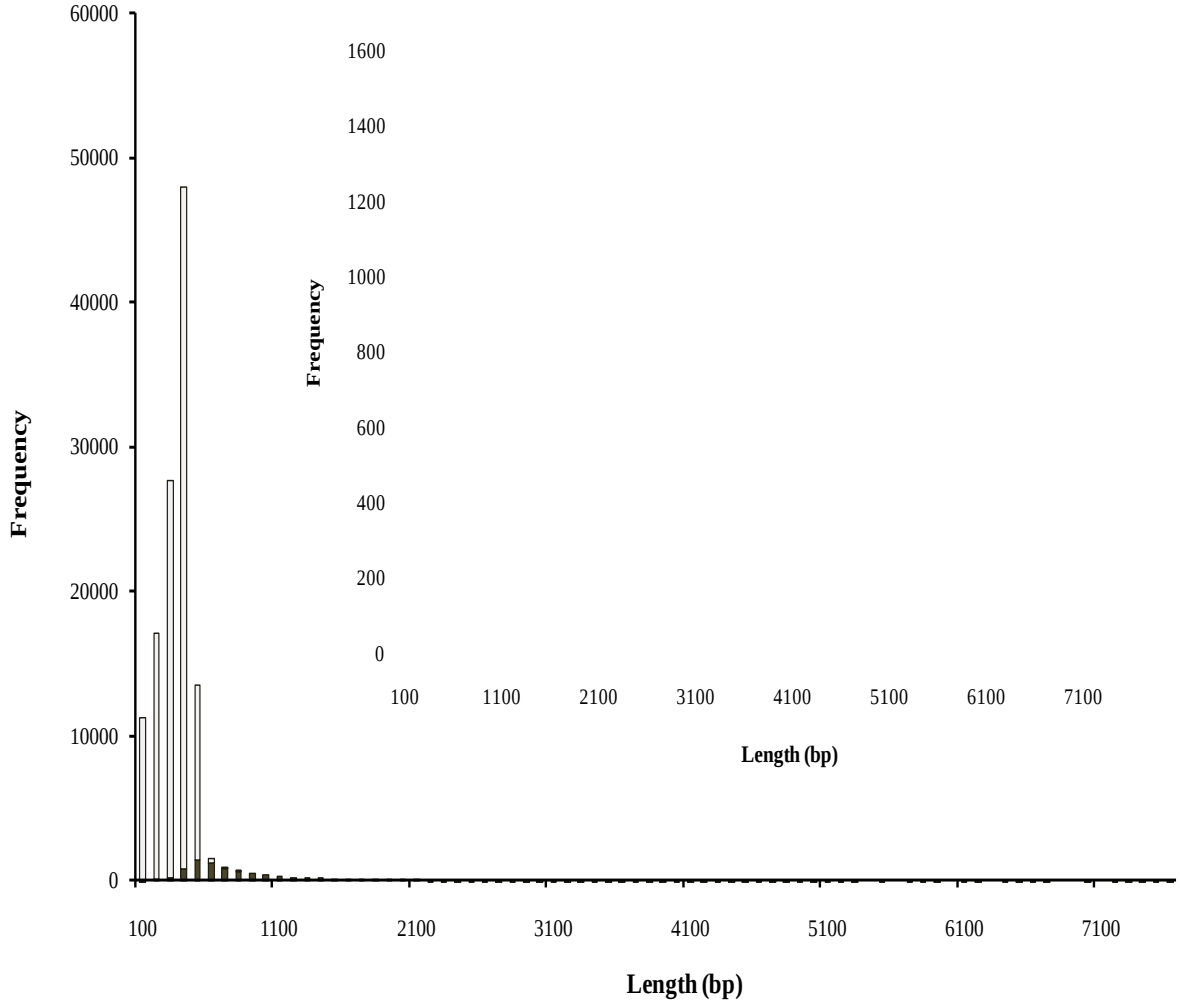


Figure 1. Summary of *M. rosenbergii* transcriptomic sequences. The contig sequences are represented by solid bars and the singleton sequences by open bars.

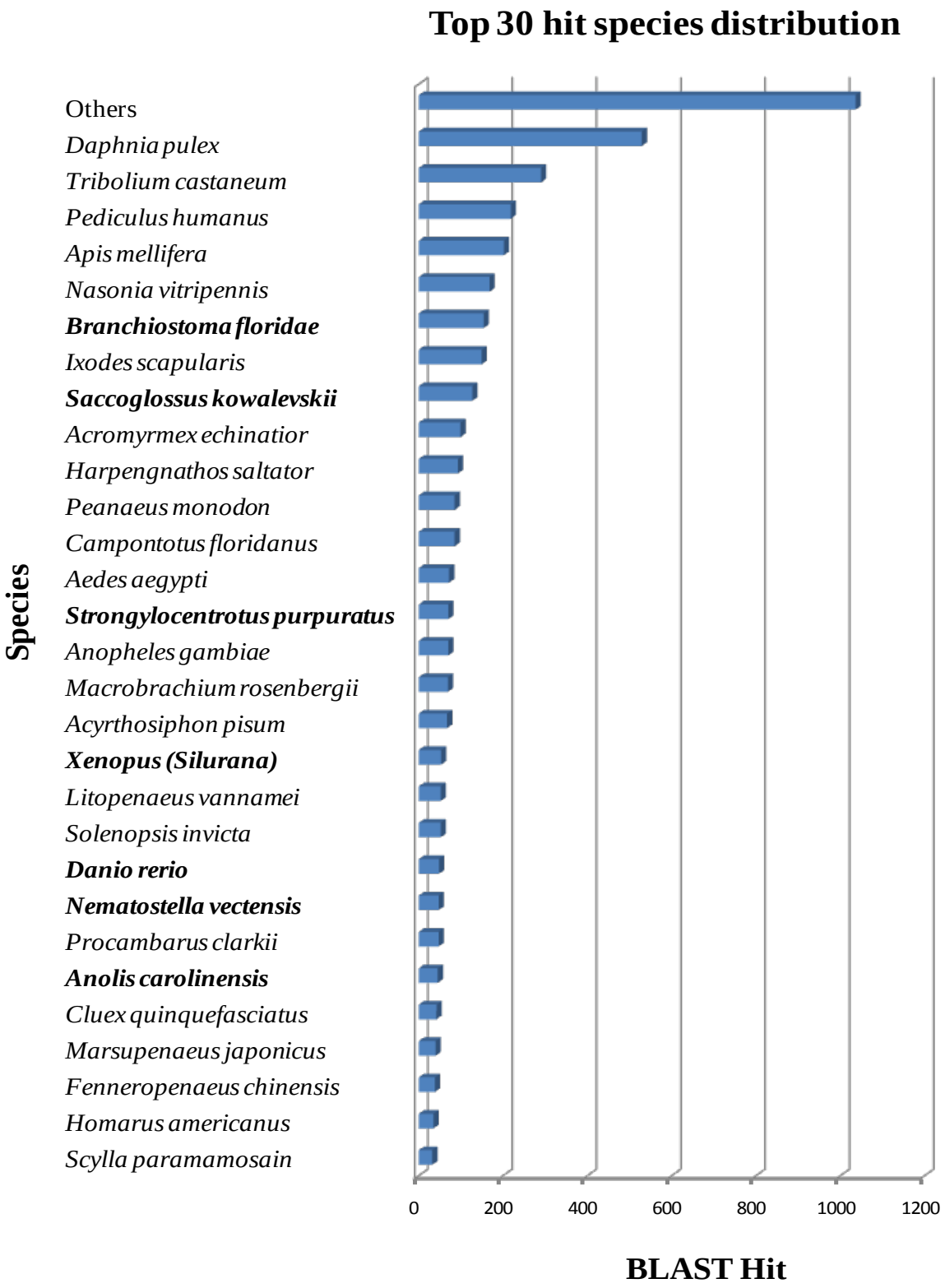


Figure 2. Top 30 hit species distribution based on BLASTx. E value cut-off is $1e^{-5}$ and top 30 hit species distribution of gene annotations showing high homology to the Arthropoda (Insecta and Crustacea) phylum with known genome sequences. Only contig sequences were used. Bold text indicates non-Arthropod homology.

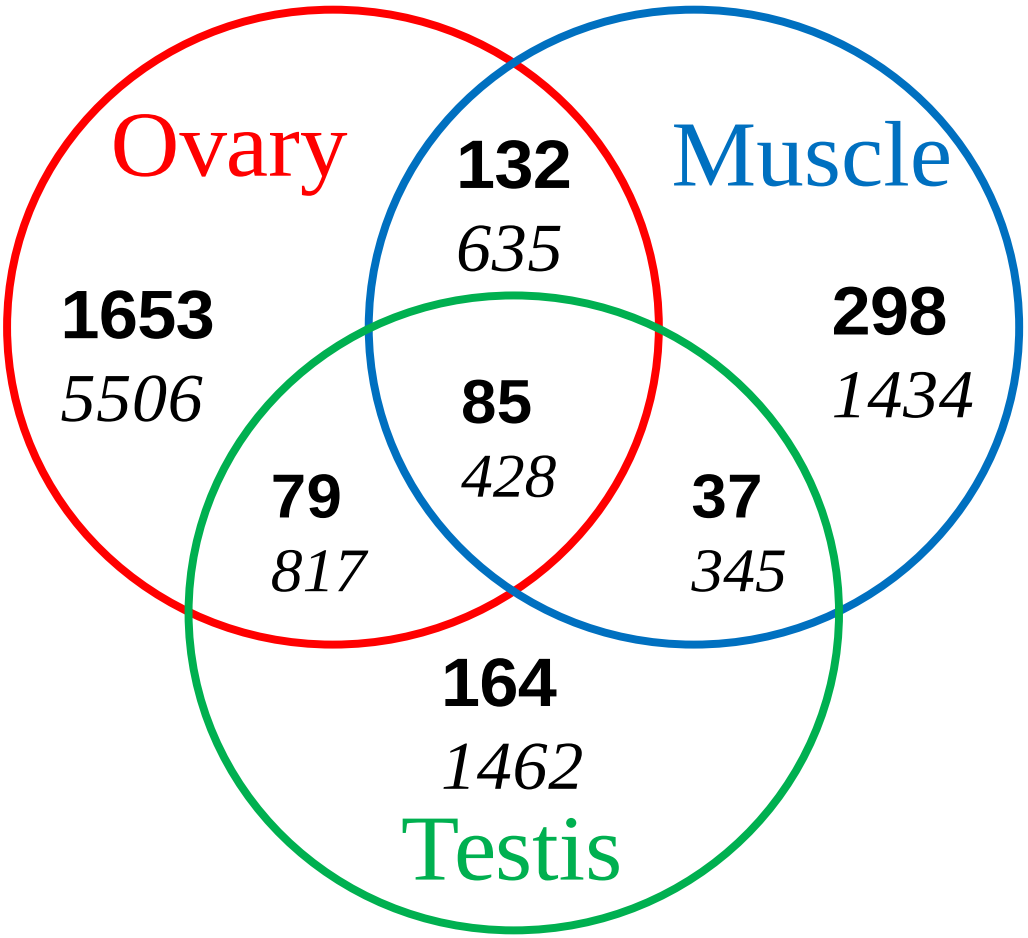


Figure 3. Comparative summary of *M. rosenbergii* transcriptomic sequences among three libraries. Putative sequence descriptions were counted using BLASTx results (E-value $<1e^{-5}$) after excluding ribosomal proteins and redundant ones. Bold numbers indicate contigs and numbers in italics indicate singletons.

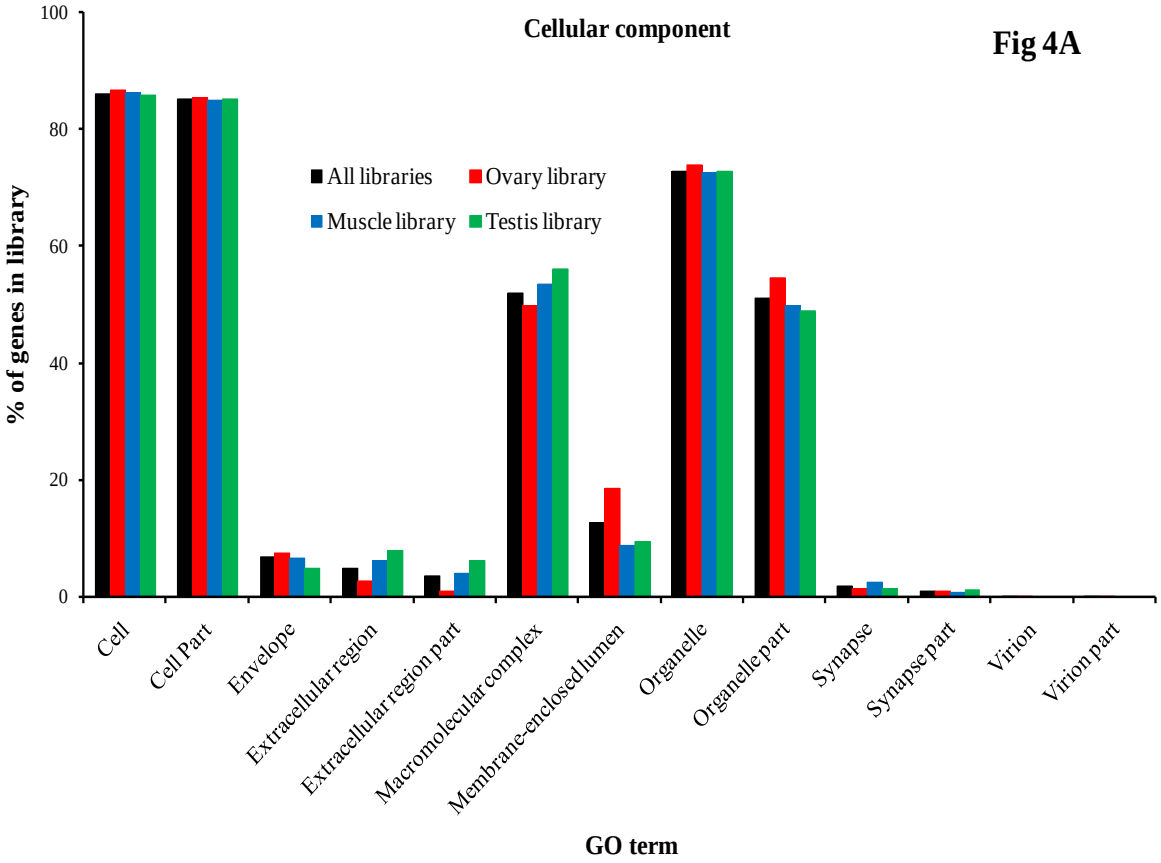

Cellular component
% of genes in library
100
80
60
40
20
0
All libraries
Ovary library
Muscle library
Testis library
Cell
Cell Part
Envelope
Extracellular region
Extracellular region part
Macromolecular complex
Membrane-enclosed lumen
Organelle
Organelle part
Synapse
Synapse part
Virion
Virion part
GO term

Fig 4A

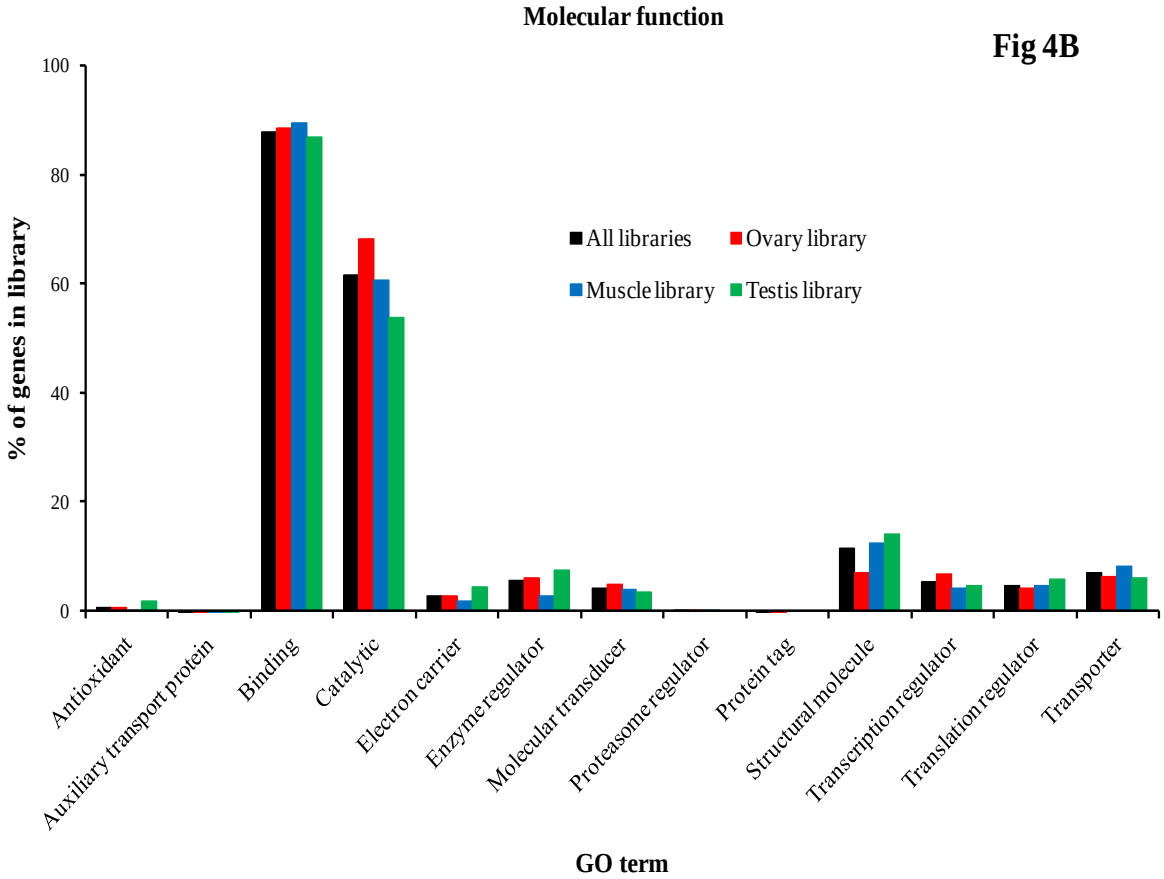

Molecular function
% of genes in library
100
80
60
40
20
0
All libraries
Ovary library
Muscle library
Testis library
Antioxidant
Auxiliary transport protein
Binding
Catalytic
Electron carrier
Enzyme regulator
Molecular transducer
Proteasome regulator
Protein tag
Structural molecule
Transcription regulator
Translation regulator
Transporter
GO term

Fig 4B

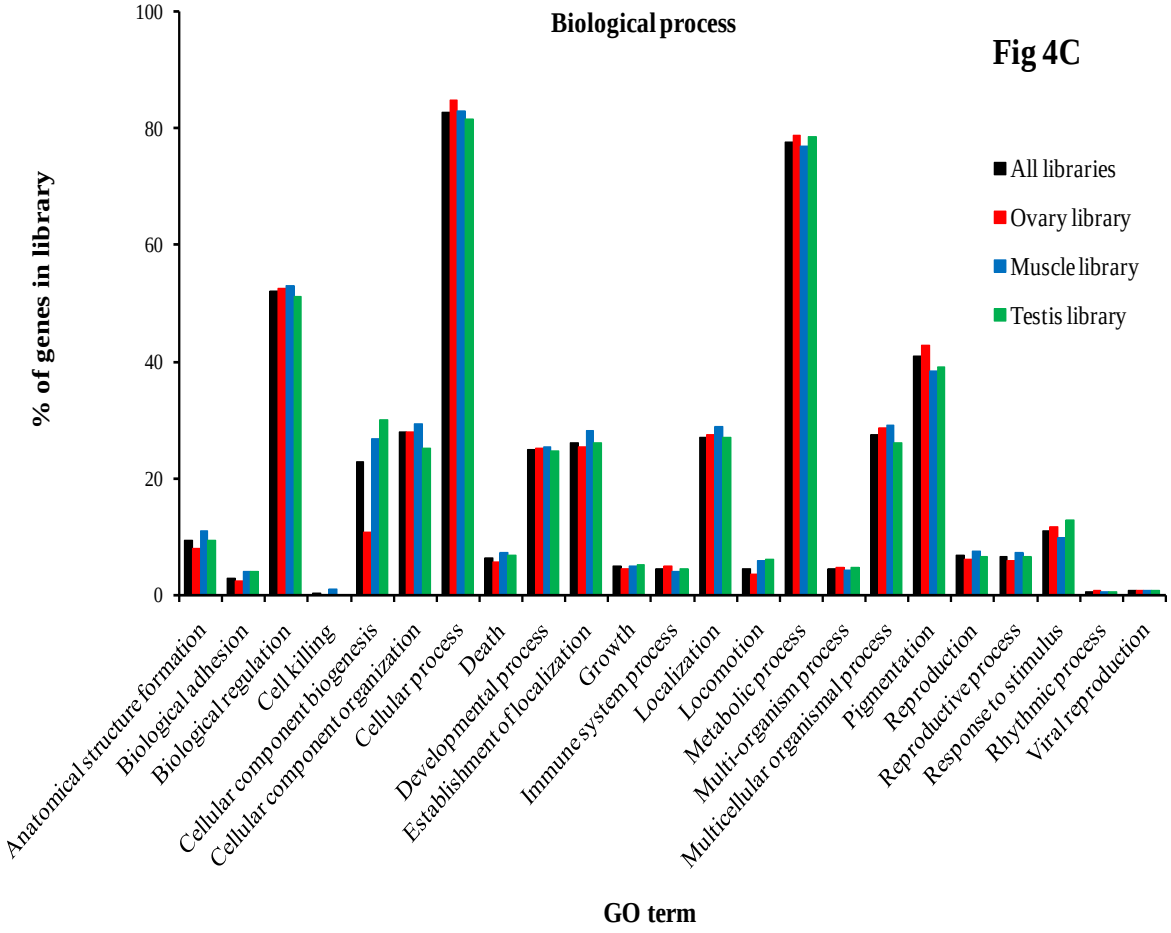


Figure 4. Gene ontology (GO) terms for the transcriptomic sequences of *M. rosenbergii* and comparison of among libraries. (A) cellular component, (B) molecular function and (C) biological process.

Table 2. Summary of top 20 domains predicted in *M. rosenbergii* sequences.

IPR accession	Domain name	Domain description	Occurrence
IPR000504	RRM_dom	RNA recognition motif domain	188
IPR017442	Se/Thr_prot_kinase-like_dom	Serine/threonine-protein kinase-like domain	180
IPR000719	Prot_kinase_cat_dom	Protein kinase, catalytic domain	180
IPR007087	Znf_C2H2	Zinc finger, C2H2-type	179
IPR004000	Actin-like	Actin-like	150
IPR002290	Ser/Thr_prot_kinase_dom	Serine/threonine-protein kinase domain	144
IPR012677	Nucleotide-bd_a/b_plait	Nucleotide-binding, alpha-beta plait	112
IPR016024	ARM-type_fold	Armadillo-type fold	108
IPR015943	WD40/YVTN_repeat-like_dom	WD40/YVTN repeat-like-containing domain	104
IPR011009	Kinase-like_dom	Protein kinase-like domain	104
IPR015880	Znf_C2H2-like	Zinc finger, C2H2-like	102
IPR017986	WD40_repeat_dom	WD40-repeat-containing domain	90
IPR011046	WD40_repeat-like_dom	WD40 repeat-like-containing domain	88
IPR002041	Ran_GTPase	Ran GTPase	86
IPR008271	Ser/Thr_prot_kinase_AS	Serine/threonine-protein kinase, active site	84
IPR013783	Ig-like_fold	Immunoglobulin-like fold	84
IPR011989	ARM-like	Armadillo-like helical	84
IPR023796	Sepin_dom	Serpin domain	79
IPR013083	Znf_RING/FYVE/PHD	Zinc finger, RING/FYVE/PHD-type	72
IPR016040	NAD(P)-bd_dom	NAD(P)-binding domain	72

Table 3. Genes of interest for growth and muscle development in *M. rosenbergii* sequences.

Candidate genes	Contig IDs	Length (bp)
Actin	A000585; A000586; A000587; A000588; A000763; A000764; A000765; A000766; A001338; A001339	1329; 1318; 1306; 1295; 1679; 1676; 1641; 1638; 930; 738
Alpha skeletal muscle	A000008; A000407; A000408; A000807; A002601; A002969	710; 1141; 1016; 1110; 1474; 1166
Calponin/calponin transgelin	A002718; A002875; A006133	1383; 1232; 518
Cyclophilin a	A001348; A001349	811; 850
Farnesoic acid O-methyltransferase	A002527	1587
Fatty acid binding protein	A004382	728
Lim domain binding	A000448	2610
Muscle lim protein	A000421; A000422; A000423; A000424; A000425	5788; 5694; 4595; 4501; 1716
Myosin heavy chain	A000009; A000016; A001103; A001282; A001283; A002073; A003870; A004348; A004442; A007715; A008193	612; 1510; 672; 942; 916; 711; 828; 733; 717; 383; 277
Myosin heavy nonmuscle or smooth muscle	A000018; A000968; A000969; A001363; A008338; A008352	1512; 5609; 2201; 730; 156; 148
Myosin light chain	A008264; A008271; A008339	220; 209; 155
Myosin light chain smooth muscle	A000639; A000783; A000785	3309; 2919; 1544
Profilin	A002454; A003703	1696; 872
Skeletal muscle actin 6	A000022; A000409; A005595; A006187; A007119; A008308; A008374	853; 522; 574; 512; 433; 177; 122
Transforming growth factor beta regulator 1	A006817	458
Tropomyosin	A000105; A000106; A000107; A000108; A000109; A000110; A000111; A000112; A000113; A000114; A000115; A000116; A001463; A002025; A002026; A007719	2777; 2769; 2770; 2768; 2762; 2760; 2773; 2775; 2765; 2767; 1962; 1954; 377; 1391; 110; 383

* Prefix "A" in ContigIDs indicates all merged contig from three libraries.

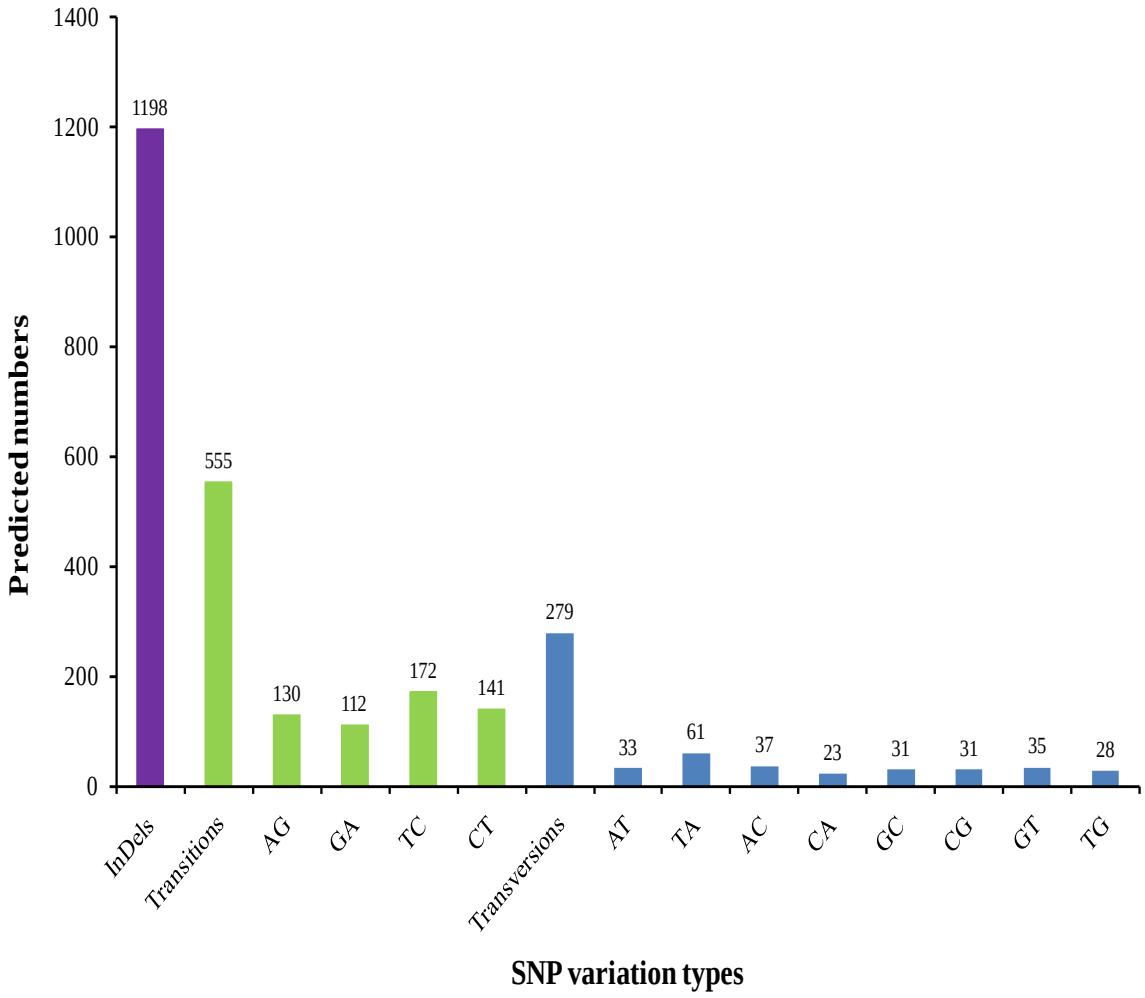


Figure 5. Distribution of putative single nucleotide polymorphisms (SNP) and indels in *M. rosenbergii* sequences.

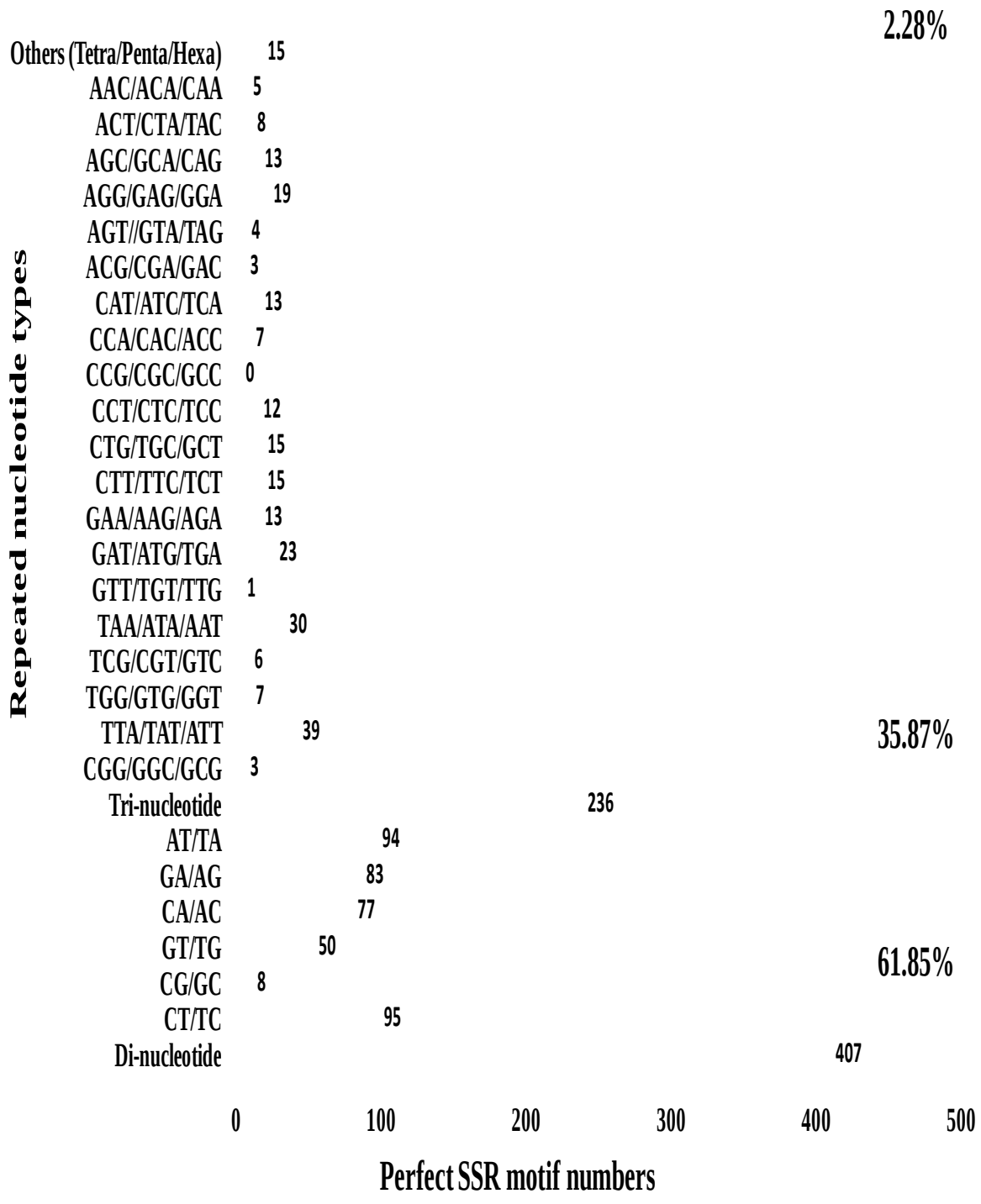


Figure 6. Distribution of simple sequence repeat (SSR) nucleotide classes among different nucleotide types found in *M. rosenbergii* sequences. Both contig and singleton sequences are used to predict the SSR loci.

CAHPTER 4: A candidate gene association study for growth performance in an improved giant freshwater prawn (*Macrobrachium rosenbergii*) culture line

- Presented at the 2012 World Aquaculture Society 'Global aquaculture securing our future' 1-5 September 2012, Prague, Czech Republic.

- Accepted in Marine Biotechnology (2013).

Preface to Chapter 4

This study was undertaken to assess correlations between a set of candidate gene SNP markers (CGSMs) in growth-related candidate genes with individual growth performance in an improved GFP culture line. 28 SNPs from 23 growth-related candidate genes were confirmed using SEQUENOM in broodstock representing fast and slow growth performance GFP lines in Vietnam (N=200). Significant correlations with expected breeding values (EBVs) within subjected samples by analysing interactions of multiple CGSMs with pairwise interactions for markers in additive model were observed. This is the first largest set of CGSMs reported for *M. rosenbergii* and will be useful in future genetic analyses including for dissecting of complex traits, tracing relationships in breeding programs, and monitoring genetic diversity in commercial and wild populations of *M. rosenbergii*. In particular, some unique SNPs identified across diverse geographical panels may also be useful in other *Macrobrachium* species and provide nuclear markers for systematic studies and for detection of interspecific hybridization among congeners.

A candidate gene association study for growth performance in an improved giant freshwater prawn (*Macrobrachium rosenbergii*) culture line

Jung H[1*], Lyons RE[2], Li Y[2], Thanh NM[3], Dinh H[1, 4], Hurwood DA[1], Mather PB[1]

[1]Earth, Environmental and Biological Sciences, Science and Engineering Faculty, Queensland University of Technology, Brisbane, Queensland 4000, Australia

[2]Commonwealth Scientific and Industrial Research Organisation, Livestock Industries, St Lucia, Queensland 4067, Australia

[3]School of Biotechnology, International University of Vietnam National University, Ho Chi Minh City, Vietnam

[4]Research Institute for Aquaculture No.2, Ho Chi Minh City, Vietnam

ABSTRACT

A candidate gene approach using Type I single nucleotide polymorphism (SNP) markers can provide an effective method for detecting genes and gene regions that underlie phenotypic variation in adaptively significant traits. In the absence of available genomic data resources, transcriptomes were recently generated in *Macrobrachium rosenbergii* to identify candidate genes and markers potentially associated with growth. The characterisation of 47 candidate loci by ABI re-sequencing of 4 cultured and 8 wild samples revealed 342 putative SNPs. Among these, 28 SNPs were selected in 23 growth-related candidate genes to genotype in 200 animals selected for improved growth performance in an experimental GFP culture line in Vietnam. The associations between SNP markers and individual growth performance were then examined. For additive and dominant effects, a total of three exonic SNPs in glycogen phosphorylase (additive), heat shock protein 90 (additive and dominant) and peroxidasin (additive), and a total of six intronic SNPs in ankyrin repeats-like protein (additive and dominant), rolling pebbles (dominant), transforming growth factor-β induced precursor (dominant) and UTP-glucose-1-phosphate uridylyltransferase 2 (dominant) genes showed significant associations with the estimated breeding values (EBVs) in the experimental animals ($P = 0.001−0.031$). Individually, they explained 2.6−4.8% of the genetic variance ($R^2 = 0.026−0.048$). This is the first large set of SNP markers reported for *M. rosenbergii* and will be useful for confirmation of associations in other samples or culture lines as well as having applications in marker-assisted selection in future breeding programs.

Keywords: *Macrobrachium rosenbergii*, transcriptome, ABI re-sequencing, growth-related candidate genes, SNPs, associations, estimated breeding values

INTRODUCTION

Growth is an important fitness trait fundamental to proliferation in all living cells and animal development, and is a factor that can affect survival and fecundity in a wide range of taxa (Arendt 1997). Selecting for optimal growth rate or body weight is a key breeding objective in many aquaculture stock improvement programs. Growth and/or growth rate is a quantitative trait that can be correlated with body weight under a polygenic inheritance model, and can show moderate heritability. It can also be influenced by both prenatal and postnatal maternal effects (Falconer 1996), life-history traits including development rate and size at specific ages (Mangel & Stamps 2001), and environmental factors (Sundstrom *et al*. 2005). Currently a major challenge in biology is to understand the genetic basis of variation in quantitative traits because clearly defined relationships between variation in genotypes and phenotypes can provide crucial information for increasing the rate of genetic gains in selective breeding programs in agriculturally important organisms and for predicting adaptive evolution (Mackay *et al*. 2009). In crustaceans, a number of internal and external factors in addition to genetic factors can also contribute to phenotypic variation for growth. These include nutritional status, stocking density, water temperature and salinity, dissolved oxygen levels, age, sex, social rank (i.e. competition), moulting stage, seasonal changes, and/or hormonal expression (Jung *et al*. 2013). Most factors have not, however, been studied in any detail in most species produced in commercial aquaculture. A few potential candidate genes associated with individual growth performance have been identified, with only actin and crustacean neuropeptide family (crustacean hyperglycaemic hormone [CHH] and moult-inhibiting hormone [MIH]) having been studied with respect to growth in more than a single crustacean species (De-Santis & Jerry 2007; Tangprasittipap *et al*. 2010; Jung *et al.* 2013). Despite rapid progress with DNA sequence analysis, little is known currently about the genetic control of growth in crustaceans.

Understanding multi-gene interactions affecting growth will require a large number of gene-associated markers. For many species, ESTs provide a primary resource of nuclear DNA sequence data for developing Type I molecular markers, in particular SNPs (Bouck & Visoion 2007). SNPs are biallelic gene markers that have become the marker of choice for many applications in genetics and genomics due to their abundance, high genotyping efficiency and data quality, and analytical simplicity (Kruglyak & Nickerson 2001; Morin *et*

al. 2004; Seeb *et al.* 2011). Recent advances in sequencing technologies have enabled the cost-effective and rapid generation of massive amounts of genomic data, including ESTs derived from transcriptomic studies in non-model organisms (Metzker 2010; Ekblom & Galindo 2011). SNP markers are becoming an increasingly important addition to molecular genetics and molecular ecology toolkits in many non-model organisms (Seeb *et al.* 2011) as well as for evaluating variation in quantitative trait loci (QTL).

QTLs are regions of the genome that affect polygenic phenotypic traits that show continuous variation (McClelland & Naish 2010) and can usually be located in the genome by conducting marker-trait association analysis using specific marker genotypes (Lynch & Walsh 1998). As a first step to identifying gene(s) with functional roles in growth, genome scans provide an effective method for detecting QTLs for growth traits when a fine linkage map is available. When a linkage map is not available, however, a candidate gene approach (CGA) is more practical for detecting genetic correlations between traits. In general, CGA can more precisely assess genotype-phenotype associations (Hu *et al*. 2009) than using anonymous markers (Lynch & Walsh 1998). The targeting of key candidate genes of known physiological function can enhance identification and characterisation of specific functional mutations that affect variation in a specific phenotype (De-Santis & Jerry 2007; Jung *et al.* 2013). Furthermore, a statistical correlation found between specific alleles in a candidate gene of known function, increases the likelihood of finding meaningful trait associations. Where significant associations are detected, this indicates that the gene is either involved directly in genetic control of the specified trait or that the functional polymorphism is sufficiently close to the marker so that the two loci are in linkage disequilibrium (Lynch & Walsh 1998). While CGA has been used to detect genotypic associations with phenotypic performance in several well-studied livestock and a few cultured fish species (Goddard & Hayes 2009; McClelland & Naish 2010; Wang *et al*. 2010), few genetic variants associated with quantitative traits have been identified in organisms where genetic and/or genomic information is limited or is not currently available. A recent study, however, applied 418 SNP markers in Pacific white shrimp, (*Litopenaeus vannamei*), to develop a gene map and to identify QTLs for economically important traits (Du *et al*. 2010). A number of SNPs in anonymous genes were reported to show significant associations with weight gain, growth rate, survival, and/or pathogen resistance in these populations (Ciobanu *et al.* 2010). This study demonstrates the value of identifying markers located within functional genes. While genomic data for marine shrimps has been expanding rapidly, the same is not true for

freshwater prawns used in culture. A recent practical strategy, however, suggested by Jung *et al.* (2013) for identifying a number of growth-related candidate genes has provided an excellent starting point for future genomic studies of giant freshwater prawn (GFP) growth performance.

M. rosenbergii (GFP) supports one of the largest culture industries for any crustacean species farmed in inland waters across the world with production value in 2009 exceeding US$ 1 billion annually across Asia alone (FAO 2010). Recently, research has been directed at improving the relative productivity of GFP culture lines to assist industry development (Schwantes *et al.* 2009; Thanh *et al.* 2010a). While a number of stock improvement programs have been initiated recently for GFP, notably in Vietnam and India (Dinh *et al.* 2012, 2013; Pillai *et al.* 2011), to date genetic information on this species is very limited. A traditional phenotype-based selection approach is the accepted method used to select for growth traits, but requires several generations to optimize genetic gains. To date, only a single CGA study has applied SNP markers to evaluate variation in quantitative trait loci (QTL) in a freshwater prawn species (Thanh *et al*. 2010b). This study identified four SNPs in the CHH gene that were associated significantly with three growth traits (body weight, carapace length and standard length). More gene-based markers will be required, however, to increase the accuracy of estimation of phenotypic traits. Recently, transcriptomes were generated from *M. rosenbergii* using 454 pyrosequencing and the analysis has identified a large number of ESTs as putative functional genes, including some potentially involved with growth and muscle development traits (Jung *et al.* 2011). These markers will need to be validated and tested, however, before being applied potentially to stock improvement.

The main objectives of the current study were: a) to validate SNPs previously identified in candidate genes and further characterise these candidate gene regions to identify other potentially useful markers; b) to generate a SNP panel for genotyping animals in a GFP culture stock; c) to use the resulting multi-locus genotypic data to undertake an association analysis of validated SNPs with growth traits in a GFP line selected for fast growth in Vietnam.

MATERIALS AND METHODS

Study populations and morphological data

A test panel comprising 12 GFP individuals was used to test SNP primers and for DNA re-sequencing. These included eight GFP individuals collected in Indonesia (Bengawan R), Thailand (Tapi R), Malaysia (Bahand R), Bangladesh (Meghna R), India (Gujarat Narmada R), India (Chalakudy R), and Vietnam (Dongnai R), and four individuals from a hatchery stock in Vietnam (characterised as fast and slow growth performance individuals) (Dinh *et al*. 2012, 2013).

In 2007, three GFP culture strains (two native Vietnamese, Dong Nai and Mekong, and an exotic strain from Malaysia) were sourced as founder populations at the Research Institute for Aquaculture No. 2 in Vietnam (RIA2). They were used to produce the G0 generation (year 2008), the G1 (year 2009), G2 (year 2010) and G3 (year 2011) generations, and all matings were made between genetically unrelated brood stock (in order to minimize inbreeding levels) to produce full-sib and (paternal) half-sib families using a complete 3 × 3 diallel cross (Dinh *et al*. 2012, 2013). In each generation, approximately 70 families that showed highest mean breeding value or estimated breeding values (EBVs) for harvest weight were chosen as founders for the next generation. From each family, five to eight males and females with the highest EBVs were selected as parents (the Selection line) for the following generation. Controls in each generation in the current study consisted of individuals within the selected line that had the mean EBVs for that generation.

Following the above strategy, approximately 1,200 prawns from the G2 (same age class) were cultured in a single pond (in order to minimize environmental effects) and were assessed for six growth traits at harvest, including body weight, body length, cephalothorax length, abdominal length, cephalothorax width and abdominal width (Dinh *et al*. 2013). The linear mixed animal model in the ASReml program (Gilmour *et al.* 2009) was used to examine the magnitude of both genetic and environmental effects (Dinh *et al*. 2013). Complete details of the environmental effects fitted as variables in the model, founder populations, mating design and family selection are available in Dinh *et al*. (2013). As a result, EBVs were derived for all G2 animals (Dinh *et al*. 2013). EBVs were then used to rank all individuals from families chosen at random. For the current study, a total of 200

extreme growth performance individuals, 100 individuals with the highest EBVs and 100 individuals with the lowest EBVs from the selected line, were sampled for the SNP association studies. The 200 individuals were taken from the 89 families in the selected line based on EBV ranking. Individuals were then genotyped as growth-rate extreme classes (top and bottom families within the selected line) for the SNP trait association study. Each family contributed 1 to 7 individuals to the 200 individuals genotyped. An analysis of skewness produced an estimate of 0.31 with a Kurtosis value of -0.91 (P = 0.04). Further statistical analyses were conducted using only the highest and lowest growth performance individuals ranked by EBV within the selected line. Thus, the Control for high growth individuals was individuals with the mean population performance in each generation at harvest.

PCR primer design

Two criteria were used to select sequences for design of PCR primers from EST sequences taken from a recent GFP transcriptome dataset (Jung *et al.* 2011). They were; the contigs must contain SNPs identified in the initial 454 pyrosequencing run, and must have a putative function related to muscle development and growth (Jung *et al.* 2011, 2013). All selected contigs were annotated provisionally (i.e. exon, intron and untranslated regions identified) via BLASTx and ORF searches of public databases. Oligo 7 software (Molecular Biology Insights) was used to design primers to amplify short products (100-600 bp) including the SNP sites screened initially in the 454 sequencing runs. For longer contigs that had putative functions related to muscle development and growth, the forward primer was designed in the 3' end of the coding sequence and the reverse primer in the 3' untranslated region (UTR) to minimise the potential for amplifying anonymous introns. Primer pairs were designed to amplify 100 SNP loci.

DNA extraction and sequencing

Tissues were preserved in 95 % ethanol and stored at room temperature. DNA was extracted from periopods (walking legs) tissue using a modified salt extraction method (Miller *et al.* 1998). All amplification reactions were performed in 35 µl reaction volumes in an Eppendorf Mastercycler EPgradientS thermocycler (Eppendorf). Each reaction contained 25 µl dH_2O, 8 µl 5× MyTaq Red Reaction Buffer (BIOLINE), 0.5 µl each primer (10 µM), and 0.1 µl

MyTaq HS DNA Taq polymerase (BIOLINE). All PCR reactions consisted of an initial denaturation of 5 min at 95 °C, followed by 35 cycles of denaturation at 95 °C for 45 s, annealing at 57−64 °C (depending on the amplicon, Table 1) for 45 s, extension at 72 °C for 45 s, and a final extension of 10 min at 72 °C. PCR products were examined by electrophoresis on 1.5 % agarose gels stained with Gel Red Nucleic Acid Stain (Biotium) to confirm amplification of the target region. Primer sequences and amplicon details are summarised in Table 1.

Amplicons were purified for sequencing using PCR product clean up kit (BIOLINE) according to the manufacturer's protocol. Sequencing of each sample was conducted in both directions with the primers used for amplification and the ABI BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystem) under the following conditions: initial denaturation of 1 min at 96 °C, followed by 30 cycles of 96 °C for 10 s, 50 °C for 5 s, and 60 °C for 4 min. Sequences were analysed on a ABI 3500 capillary sequencer (Applied Biosystem).

Sequences were edited and aligned using Lasergene v10 software (DNASTAR). Forward and reverse sequences were aligned to produce a consensus sequence including 454 contigs and haplotypes. A multiple sequence alignment of individual consensus sequences including the sequence used to design the primers was undertaken to detect SNPs and insertions/deletions (indels). Haplotype and nucleotide diversity (π) were estimated from diploid sequences using the PHASE algorithm (Stephens & Donnelly 2003) implemented in Dnasp 5.10 (Librado & Rozas 2009). SNPs were classified as coding, intron, or UTR SNPs. Total number of non-synonymous (Ka), synonymous (Ks), transitions (Ti) and transversions (Tv) were calculated in MEGA5 (Tamura *et al*. 2011).

Molecular data used in association studies

A total of 30 candidate gene loci were selected from transcriptomic data generated in an earlier study (Jung *et al*. 2011). Candidate gene selection was based on four criteria: 1) associations between genes and growth previously reported in crustaceans, 2) growth-related genes involved with moulting, 3) muscle development and degradation genes involved in moulting, and 4) demonstrated association with growth traits in farmed terrestrial animal species for which orthologs were present in crustacean species. Further details are provided in

Jung *et al*. (2013). SNPs were selected for screening based on three criteria: 1) SNPs in coding rather than in non-coding regions, 2) non-synonymous *vs* synonymous mutations, 3) SNPs identified in both 454 pyrosequencing and ABI-resequencing analyses. DNA samples from 200 adult females with extreme growth phenotypes (fast and slow) were sent to the Australian Genome Research Facility (AGRF) in Brisbane for allele separation. Thirty pre-selected SNPs were multiplexed using a MassARRAY platform according to the manufacturer's protocols (SEQUENOM) and manual inspection was conducted at AGRF using MassARRAY TYPERAnalyzer v3.4 (SEQUENOM). Assay information is presented in Table 2.

Statistical analyses

To examine individual SNPs from the MassARRAY platform, allele frequencies were calculated and Hardy-Weinberg equilibrium (HWE) and linkage disequilibrium (LD) were tested in GENEPOP v4.0.10 with 1000 iterations and 100 Markov Chain approximations (Raymond & Rousset 1995: Rousset 2008). Where appropriate, significance values for each test were adjusted to account for family-wise error using Bonferroni correction (Rice 1989) and a False Discovery Rate Procedure (FDR) (Benjamin & Hochberg 1995; Verhoeven *et al.* 2005) for the analysis of individual SNP marker quality test, and for multiple comparisons via ANOVA (Tukey-Kramer) (Hayter 1984; Rafter *et al.* 2002) for the analysis of SNP association test.

All significant environmental effects on body weight of all animals were taken into account in the large genetic experiment (Dinh *et al*. 2012, 2013) and the genetic values of individuals (EBVs) were used to rank animals. Therefore, the EBVs of body weight of 200 extreme growth animals were used for the SNP trait association analyses. For single SNP-trait association analyses, both arbitrary dominant and additive effect models (Genissel *et al*. 2004) were investigated using the ANOVA of the EBVs. The dominance model (Model 1) was:

$$y_{ij} = \mu + SNP_{ij} + e_{ij} \quad \text{(Model 1)}$$

where y_{ij} is the trait measured in the *i*th individual; μ is the overall mean; SNP_{ij} is the fixed effect for the SNP genotype *j* (*j* = 1, 2 or 3 corresponding to GG, GA, or AA); and e_{ij} is the error term of the *j*th individual. The additive model (Model 2) was similar to the

dominance model except that SNP genotype was fitted as a covariate, corresponding to a simple linear regression of growth data on the number of "A" alleles present in each individual animal tested (i.e. the additive effect of allelic substitution). Both additive and dominance models were performed using PROC GLM in SAS package version 8 (SAS Institute 2002). Since the SNP was the only effect in the GLM model, the R^2 value in the GLM model indicates the proportion of genetic variance explained by the SNP. In order to test multiple SNP-trait associations, the analyses were also performed with all SNPs fitted simultaneously in the models and then using the stepwise selection method (both forward and backward) to identify a subset of SNPs that contributed most to genetic variation for body weight.

RESULTS

Candidate gene markers

As unrelated animals were used in the initial pyrosequencing experiment (Jung *et al*. 2011), an essential step was validation of putative SNPs in the current study population. To this end we utilized standard Sanger sequencing (ABI 3500) to confirm the predicted SNP in genomic DNA and in the process we could identify novel SNPs not previously characterized in the pyrosequencing, as they were either not present in the earlier cohort of animals or were present within non-coding elements (introns) not present in the initial transcriptomic data.

Successful PCR amplification of a single clear product was achieved in all tested individuals for 75 of the 100 target loci screened (75 %). Re-examination of the 25 PCRs that had failed to amplify suggested three reasons for failure; 1) the original PCR primers designed from EST data were anchored on an exon-intron boundary and/or alternative splice exon regions were unable to anneal to the genomic DNA template; 2) the presence of very large and/or numerous introns in the target amplicons interrupted effective amplification under standard screening conditions; or 3) priming site polymorphisms caused inferior or erratic amplification. From 75 amplicons, 47 were chosen to survey patterns of intraspecific polymorphisms after re-sequencing and 28 loci were not included due to large intron(s) insertion or sequencing difficulties due to presence of indels. While all 47 amplicons proved to be the correct (target) loci, two amplicons showed a possible alternative splice event and another amplicon showed a possible frame shift and/or alternative splice site (Table 1). Even

with these potential changes, the putative functions of the three loci were not affected based on BLASTx hit results. Summary of the SNPs and alignment of the sequenced amplicons (454 pyrosequencing and ABI re-sequencing) including fasta files are presented in Electronic supplementary material 1, 2. Both files are ordered alphabetically based on putative gene names (same order in Table 1) and sequence region numbers presented in an excel file and are based on actual sequenced regions from ABI re-sequencing data. In the case of the fasta file, each gene name came first and abbreviation of 454Con (consensus from 454 pyrosequencing), 454Hap (haplotype from 454 pyrosequencing), TF1 (top female1), TF2 (top female2), BF1 (bottom female1), BF2 (bottom female2), I3-35 (Indonesia3-35), T2-45 (Thailand2-45), M1-43 (Malaysia1-43), B2-24 (Bangladesh2-24), IG-33 (India Gujarat-33), V2-2 (Vietnam2-2), IDF1 (India Kerala 1) and IDF2 (India Kerala 2) was followed in order. In the case of the excel file, additional sample locations with abbreviation (including SNP calling, coding/non-coding/UTR regions and amino acid change) were added immediately before the same abbreviation used in the fasta file.

Of the 47 amplicons characterized (coding sequence region 9,158 bp total), 44 contained predicted coding sequences and of these, 43 contained SNPs in coding regions (127 synonymous and 75 non-synonymous) out of 342 putative SNPs. A single indel was observed across characterised sequences in the inhibitor of growth (ING) gene as a possible frame shift or alternative exon splice (Table 1). Intron sequences obtained from 15 amplicons (3,509 bp total) contained 126 SNPs and 29 indels, respectively. Six amplicons contained UTR (1,167 bp) that showed 14 SNPs and 22 indels, respectively. As expected, the most common SNP mutation types were C↔T (39.48 %) and A↔G (27.78 %), while incidences of other SNP types were significantly lower: C↔G (8.48 %), G↔T (8.48 %), A↔C (7.89 %), and A↔T (7.89 %) (Figure 1). Average π, Ka/Ks ratio and Ti/Tv ratio across all amplicons were 0.00722, 0.59 and 2.05, respectively. Analysis of individual amplicons is presented in Table 1. Overall, SNPs were observed in all but a single amplicon (Cyclophilin A). Re-sequencing of 47 amplicons using an ABI 3500 sequencer revealed a large number of regionally specific SNPs not identified in the 454 pyrosequencing run, but only 42 of 61 putative SNPs were confirmed with the ABI sequencer. In addition, two amplicons (β-actin and glycogenin) showed a possible alternative splice event and/or transposable element, and a single amplicon (ING) showed a possible frame shift and/or alternative splice site. In a comparison of SNP heterozygosity among samples, four cultured samples from the G2 showed signs of reduced

diversity compared with eight wild samples as 0.68/1 ratio (See Electronic supplementary material 2).

When PCR primers were designed in the middle sections of coding regions, this resulted in a large number of intron insertions as exon-primed intron-crossing (EPIC) markers. Of the 47 amplicons re-sequenced, 17 introns were observed in 15 amplicons and this included some amplicons that contained multiple introns (ANK-like protein and glutamate-gated chloride channel) (Table 1). Of the 17 introns identified as EPIC markers, 5 were inserted in the first codon, 2 were inserted in the second codon, and 10 were inserted in the third codon position, but exon-intron boundaries for all amplicons followed Chambon's rule (GT/AG) for splice donor and acceptor sites (Table 1).

SNP genotyping

Sequences for 28 of 30 pre-selected SNPs in candidate gene markers were successfully amplified and they were used to genotype 200 GFP female brood-stock individuals representing fast and slow growth phenotypes. The detailed SNP sources and positions are summarised in Table 2. Of the 28 SNPs screened, only a single SNP (C SNP in Ferritin) failed to show polymorphisms and the remaining 27 polymorphic SNPs were located in 18 exons (15 synonymous and 3 non-synonymous), eight in introns and one in a 3' UTR.

Results of observed and expected homozygosity and heterozygosity estimation at each SNP and the significance of HWE are summarised in Table 3. The majority of SNPs detected in the current study did not deviate ($P > 0.05$) from HWE except for ANK-like protein_A1 ($P < 0.05$) in the slow growth performance population, GluCl_A1 ($P < 0.001$), TGF-βIP_A2 ($P < 0.001$) in both slow and fast growth performance populations, and Zasp52 ($P < 0.001$) in the fast growth performance population. While the actual positions of each SNP in the GFP genome are not known, five pairs from different genes (NHCP/VAIPRCP, ANK-like protein_A2/Cathepsin L, Cathepsin L/Calreticulin, GP/ING, and ING/TGF-βIP_A2) and four pairs from the same genes (ANK-like protein, GluCl, TGF-βIP, and UGP2) showed significant LD ($P < 0.001$).

When comparing SNP allele frequencies between fast and slow growth individuals (Table 3), most observed SNP allele distributions were not significantly different between the two groups except for ANK-like protein_A1 ($P = 0.001$), ING ($P = 0.03$) and Zasp52 ($P =$

0.04) (Fisher's exact test). Interestingly more than three genotypes were observed at the ANK-like protein_A2 and TGF-βIP_A2 (high homology with Fasciclin) loci.

SNP-trait associations

Both additive and dominant genetic inheritance models were examined for associations between individual SNPs and the genetic values (EBVs) for body weight in 200 animals (Table 4). Under an additive model, two exonic SNPs in the GP (P = 0.022) and HSP90 (P = 0.002) genes, and two intronic SNPs in ANK-like protein_A1 (P = 0.021) and _T_A2 (P = 0.017) genes showed significant associations with the EBVs for body weight.

Individually they explained 2.6−4.7 % of the genetic variance for body weight (R^2 = 0.026 to 0.047, Table 4). Under a dominance model, the exonic SNP in the HSP90 gene (P = 0.002) and three intronic SNPs in ANK-like protein_A1 (P = 0.008) and _T_A2 (P = 0.018), and TGF-βIP_T_A2 (P = 0.017) genes were associated significantly with body weight (Table 4), explaining 4−4.8 % of the genetic variance (R^2 = 0.040 to 0.048). Both models identified that the exonic SNP in HSP90 and two intronic SNPs in ANK-like protein_A1 and _T_A2 (P = 0.017−0.021) had both significant additive and dominant genetic associations with body weight. In addition, the SNPs in ANK-like protein_A_A2, GAPDH, GP, ING and NCHP genes showed marginal associations with body weight (P = 0.057−0.070) under both models. R^2 values were relatively weak, however, and ranged from 0.017 to 0.028 (Table 4).

When all SNP effects were examined simultaneously in the same model using a multi-SNP analysis approach, the analyses showed that under the additive model, the significant associations with body weight were only supported for two SNPs (GP, P = 0.023 and HSP90, P = 0.003). Under the dominance model, apart from three SNPs already identified from individual SNP analyses (ANK-like protein_A1 [P = 0.009], ANK-like protein_T_A2 [P = 0.004], HSP90 [P = 0.001]), two new SNPs in RP (P = 0.010) and UGP2 (P = 0.013) genes also showed significant dominant associations with body weight.

When the multi-SNP regression analysis with a stepwise selection method was performed on all SNPs, the results suggested five SNPs (an exonic SNP in HSP90 gene [P = 0.002], ANK-like protein_A1 gene [P = 0.010], GP gene [P = 0.028], peroxidasin gene [P = 0.031] and TGF-βIP_A1 gene [P = 0.065]) had significant additive genetic effects on body

weight. Together they explained 12.2 % of genetic variance for body weight. There were no significant marker interactions identified among the five markers.

DISCUSSION

Candidate gene markers

In the current study we identified and validated 100 SNP markers within putative candidate genes, and examined the associations between 27 selected SNPs and individual growth performance in an improved GFP culture line. A 75 % PCR success rate was similar to that reported for Pacific white shrimp (79 %), *Litopenaeus vannamei* (Gorbach *et al*. 2009) but less than that previously reported for the same species (87 %) (Ciobanu *et al*. 2010). PCR failures mainly resulted from unexpected intron insertions. Whole genome sequencing projects offer the prospect of developing large numbers of null-free Type I loci, but it is still difficult to obtain them for most aquatic invertebrates (Audzijonyte & Vrijenhoek 2010; Kim *et al*. 2011) because of time and labour costs. In the current study of 47 loci, we observed an average of one SNP per 40 bp sequenced in the sampled gene regions in *M. rosenbergii*, and this frequency is similar to that reported in other aquatic invertebrate species including; oysters (one SNP per 20−60 bp), nematode (one per 20 bp) and ascidian (one per 20 bp) (Cutter *et al.* 2006; Sauvage *et al.* 2007; Small *et al.* 2007; Zhang & Guo, 2010; Kim *et al.* 2011). This SNP frequency was high, however, compared with the SNP frequency (one per 285 bp) reported for an improved hatchery line of Pacific white shrimp (G7 to G9) (Ciobanu *et al.* 2010). This difference may in part, be explained by 1) different sample sources including the geographical panels used (hatchery *vs* wild), 2) different location of primers in the genome and 3) different sequencing approaches (454 pyrosequencing *vs* Sanger sequencer). In the case of SNP heterozygote diversity, G2 individuals studied here already showed a sign of reduced frequency compared with wild samples despite being exposed to only a short period of selection. Abundant SNPs and indels as observed in the *M. rosenbergii* populations could provide useful information for a number of potential genetic applications, but can also be a significant obstacle to marker development and selection of individual SNPs as targets for high-throughput screening. In particular, the k-mer approach and homopolymer errors in 454 *de novo* assembly could potentially identify false putative SNPs. Applying third-generation sequencing technologies that can generate longer average read length could

be a potential solution to address these issues but its wider application requires further technology maturation.

Association studies

Among the 27 SNPs genotyped successfully here using SEQUENOM, both individual and multi-SNP analyses identified two exonic SNPs in HSP90 and GP, and an intronic SNP in ANK-like protein_A1 that showed significant additive genetic effects on body weight. Among them, HGP90 explained the most genetic variation for body weight (4.7 %). In addition both individual and multi-SNP analyses also showed that SNP markers in ANK-like A1, T_A2, HSP90, RP, and UGP2 had significant dominant effects on body weight. Since the success of a genetic improvement program applying marker-assisted selection largely relies on additive effects of genetic markers, the markers with both favourable additive and dominant effects have potential to contribute most to genetic gains in a selection program.

The genes identified here represent clear candidates for possessing roles in growth in *M. rosenbergii*. All findings in the current study represent new discovery of direct and/or indirect evidence of these candidate genes related to growth in crustacean species. For example, ankyrin repeats (ANK) are necessary for providing effective protein-protein interactions (Rubtsov & Lopina, 2000; Mosavi *et al*. 2004; Barrick *et al*. 2008). In pig, polymorphisms in ANK-like protein showed significant associations with growth and carcass traits (porcine loin depth and meat firmness) (Sun *et al*. 2011), and in crustaceans they have been implicated in moulting (Chithr & Devaraj 2012). While an ANK gene was reported to be involved in early developmental regulation, especially brachyurization regulation in Chinese mitten crab (Li *et al*. 2009), the current study has provided the first direct evidence for significant association with growth in a crustacean. HSPs protect endogenous proteins during stressful conditions, preventing their irreversible aggregation and assist in protein refolding (Feder & Hofmann 1999; Hartl & Hayer-Hartl 2002; Liberek *et al*. 2008). In crustaceans, most studies of HSP forms have been conducted to assess relationships with innate immune responses and/or environmental stress, but some studies have provided substantial evidence for their involvement in crustacean muscle atrophy over the moult cycle (muscle degradation) in American lobster (Spees *et al*. 2003) and Pacific white shrimp (Cesar & Yang 2007). The first direct evidence for HSPs related to growth in crustacean suggests that putative functional role of HSPs should not be limited to only innate immune responses

and/or environmental stress but must be expanded to include moulting and growth. TGF-βIP (high homology with Fasciclin) is believed to be involved in mediating cell growth and differentiation, and also inhibits epidermal growth factor receptor signalling during insect development (Mao & Freeman 2008). The first confirmed evidence for an association between TGF-βIP with growth in a crustacean was consistent with previous studies in livestock, where this gene was suggested to be associated significantly with some production traits (Bennett *et al*. 2007; Chen *et al*. 2012a).

Glycogen phosporylase (GP) is an important regulator of muscle glycogen levels (Greenberg *et al*. 2006; Irimia *et al*. 2012). A recent study suggests that GP may also play a similar role in crustaceans by meeting heightened energy demand and metabolite state during environmental fluctuations (Oliveira *et al*. 2004; Bacca *et al*. 2005), and may influence glycogen accumulation in specific tissues, including during moulting (Reddy *et al*. 2007; Nagai *et al*. 2011; Babu *et al*. 2012). Discovery of a new GP identified here provides a starting point for exploring its physiological role(s) in crustaceans.

Peroxidasin (PXDN), closely related to chordate peroxidase and peroxinectin, has not been identified in crustaceans previously but has potential roles in antimicrobial defence, extracellular matrix formation and/or extracellular inhibitors of axonal regeneration, and consolidation at various developmental stages (Tindall *et al*. 2005; Gotenstein *et al*. 2010; Soudi *et al*. 2012). Peroxinectin deduced peptide has previously been characterised in crustaceans during innate immune defence experiments, and has as an overall similarity with PXDN (Sritunyalucksana *et al*. 2001; Hsu *et al*. 2005; Dong *et al*. 2009). While results of the current study suggest the first direct evidence for a role in growth, the exact physiological role(s) of PXDN in crustacean species remain to be resolved. UTP-glucose-1-phosphate uridylyltransferase (UGP), also known as UDP-glucose pyrophosphorylase (UDPGP), is a ubiquitous enzyme in nature, and an obvious gene candidate for growth traits given its central roles in carbohydrate interconversion and glycogen synthesis (Thoden & Holden 2006). Rolling pebbles (RP) or Rols isoform are recognised as muscle-specific founders. Recent studies suggest it has molecular interactions during myogenesis and may regulate cell fusion, in the tubules that lead to cell intercalation during organ formation (Denholm & Skaer 2009), myotube enlargement (Menon *et al*. 2005) and/or myoblast fusion (Avirneni-Vadlamudi *et al*. 2012). While results from the current study represent new findings for GP, UGP and RP genes, considering their known physiological roles, significant associations with crustacean growth is not surprising. Despite requiring further validation, markers identified in candidate

genes here will be informative not only for pinpointing signals in biological pathways mapping (i.e. growth) in the future via molecular networks, but also potentially could be useful in GFP breeding because of their important physiological roles. As it currently stands, this study represents the first direct evidence for genetic variation in the genes identified here being associated with growth in crustacean species.

Since there is neither whole genome sequence information nor gene locations on chromosomes available for GFP, it is impossible to locate the SNPs examined here to specific GFP chromosomes. While it is possible that one or more of the SNPs identified here could be potentially located on sex chromosomes because only females were studied here, it is not possible to determine their chromosome locations. However, evidence of no interactions among significant SNPs with additive effects suggests that the SNPs do not occur on the same chromosome(s) and therefore, can be selected independently or jointly. Despite this fact, any SNPs associated with individual growth performance potentially can be applied in future GFP breeding programs and used to increase the efficiency of the selection process in this species. Furthermore, the markers can contribute to future linkage mapping, population genetic studies, and phylogenetic analyses of wild and cultured GFP genetic resources.

Although *M. rosenbergii* culture has been practiced since the 1970's, information on response to selection and/or the heritability of growth-related traits, defined as the proportion of additive genetic variance over the total phenotypic variance, are very rare because stock improvement of GFP culture lines has only been trialled recently (Thanh *et al*. 2009; Aflalo *et al*. 2012; Dinh *et al*. 2012). In addition, most research has focused on traditional quantitative genetics studies of genetic parameters (heritabilities, genetic correlations) of phenotypic traits and selection response per generation (Aflalo *et al*. 2012; Dinh *et al*. 2013; Kitcharoen *et al*. 2012; Luan *et al*. 2012). In a recent study of *Fenneropenaeus indicus*, however, the proportion of genetic variance (R^2) for growth was estimated to be 0.07 based on random amplified polymorphic DNA (RAPD) markers (Rezvani Gilkolaei *et al*. 2011). In addition, R^2 was estimated at 0.02−0.30 for body mass, 0.02−0.25 for standard length and 0.03−0.36 for body height using microsatellite markers in yellow croaker *Larimichthys crocea*, (Liu *et al*. 2012). R^2 estimation identified for significant candidate loci in the current study (Type I EST markers) ranged from 0.026−0.048, a level that is similar or lower than previously reported for other Type II markers (Rezvani Gilkolaei *et al*. 2011; Liu *et al*. 2012).

Candidate gene association analyses have been successful approaches for identifying the genetic basis of phenotypic correlations between hatch timing, shape, skin, weight, length and growth rate in some aquaculture species including; Arctic charr (Tao & Boulding 2003), Kuruma prawns (Lyons *et al*. 2007), Atlantic salmon (Boulding *et al*. 2008), coho salmon (McClelland & Naish 2010), Pacific white shrimp (Ciobanu *et al*. 2010), and GFP (Thanh *et al*. 2010b). Exonic SNPs from ESTs have been the main target of association studies in the past because SNPs are likely to affect the physiological functions of target genes by altering amino acids present in resulting proteins (i.e non-synonymous mutations) (Genissel *et al*. 2004). Intronic SNPs have been suggested to be potential regulatory elements that can affect splicing, thereby resulting in phenotypic variation, and they have also been implicated in producing phenotypic variation including in production traits in aquatic species (Tao & Boulding 2003; Thanh *et al*. 2010b) as well as genetic disorders in humans (Pagani & Barralle 2004; Hastings *et al*. 2005; Hull *et al*. 2007). Since most conserved motifs near exon-intron boundaries act as critical splice signal sites, a number of intronic SNPs have been suggested to be involved in regulating normal splicing of exonic sequences (Pagani & Baralle 2004). A few intronic SNPs identified at exon-intron boundaries in the current study (< 50bp from exon) were associated with genetic values for body weight, including ANK-like protein A1, T_A2 and TGF- βIP_T_A2. Therefore, intronic SNPs are likely to be an important component of future association studies in GFP due to their relative abundance compared with exonic SNPs (Liu 2007; De-Santis & Jerry 2007; Jung *et al*. 2013). A recent study has suggested that both synonymous and non-synonymous SNPs in exonic regions can be equally useful for genetic applications (i.e. mapping and association studies) because synonymous SNPs also can contribute to phenotypic variation (Kimchi-Sarfaty *et al*. 2007). Synonymous polymorphisms in HSP90 and GP showed significant correlations with EBVs in the current study but require further validation to provide robust evidence for a direct role in growth in GFP.

It is worthwhile to mention that the SNP effects identified here were estimated using the polygenic component of genetic variance for body weight (i.e. EBVs) in extreme growth phenotypes for 200 adult GFP females. Therefore R^2 was in fact the proportion of additive genetic variance explained. The estimates could have been different if the residuals for body weight from the mixed animal model in the large experiment (Dinh *et al*. 2012, 2013) had been used for the association study, and the variance from the residuals of an animal model would explain all other non-polygenetic genetic variation in the population. The majority of

SNPs detected in the current study did not deviate ($P > 0.05$) from HWE, but ANK-like protein_A1, GluCl_A1, TGF-βIP_A2 and Zasp52 indicated a potential for selection pressure on growth performance between slow and fast groups even though this only involved effects over two generations (only G2). Relatively intense selection for growth rate was used in the Dinh *et al*. study (2012, 2013), so detection of significant deviations from HWE for individual SNPs may provide a marker for specific sites in candidate genes where selection has had a major effect. While several SNP markers in the current study were in significant LD (FAP, FAMeT, TGF-βIP, GluCl, ING, PHLCFM2, RP, and UGP2) ($P < 0.001$), the multi-SNP trait analysis did not identify any significant linkage among these SNPs and body weight. While this shows that much larger validation populations would be required to accurately study the associations in the future, this is the first large set of genetic markers reported for *M. rosenbergii* and will be useful for confirmation of associations in other samples or lines.

Evolutionary pressures

Evolutionary response is generally determined by a combination of phenotypic variation, the strength and direction of selection, the underlying genetic variation within a population and the genetic covariance between traits (Falconer & Mackay 1996). Estimates of allele frequencies, non-synonymous (Ka)/synonymous (Ks) ratio and transition (Ti)/tranversion (Tv) ratio have been used widely to understand selection pressure during molecular evolution and to monitor long-term effective population size. These findings, however, will require further validation in diverse test populations (wild and several breeding generation stocks) because the current data were observed under growth selection pressure using only EBVs from G2 adult females (body weight). Therefore, future studies should consider different forms of selection driven by reproduction and at different developmental stages because of potential differential impacts on epigenetic, epistatic, pleiotropic factors among genes which often act unevenly among genes and genomic regions (McClelland & Naish 2010; Loukovitis *et al*. 2011; Chen *et al*. 2012b; Gruber *et al*. 2012).

Identification of combinations of functional molecular polymorphisms that differ in frequency among populations that utilise different environments or that possess different phenotypic ranges could facilitate MAS in the target species. Little is known, however, about natural genetic diversity in *M. rosenbergii*. To date, only a few microsatellite markers have

been utilised to evaluate genetic diversity in wild and cultured strains (Hurwood *et al*. 2012; Schneider *et al*. 2012) in *M. rosenbergii*. Genetic diversity is a critical resource in animal breeding because highly inbred populations are likely to respond much more slowly or potentially not at all to stock improvement (Ciobanu *et al*. 2010). Significant differences in polymorphism levels at marker sites between fast and slow growth lines can potentially be used as tags for monitoring stock productivity over time (i.e. genetic diversity, inbreeding and genetic drift). In addition, it would be interesting to identify whether the sources of large differences in allelic frequency are signatures of selection acting on a particular trait or simply random effects of genetic drift. A few studies conducted in shrimps and prawns have reported that some negatively correlated traits (pathogen resistance and growth) may have resulted from selection of certain allelic variants acting in opposing ways (Gitterle *et al*. 2005; Ciobanu *et al*. 2010; Panphut *et al*. 2011) that result from possible mutations and drift forces. The markers (i.e. EST-SNP and EPIC) identified here could be useful in future analyses of other complex traits, for tracing pedigree relationships in breeding programs, detection of inter-/intra-specific variation among congeners, and for monitoring genetic diversity in commercial and wild populations of *M. rosenbergii*.

Conclusions

A major result of the current study was identification of polymorphisms in ANK-like protein, GP, HSP90, Peroxidasin, RP, TGF-βIP and UGP2 genes that showed associations with individual body weight in female GFP. While the candidate gene SNP markers need to be confirmed using different breeding lines to overcome two criticisms of CGA (lack of replication and thoroughness), SNPs associated with growth performance have potential applications in MAS/GS programs not only for developing fast-growth GFP lines but also to identify brood-stock that possess the best genetic potential for fast growth.

ACKNOWLEDGEMENTS

Authors wish to thank to Dr. Woo-Jin Kim, Dr. Peter Prentis and two anonymous reviewers for useful comments. This research forms a part of Hyungtaek Jung's PhD project, and is supported by an International Postgraduate Research Scholarship (Australia) and a

Queensland University of Technology Postgraduate Award (N7333978). Additional funding for this work was provided by Queensland University of Technology awarded to Peter B. Mather (VC NPSG award).

ELECTRONIC SUPPLEMENTARY MATERIAL

E1. Alignment of the sequenced 47 loci amplicons (454 pyrosequencing and ABI re-sequencing) – Fasta file format

E2. The number of polymorphic sites detected in the sequenced 47 loci amplicons and their annotation results (455 pyrosequencing and ABI re-sequencing) – Excel file format

REFERENCES

Aflalo ED, Raju DVSN, Bommi NA, Verghese JT, Samraj TYC, Hulata G *et al*. (2012) Toward a sustainable production of genetically improved all-male prawn (*Macrobrachium rosenbergii*): Evaluation of production traits and obtaining neo-females in three Indian strains. *Aquaculture* **338-341**: 197–207.

Arendt JD (1997) Adaptive intrinsic growth rates: an integration across taxa. *Quarterly Review of Biology* **72**: 149–177.

Audzijonyte A, Vrijenhoek RC (2010) Three nuclear genes for phylogenetic, SNP and population genetic studies of molluscs and other invertebrates. *Molecular Ecology Resources* **10**: 200–204.

Avirneni-Vadlamudi U, Galindo KA, Endicott TR, Paulson V, Cameron S, Galindo RL (2012) *Drosophila* and mammalian models uncover a role for the myoblast fusion gene TANC1 in rhabdomyosarcoma. *Journal of Clinical Investigation* **122**: 403–407.

Babu KN, Pallavi PN, Reedy DC, Kalarani V (2012) Effect of 5-hydroxytryptamine and dopamine on the carbohydrate metabolism in the shrimp, *Penaeus monodon* (*Fabricius*). *World Journal of Fish and Marine Sciences* **4**: 586–593.

Bacca H, Huvet A, Fabioux C, Daniel JY, Delaporte M, Pouvreau S *et al*. (2005) Molecular cloning and seasonal expression of oyster glycogen phosphorylase and glycogen synthase genes. *Comparative Biochemistry and Physiology* **140B**: 635–646.

Barrick D, Ferreiro DU, Komives EA (2008) Folding landscapes of ankyrin repeat proteins: experiments meet theory. *Current Opinion in Structural Biology* **18**: 27–34.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B* **57**: 289–300.

Bennett AK, Hester PY, Spurlock DM (2007) Relationships of a transforming growth factor-β2 single nucleotide polymorphism and messenger ribonucleic acid abundance with bone and production traits in chickens. *Poultry Science* **86**: 829–834.

Bouck A, Vision T (2007) The molecular ecologist's guide to expressed sequence tags. *Molecular Ecology* **16**: 907–924.

Boulding EG, Culling M, Glebe B, Berg PR, Lien S, Moen T (2008). Conservation genomics of Atlantic salmon: SNPs associated with QTLs for adaptive traits in parr from four trans-Atlantic backcrosses. *Heredity* **101**: 381-391.

Cesar JRO, Yang J (2007) Expression patterns of ubiquitin, heat shock protein70, α-actin and β-actin over the molt cycle in the abdominal muscle of marine shrimp *Litopenaeus vannamei*. *Molecular Reproduction and Development* **74**: 554-559.

Chen K, Hawken R, Flickinger GH, Rodriguez-Zas SL, Rund LA, Wheeler MB *et al*. (2012a) Association of the porcine transforming growth factor beta type I receptor (TGFBR1) gene with growth and carcass traits. *Animal Biotechnology* **23**: 43–63.

Chen R, Irwin DM, Zhang YP (2012b) Differences in selection drive olfactory receptor genes in different direction in dogs and wolf. *Molecular Biology and Evolution* **Online version**.

Chithra C, Devaraj H (2012) Purification and characterisation of a 48 kDa protein involved in the moulting cycle of *Penaeus monodon*. *Food Chemistry* **134**: 15–20.

Ciobanu DC, Bastiaansen JWM, Magrin J, Jiang DH, Yu N, Geiger B *et al*. (2010) A major SNP resource for dissection of phenotypic and genetic variation in Pacific white shrimp (*Litopenaeus vannamei*). *Animal Genetics* **41**: 39–47.

Cutter AD, Baird SE, Charlesworth D (2006) High nucleotide polymorphism and rapid decay of linkage disequilibrium in wild populations of *Caenorhabditis remanei*. *Genetics* **174**: 901–913.

Denholm B, Skaer H (2009) Bringing together components of the fly renal system. *Current Opinion in Genetics and Development* **19**: 526–532.

De-Santis C, Jerry DR (2007) Candidate growth genes in finfish: Where should we be looking? *Aquaculture* **272**: 22–38.

Dinh H, Coman G, Hurwood DA, Mather PB (2012) Experimental assessment of the utility of visible implant elastomer tags in a stock improvement programme for giant freshwater prawn (*Macrobrachium rosenbergii*) in Vietnam. *Aquaculture Research* **43**: 1471–1479.

Dinh H, Vu NT, Nguyen NH, Ponzoni RW, Hurwood DA, Mather PM (2013) Quantitative genetic parameter estimates for body and carcass traits in a cultured stock of giant freshwater prawn (*Macrobrachium rosenbergii*) selected for harvest weight in Vietnam. *Aquaculture* **404-405**: 122–129.

Dong B, Liu F, Gao H, Wang B, Xiang J (2009) cDNA cloning and gene expression pattern following bacterial challenge of peroxinectin in Chinese shrimp *Fenneropenaeus chinensis*. *Molecular Biology Reports* **36**: 2333–2339.

Du ZQ, Ciobanu DC, Onteru SK, Gorbach D, Mileham AJ, Jaramillo G *et al*. (2010) A gene-based SNP linkage map for pacific white shrimp, *Litopenaeus vannamei*. *Animal Genetics* **41**: 286–294.

Ekblom R, Galindo J (2010) Application of next generation sequencing in molecular ecology of non-model organisms. *Heredity* **107**: 1–15.

Falconer DS, Mackay TFC (1996) Introduction to Quantitative Genetics. Prentice Hall: San Francisco, CA

FAO (2010) Fisheries Statistical Database, Global Aquaculture Production (Fisheries Global Information System, online query).

Feder ME, Hofmann GE (1999) Heat−shock proteins, molecular chaperones, and the stress response: evolutionary and ecological physiology. *Annual Review of Physiology* **61**: 243–282.

Genissel A, Pastinen T, Dowell A, Mackay TFC, Long AD (2004) No evidence for an association between common nonsynonymous polymorphisms in Delta and bristle number variation in natural and laboratory populations of *Drosophila melanogaster*. *Genetics* **166**: 291–306.

Gilmour AR, Gogel BJ, Cullis BR, Thompson R (2009) ASReml user guide release 3.0. VSN International Ltd., Hemel Hampstead, HP11ES UK.

Gitterle T, Salte R, Gjerde B, Cock J, Johansen H, Salazar M *et al*. (2005) Genetic (co)variation in resistance to White Spot Syndrome Virus (WSSV) and harvest weight in *Penaeus* (*Litopenaeus*) *vannamei*. *Aquaculture* **246**: 139–149.

Goddard ME, Hayes BJ (2009) Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Review Genetics* **10**: 381–391.

Gorbach DM, Hu ZL, Du ZQ, Rothschild MF (2009) SNP discovery in *Litopenaeus vannamei* with a new computational pipeline. *Animal Genetics* **40**: 106–109.

Gotenstein JR, Swale RE, Fukuda T, Wu Z, Giurumescu CA, Goncharov A *et al*. (2010) The *C. elegans* peroxidasin PXN-2 is essential for embryonic morphogenesis and inhibits adult axon regeneration. *Development* **137**: 3603–3613.

Greenberg CC, Jurczak MJ, Danos AM, Brady MJ (2006) Glycogen branches out: new perspectives on the role of glycogen metabolism in the integration of metabolic pathways. *American Journal of Physiology Endocrinology and Metabolism* **291**: E1–E8.

Gruber JD, Vogel K, Kalay G, Wittkopp PJ (2012) Contrasting properties of gene-specific regulatory, coding, and copy number mutations in Saccharomyces cerevisiae: Frequency, effects, and dominance. *PLoS Genetics* **8**: e1002497.

Hartl FU, Hayer-Hartl M (2002) Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science* **295**: 1852–1858.

Hastings ML, Resta N, Traum D, Stella A, Guanti G, Krainer AR, (2005) An LKB1 ATAC intron mutation causes Peutz–Jeghers syndrome via splicing at noncanonical cryptic splice sites. *Nature Structural and Molecular Biology* **12**: 54–59.

Hayter AJ (1984) A proof of the conjecture that the Tukey-Kramer multiple comparisons procedure is conservative. *Annals of Statics* **12**: 61–75.

Hsu PI, Liu CH, Tseng DY, Lee PP, Cheng W (2006) Molecular cloning and characterisation of peroxinectin, a cell adhesion molecule, from the giant freshwater prawn *Macrobrachium rosenbergii*. *Fish and Shellfish Immunology* **21**: 1–10.

Hu X, Gao Y, Feng C, Liu Q, Wang X, Du Z *et al*. (2009) Advanced technologies for genomic analysis in farm animals and its application for QTL mapping. *Genetica* **136**: 371–386.

Hull J, Campino S, Rowlands K, Chan MS, Copley RR, Taylor MS *et al*. (2007) Identification of common genetic variation that modulates alternative splicing. *PLoS Genetics* **3**: 1009–1018.

Hurwood DA, Singh S, Dammannagoda ST, Nandlal S, Mather PB (2012) Experimental evaluation of the culture performance of three exotic and a 'local' giant freshwater prawn (*Macrobrachium rosenbergii*, De Man 1879) culture strain in Fiji. *Aquaculture Research* **Online Version**.

Irimia JM, Rovira J, Nielsen JN, Guerrero M, Wojtaszewski JFP, Cussó R (2012) Hexokinase 2, glycogen synthase and phosphorylase play a key role in muscle glycogen supercompensation. *PLoS One* **7**: e42453.

Jung H, Lyons RE, Dinh H, Hurwood DA, McWilliam S, Mather PB (2011) Transriptomics of giant freshwater prawn (*Macrobrachium rosenbergii*): *De novo* assembly, annotation and marker discovery. *PLoS One* **6**: e27938.

Jung H, Lyons RE, Hurwood DA, Mather PM (2012) Genes and growth performance in crustacean species: A review of relevant genomic studies in crustaceans and other taxa. *Reviews in Aquaculture* **In Press**.

Kim WJ, Jung HT, Gaffney PM (2011) Development of Type I genetic markers from expressed sequence tags in highly polymorphic species. *Marine Biotechnology* **13**: 127–132.

Kimchi-Sarfaty C, Oh JM, Kim IW, Sauna ZE, Calcagno AM, Ambudkar SV *et al*. (2007) A "silent" polymorphism in the MDR1 gene changes substrate specificity. *Science* **315**: 525–528.

Kitcharoen N, Rungsin W, Koonawootrittriron S, Na-Nakorn U (2012) Heritability for growth traits in giant freshwater prawn, *Macrobrachium rosenbergii* (de Mann 1879) based on best linear unbiased prediction methodology. *Aquaculture Research* **43**: 19–25.

Kruglyak L, Nickerson DA (2001). Variation is the spice of life. *Nature Genetics* **27**: 234-236.

Li P, Zha J, Huang H, Zhang Z, Sun H, Song D *et al*. (2009) Identification, mRNA expression and characterization of a novel ANK-like gene from Chinese mitten crab *Eriocheir japonica sinensis*. *Comparative Biochemistry and Physiology* **153B**: 332–339.

Liberek K, Lewandowska A, Ziętkiewicz S (2008) Chaperones in control of protein disaggregation. *EMBO Journal* **27**: 328–335.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**: 1451–1452.

Liu Z (2007) Single nucleotide polymorphism (SNP). In: Liu, Z. (Ed.), Aquaculture Genome Technologies. Blackwell, USA, pp. 59–72.

Liu XD, Zhao GT, Cai MY, Wang ZY (2012) Estimated genetic parameters for growth-related traits in large yellow croaker *Larimichthys crocea* using microsatellites to assign parentage. *Journal of Fish Biology* **Online Version**.

Loukovitis D, Sarropoulou E, Tsigenopoulos CS, Batargias C, Magoulas A, Apostolidis AP *et al*. (2011) Quantitative trait loci involved sex determination and body growth in the gilthead sea bream (*Sparus aurata* L.) through targeted genome scan. *PLoS One* **6**: e16599.

Luan S, Yang G, Wang J, Luo K, Zhang Y, Gao Q *et al*. (2012) Genetic parameters and response to selection for harvest body weight of the giant freshwater prawn *Macrobrachium rosenbergii*. *Aquaculture* **362-363**: 88–96.

Lynch M, Walsh B (1998) Genetics and Analysis of Quantitative Traits. Sinauer Assoc Inc., Sunderland, MA, USA.

Lyons RE, Dierens LM, Tan SH, Preston NP, Li Y (2007) Characterization of AFLP markers associated with growth in the Kuruma prawn, *Marsupenaeus japonicas*, and identification of a candidate gene. *Marine Biotechnology* **9**: 712–721.

Mackay TFC, Stone EA, Ayroles JF (2009) The genetics of quantitative traits: challenges and prospects. *Nature Reviews Genetics* **10**: 565–577.

Mangel M, Stamps J (2001) Trade-offs between growth and mortality and the maintenance of individual variation in growth. *Evolutionary Ecology Research* **3**: 583–593.

Mao Y, Freeman M (2008) Fasciclin 2, the Drosophila orthologue of neural cell adhesion molecule, inhibits EGF receptor signalling. *Development* **136**: 473–481.

McClelland EK, Naish KA (2010) Quantitative trait locus analysis of hatch timing, weight, length and growth rate in coho salmon, *Oncorhynchus kisutch*. *Heredity* **105**: 562–573.

Menon SD, Osman Z, Chenchill K, Chia W (2005) A positive feedback loop between Dumbfounded and Rolling pebbles leads to myotube enlargement in *Drosophila*. *Journal of Cell Biology* **169**: 909–920.

Metzker M (2010) Sequencing technologies: the next generation. *Nature Reviews Genetics* **11**: 31-46.

Miller SA, Dykes DD, Polesky HF (1998) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Research* **16**: 1215.

Morin PA, Luikart G, Wayne RK, Grp SW (2004). SNPs in ecology, evolution and conservation. *Trends in Ecology and Evolution* **19**: 208-216.

Mosavi LK, Cammett TJ, Desrosiers DC, Peng ZY (2004) The ankyrin repeat as molecular architecture for protein recognition. *Protein Science* **13**: 1435–1448.

Nagai C, Nagata S, Nagasawa H (2011) Effects of crustacean hyperglycaemic hormone (CHH) on the transcript expression of carbohydrate metabolism-related enzyme genes in the kuruma prawn, *Marsupenaeus japonicas*. *General and Comparative Endocrinology* **172**: 293–304.

Oliveira G T, Rossi IC, Kucharski LC, Da Silva RS (2004) Hepatopancreas gluconeogenesis and glycogen content during fasting in crabs previously maintained on a high-protein or carbohydrate-rich diet. *Comparative Biochemistry and Physiology* **137A**: 383–390.

Pagani F, Baralle FE (2004) Genomic variants in exons and introns: indentifying the splicing spoilers. *Nature Reviews Genetics* **5**: 389–396.

Panphut W, Senapin S, Sriurairatana S, Withyachumanarnkul B, Flegel TW (2011) A novel integrase-containing element may interact with Laem-Singh virus (LSNV) to cause slow growth in giant tiger shrimp. *BMC Veterinary Research* **7**: 18.

Pillai BR, Mahapatra KD, Ponzoni RW, Sahoo L, Lalrinsanga PL, Nguyen NH *et al*. (2011) Genetic evaluation of a complete diallel cross involving three populations of freshwater prawn (*Macrobrachium rosenbergii*) from different geographical regions of India. *Aquaculture* **319**: 347–354.

Rafter JA, Abell ML, Braselton JP (2002) Multiple comparison methods for means. *Society of Industrial and Applied Mathematics* **44**: 259–279.

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86**: 248–249.

Reddy PS, Reddy PR, Pushpalatha T (2007) Differential mode of action of hyperglycaemic hormones obtained from crab (Crustacea) and millipede (Myriapoda) in the crab, *Oziotelphusa senex senex*. *Journal of Biological Sciences* **7**: 141–152.

Rezvani Gilkolaei S, Safari R, Laloei F, Taqavi J, Matinfar A (2011) Using RAPD markers potential to identify heritability for growth in *Fenneropenaeus indicus*. *Iranian Journal of Fisheries Sciences* **10**: 123–134.

Rice WR (1989) Analyzing tables of statistical tests. *Evolution* **43**: 223–225.

Rousset F (2008) Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Molecular Ecology Resources* **8**: 103–106.

Rubtsov AM, Lopina OD (2000) Ankyrins. *FEBS Letters* **482**: 1–5.

SAS Institute (2002). SAS OnlineDoc, Version 8.2. SAS Institute Inc.: Cary, NC, USA.

Sauvage C, Bierne N, Lapègue S, Boudry P (2007) Single nucleotide polymorphisms and their relationship to codon usage bias in the Pacific oyster *Crassostrea gigas*. *Gene* **406**: 13–22.

Schneider KJ, Tidwell JH, Gomelsky B, Pomper KW, Waldbieser GC, Saillant E *et al*. (2012) Genetic diversity of cultured and wild populations of the giant freshwater prawn *Macrobrachium rosenbergii* (de Man, 1879) based on microsatellite analysis. *Aquaculture Research* **Online Version**.

Schwantes VS, Diana JS, Yi Y (2009) Social, economic, and production characteristics of giant freshwater prawn *Macrobrachium rosenbergii* culture in Thailand. *Aquaculture* **287**: 120–127.

Seeb JE, Carvalho G, Hauser L, Naish K, Roberts S, Seeb LW (2011) Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. *Molecular Ecology Resources* **11S1**: 1–8.

Small KS, Brudno M, Hill MM, Sidow A (2007) Extreme genomic variation in a natural population. *Proceedings of the National Academy of Sciences of the USA* **104**: 5698–5703.

Soudi M, Zamocky M, Jakopitsch C, Furtmüller PG, Obinger C (2012) Molecular evolution, structure, and function of peroxidasins. *Chemistry and Biodiversity* **9**: 1776–1793.

Spees JL, Chang SA, Mykles DL, Snyder MJ, Chang ES (2003) Molt cycle-dependent molecular chaperone and polyubiquitin gene expression in lobster. *Cell Stress and Chaperones* **3**: 258–264.

Sritunyalucksana K, Wongsuebsantati K, Johansson MW, Söderhäll (2001) Peroxinectin, a cell adhesive protein associated with the proPO system from the black tiger shrimp, *Penaeus monodon*. *Developmental and Comparative Immunology* **25**: 353–363.

Stephens M, Donnelly P (2003) A comparison of Bayesian methods for haplotype reconstruction. *American Journal of Human Genetics* **73**: 1162–1169.

Sun L, Dong X, Fan B, Liu B (2011) The association of ANKRD2 with loin depth and muscle firmness in pigs. *Journal of Animal Veterinary Advances* **10**: 1462–1468.

Sundstrom LF, Lohmus M, Devlin RH (2005) Selection on increased intrinsic growth rates in coho salmon, *Oncorhynchus kisutch*. *Evolution* **59**: 1560–1569.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* **28:** 2731–2739.

Tangprasittipap A, Tiensuwan M, Withyachumnarnkul B (2010). Characterization of candidate genes involved in growth of black tiger shrimp *Penaeus monodon*. *Aquaculture* **307**: 150-156.

Tao W, Boulding EG (2003). Associations between single nucleotide polymorphisms in candidate genes and growth rate in Arctic charr (*Salvelinus alpines* L.). *Heredity* **91**: 60-69.

Thanh NM, Ponzoni Rw, Nguyen NH, Vu NT, Barnes A, Mather PB (2009) Evaluation of growth performance in a diallel cross of three strains of giant freshwater prawn (*Macrobrachium rosenbergii*) in Vietnam. *Aquaculture* **287**: 75–83.

Thanh NM, Nguyen NH, Ponzoni RW, Vu NT, Barnes AC, Mather PB (2010a). Estimates of strain additive and non-additive genetic effects for growth traits in a diallel cross of three strains of giant freshwater prawn (*Macrobrachium rosenbergii*) in Vietnam. *Aquaculture* **299**: 30–36.

Thanh NM, Barnes AC, Mather PB, Yutao L, Lyons RE (2010b). Single nucleotide polymorphisms in the actin and crustacean hyperglycemic hormone genes and their correlation with individual growth performance in giant freshwater prawn *Macrobrachium rosenbergii*. *Aquaculture* **301**: 7–15.

Thoden JB, Holden HM (2006) The molecular architecture of glucose-1-phosphate uridylyltransferase. *Protein Science* **16**: 432–440.

Tindall AJ, Elizabeth M, Pownall E, Morris ID, Issacs HV (2005) *Xenopus tropicalis peroxidasin* gene is expressed within the developing neural tube and pronephric kidney. *Developmental Dynamics* **232**: 377–384.

Verhoeven KJ, Simonsen KL, McIntyre LM (2005) Implementing false discovery rate control: increasing your power. *Oikos* **108**: 643–647.

Wang S, Peatman E, Abernathy J, Waldbieser G, Lindquist E, Richardson P *et al*. (2010) Assembly of 500,000 inter-specific catfish expressed sequence tags and large scale gene-associated marker development for whole genome association studies. *Genome Biology* **11**: R8.

Zhang L, Guo X (2010) Development and validation of single nucleotide polymorphism markers in the eastern oyster *Crassostrea virginica* Gmelin by mining ESTs and resequencing. *Aquaculture* **302**: 124–129.

TABLES and FIGURES

Table 1. Summary of SNPs identified in *M. rosenbergii* amplicons from 454 transcriptome and ABI re-sequencing data.

Locus [a]	Amplicon length (bp)	Primers (5' → 3')	Ta (°C)[b]	Amplicon content [c]	Intron features [d]	π[e]	# of Ka/Ks[f]	# of Ti/Tv[g]	GenBank reference[h]	E-value[i] (BLASTx)	Similarity [i] (%)
Actin 1	158	GTCCTATGAACTTCCCGATGGTC TCCTGATATCAATGTCGCACCTC	59	E		0.02363	6/8	11/3	AF100986	0	99
Actin 6	174	CTTTGCTACATTGCCCTGGACTAC ACCGGCAGATTCCATACCAAG	62	E		0.02057	2/7	5/4	JN006040	2.4E-146	98
ADP/ATP translocase	277	ATCTCGGGCATCATCTCGTA GGAGGATAGAGGGAGGGAGTTACT	62	E		0.00781	3/4	3/4	DQ874397	0	95
Adrift-like (FTSJ) [a]	252	CATGCTGGAAGTACTTTGTCTGA TTCCAGCCAGTTCTCCCTAA	61	E		0.00587	2/2	5/0	GJ062549 XM001506876	4.83E-94	55
ANK-like protein A	614	GGCACCCATTGAGGAACTGAA GAGCAGTGTACCCAATTTCGTCCC	59	E/I/E/I/E	3^{rd} /1^{st} GT-AG	0.00722	0/1	21/6	EU999949	8.01E-180	88
ANK-like protein B	353	GACGCCAGTGATGAGCCCCTTAGA GGATTGGCGCCGTATGAAAGCA	59	E/I/E	3^{rd} GT-AG	0.00490	0/3	8/3	EU999949	8.01E-180	88
β-actin (AS)[†]	270	CCTCGCGCCGTCTTCCCTTCC GCGGTTGGCCTTTGGGTTGAG	62	E		0.02158	4/15	11/7	AM886165	9.31E-68	86
Calreticulin	371	GTCCAGGCACTAAAAAGGTTCAT CCATTCACCGTCCATTTCAT	62	E		0.00463	1/5	4/2	JQ682618	0	97
Cathepsin L	523	GATGATAGGGGCTTTGTGGATGTC AAACCAAAGGATAGGATGCAGCAC	62	E/I/E	1^{st} GT-AG	0.00208	0/4	6/3	JQ682618	1.76E-180	98
CHF	231	CCGTAAGGGCGTAGGTGAAGAGTG CGTCGCGGAAGATTCAGTCAAAC	62	E/3'U		0.00302	1/2	3/0	GQ497446	6E-10	53
Cyclophilin A	284	CTAATGCTGGACCCAACACC CCTCCACTCCAATTCTAGCTGTAA	62	E		0.00000	0/0	0/0	JX258127	4.79E-87	90
ELOVL6 A	225	CGTCTGCATGAACTACCTGGTG GTGTACATGAGGAGGGAAATCTTG	61	E		0.00187	1/1	1/1	GL453806	8.04E-105	76
ELOVL6 B	252	CAAGGACACGGCTATTCAGAAC ATGCTCTTATGCTCGAATCGTTT	61	3'U		0.00509	0/0	1/3	GL453806	8.04E-105	76
FAMeT	430	GGAGTACCGAGAGTTCTGGATTG ATCGGTGTGGTTTCCTCAG	64	E/I/E	1^{st} GT-AG	0.00481	0/3	5/4	JN704303	0	98

Ferritin	214	CGAGCTGGACCAACTGGACTT CATTCAGCACCAAAGAACCACAG	60	E/3'U		0.01167	1/4	7/1	JQ670927	3.34E-108	100
FLCS	275	CATAAGCAAAAGAGGAGGAGTGAC GGCCACAGCTCTTATGAACAAA	62	3'U		0.00510	0/0	4/1	FJ446525	1.75E-69	76
FAP	289	AGGAAGAGGCATCTGCAAAA GCAGTGGACTGTTCCTTTCC	59	E		0.00647	6/2	4/4	XM001310210	1.18E-8	47
GAPDH	175	GCCGGCATTCAGCTCAGCAAAAC CCCTCCTTTCATTCGCATCCACC	62	E/3'U		0.00499	1/2	2/2	HM157285	0	98
GluCl	731	CGGAAGTGTTGTCGAAGTTGG ATCCGGAATGTCATGTCAAGAGT	59	E/I/E/I/E	1st/3rd GT-AG	0.00801	0/3	20/3	NM001177762	1.71E-55	61
GS	221	CCTCCTCCATCCACGACTTC AGGTAGGCAGGTGTATGATTGTTG	62	E		0.00366	0/3	2/1	JF738076	0	96
Glycogenin (AS) †	487	GACTCTGCCCACCTTTCACTACTT CAAGAGCTGTCCATATGTGTCG	57	E/I/E	2nd GT-AG	0.00159	0/1	2/2	KB030947	1.43E-59	57
GP	222	GAGTCTCCATCAACCCTTCTTCTA GTCCAACTGCACAAATCAATC	59	E		0.00388	3/1	2/2	AB596876	0	96
HSP90	236	TGAAGGTGCAGATGGTGTTGAGA GTACGAAGAGGAGGGCACGAAAT	61	E		0.00308	0/2	2/0	HQ162267	0	98
HMGBa	236	GCCTTCTTCTTCTACGCCAACGAC GGGTGTCCATTGCTAGCCTTCATC	59	E		0.00648	1/5	4/2	HQ228174	6.66E-89	85
ING (FS/AS) †	449	CCGGGTCGCAAGAAACAGAAAAG TGTCAACCCAACGCATGCAAAA	61	E/I/E	3rd GT-AG	0.00942	0/4	12/5	XM002428624	1.82E-42	88
LDLR	230	GGTCCAAAGGGTTATATGAAGTGC CTCCACCACTAACTTTTGCGTATC	62	E		0.00169	0/1	1/0	XM001867854	2.29E-52	55
METTL7A	217	TCTACGAGCACATCCACGAGTTC GGCTTGATGATCTGGAATATGAGG	62	E		0.00575	2/3	4/1	NM001193711	4.21E-17	62
MHCTa A1	347	CAGCGAGATGTTTGAGAAGCAG TCCTTGGTAGCCTTGTCCTCC	59	E/I/E	3rd GT-AG	0.00193	1/0	0/2	AB758443	0	87
MHCTa A2	307	CCGTGTCATCTCCCAGCAGCCTCT TGCATGTCCTCGCCGTCGTCAAT	59	E/I/E	1st GT-AG	0.00272	1/2	2/1	AB758443	0	87
MHCTa A3	162	CTGCTGGGTCCCTCATCCTAC GACATGTCCTCGCACTTCTCGTAT	60	E		0.00294	1/0	0/1	AB758443	0	87
MHCTb B1	158	TGCTGTCAAGACCGGTAAGAAG CGTCGGACTGGAAGGTGA	62	E		0.01000	4/0	2/2	AB758442	0	87
MHCTb B2	411	AGACCCTGACTGCTTCCAATGC GATGCTGTTCACCTCCTGCTCA	59	E/I/E	3rd GT-AG	0.00400	2/0	2/5	AB758442	0	87
MLC2	171	AAGTCTGGAGGAGGAAGCAATGTC	59	E		0.00795	6/3	5/4	HM486526	9.9E-28	79

		GTCGAGCTCCTGGTCTGTGG									
NHCP	343	ATGAGAATAAACCCAAGCGTC CCTTCATAGCCTTTTCATACTCTG	61	E/I/E	3rd GT-AG	0.00246	1/0	3/1	XM001741858	1.89E-9	85
Peroxidasin	158	CAACTGAAGGAGCTGCACAAAG GGGCTTAAGCTGTAGTCTCCAATC	62	E		0.00566	2/2	4/0	GL436239	8.62E-114	58
PHLCFM2	623	TAGGAGGTGGTGAAGCATCTGGAC TAGTAGCTGGAAGGCATGAGGTCG	64	E/I/E	2nd GT-AG	0.00713	0/2	6/7	GL440840	2.24E-18	57
PRDM5	208	TCGCCAGCCACTTGAAGACC TGCATCTCCACAGTCACCGAATAC	62	E		0.00238	1/1	2/0	CH471056	5.86E-8	56
RhoA	174	TCTGCGAAATGATGCCACT GCAGCCTTTGTAGCAGTTTCA	60	E		0.00395	3/2	1/4	HM581521	8.37E-114	97
RP	451	GTTGCTGCTGCTTCTAATGGACAT GCTGCGACTAATGCTGTGTGG	62	E/I/E	3rd GT-AG	0.00414	4/2	11/3	XM964803	0	71
SERCA	175	GTGATTATGGCTGTGGTCGTGAC AATACTTCTTTGCGGGGAGGGT	59	E		0.01164	4/1	3/2	AF025849	0	95
SMA 3/5	200	GATGAACGTTGCTGCTGCCTCCTC GGTCCTTCCTGATGTCGATGTCGC	62	E		0.01522	2/6	6/2	FJ217208 FJ217210	7.21E-163	96
SMA 6/8	162	ATGAGGCCCAGAGCAAGAGAG GGGGAGCCTCAGTAAGCATTGTA	59	E		0.02637	4/7	10/1	FJ217211 FJ217213	4.35E-165	96
TGF-βIP	398	TGGAGTCACCTGGGGCCTTCAC CGTTGTGTCCCTGATGCCTGGTTT	62	E/I/E	3rd GT-AG	0.00521	0/2	7/3	XM002402435 XM003494674	5.18E-118	58
Tropomyosin	312	CCCACTACTATAACTCCTCCTCCT CGGAAGCTGATGATACTAAAACAC	60	3'U		0.00149	0/0	0/1	GU369816	8.37E-81	100
UGP2	360	GCCTTCTGTGGAGTGGAGTGATA GCTGGACAGTGAGATCTAAAAACG	62	E/I/E	3rd GT-AG	0.01479	3/7	13/6	XM001847430	0	84
VAIPRCP	177	TTCGAGCTAGCTAGGCAGTTTGAG CACTTGTTCCATTGTCGCTTCTG	60	E		0.00380	2/3	3/2	XM001024253	4.27E-11	47
Zasp52	141	AGGTAGTGCCGTGTTTGTCTTGTG GCCTGAAGCCCCATGATGC	62	E		0.00360	0/1	0/1	JN034422 XM001862132	1.26E-53	63

Locus[a]: Locus symbol is abbreviated according to the gene name of the best BLAST match for each sequence. ANK-like protein, Ankyrin repeats-like protein; CHF, Crustacean hematopoietic factor; ELOVL6, Elongation of very long chain fatty acid protein 6; FAMeT, Farnesoic Acid O-methyltransferase; FAP, Fimbriae-associate protein; FLCS, Ferritin light chain subunit; FTSJ, methyltransferase domain-containing protein 1-like; GluCl, Glutamate-gated chloride channel; GAPDH, Glyceraldehyde-3-phosphate dehydrogenase; GP, Glycogen phosphorylase; GS, Glutamine synthetase; HMGBa, High mobility group box a; HSP90, Heat shock protein 90; ING, Inhibitor of growth; LDLR, Low-density lipoprotein receptor; METTL7A, Methyltransferase-like protein 7A; MHCTa/b, Myosin heavy chain type a/b; MLC2, Myosin light chain 2;

NHCP, Non-histone chromosomal protein; PHLCFM2, Pleckstrin-like protein domain-containing family M member 2; PRDM5, PR domain containing 5 isoform CRA_c; RP, Rolling pebbles; SERCA, Sarco endoplasmic reticulum $Ca2^{+}$ atpase; SMA, Skeletal muscle actin; TGF-βIP, Transforming growth factor beta induced precursor (high homology with Fasciclin); UGP2, UTP-glucose-1-phosphate uridylyltransferase 2; VAIPRCP, Viral A-type inclusion protein repeat containing protein, Zasp52, Z-band PDZ-motif protein 52 isoform 1.

Ta[b]: Annealing temperature

Amplicon content[c]: Exon, 5' or 3'UTR, and Intron

Intron features[d]: Intron inserted position in triple codons and start-end sequences.

π[e]: Nucleotide diversity

Ka/Ks[f]: Non-synonymous (Ka) and synonymous (Ks) mutations

Ti/Tv[g]: Transition (Ti) and tranversion (Tv)

GenBank hits[h]: Top BLASTx hit (gene accession number)in GenBank

E-value/Similarity[i]: E-value and similarity of top BLASTx hit in GenBank

[†]FS indicates frame shift and AS indicates alternative splice. 454 consensus and 454 haplotype sequences were not included for nucleotide diversity analysis.

Table 2. MassARRAY primers used for genotyping in *M. rosenbergii*.

Locus[a]	SNP position & type[b]	SNP source[c]	Location	S/NS/NA[d] Protein change	Primer sequences
ANK-like protein_A1	62_T/C	ABI	Intron	NA	PCR1: ACGTTGGATGGTGCGGTGCAAATGTTAGTG PCR2: ACGTTGGATGGACGTTTGGGAGAGAAAGTG EXT: CGGTGCAAATGTTAGTGTTAAGTAA
ANK-like protein_A2	213_A/T/G	ABI	Intron	NA	PCR1: ACGTTGGATGGTAACGGCGTCTTTCACCTG PCR2: ACGTTGGATGTGGCATATAAGATTCTCTCC EXT: AGAGTTGTTTATGCAAACAGAT
ANK-like protein_B	285_A/G	ABI/454	Exon	S	PCR1: ACGTTGGATGCGAACCCAAATGCGAGATAC PCR2: ACGTTGGATGTATGAAAGCAGGAGGTCCAG EXT: CCACTGCGAGATACTTCTTCGGCTC
Cathepsin L	351_T/C	ABI/454	Exon	S	PCR1: ACGTTGGATGATCCTCGGTAGTGCCATAAC PCR2: ACGTTGGATGTACAGTGATCCAGACTGCTC EXT: ACATCAAGAACTCCATGATCCAA
Calreticulin	290_A/G	ABI/454	Exon	S	PCR1: ACGTTGGATGGGAATGTGCTCTGGTTGATC PCR2: ACGTTGGATGAGAAGCCAGCAAACCTGACG EXT: CTCTGGTTGATCCCAATC
CHF	94_A/G	ABI	Exon	S	PCR1: ACGTTGGATGACTGCGGGACGTATCAATAG PCR2: ACGTTGGATGCTATAGGGCTGATCTGTGAC EXT: AAGAAATAGGCCCTGCAATAACCTTG
FAP	167_A/G	ABI	Exon	NS Iso → Met	PCR1: ACGTTGGATGCAGATCTTTCCTCTTTTTCAG PCR2: ACGTTGGATGCTGATGATAAATATGAAACAG EXT: CTTGCCTTTCCAAGACTTACATCTTC
FAMeT	335_T/C	ABI/454	Exon	S	PCR1: ACGTTGGATGAACAGCGATGACAGCTTGAC PCR2: ACGTTGGATGAAGCTAAGTGGGCGTCATTG EXT: CCATCGATGACAGCTTGACCTACAACTT
Fasciclin/TGF-βIP_A1	307_T/C	ABI/454	Exon	S	PCR1: ACGTTGGATGGATCCTTGAGAAGCCAGATG PCR2: ACGTTGGATGCATCCGACAGAAGTTTGGTG EXT: GTGGAACCCTAGCAATCC
Fasciclin/TGF-βIP_A2	71_A/T/C	ABI	Intron	NA	PCR1: ACGTTGGATGCGCCAACTTCACGACTTTTG PCR2: ACGTTGGATGGAAGCCTTTGACGTAAGAGC EXT: CATAAATATAAGAAGTATAATCCACCA
Ferritin	36_T/C	ABI	Exon	S	PCR1: ACGTTGGATGGGCTATGATGAAGGTTTAGTG

	All C[†]				PCR2: ACGTTGGATGATGATCACCAAGCTGAAGCG EXT: CCCAAGTGCAGTTCCTTGTC
FLCS	69_A/G	ABI/454	3' UTR	NA	PCR1: ACGTTGGATGAGCCAAATTTATAGCATGA PCR2: ACGTTGGATGGAGTGACTCATGTGGTTAAG EXT: GTTAATTACTGAGGAACAAATTTC
GAPDH	69_T/G	ABI/454	Exon	S	PCR1: ACGTTGGATGGCATGTGCTTCAAGAGATCG PCR2: ACGTTGGATGTTCGTGAAGGTCGTCTCTTG EXT: GTGCGTCAAGAGATCGATGACGCGGTG
GluCl_A1	66_T/C	ABI	Intron	NA	PCR1: ACGTTGGATGTCCTCAAGTGAGACCATCGG PCR2: ACGTTGGATGGAAAATGACTTAACAAAGA EXT: CATCGGTTAACAATGGTATG
GluCl_A2	326_A/T	ABI	Intron	NA	PCR1: ACGTTGGATGATGGACGTCGATGATGTAG PCR2: ACGTTGGATGGCCGTGCTAAGTAAAGCATT EXT: CGATGATGTAGACATGGT
GP	149_T/C	ABI/454	Exon	S	PCR1: ACGTTGGATGAGCAGTATGGTAACCAGGAG PCR2: ACGTTGGATGAAGGCTAATCCAGGTGGTTC EXT: CCAGCTTTTCCTCCAATCAT
HMGBa	87_G/C	ABI/454	Exon	S	PCR1: ACGTTGGATGTCTCGTACTTCAGCTTCTCG PCR2: ACGTTGGATGAACCCGGACTTCTCCGTAG EXT: GGGAGGTCAATTCGTTCCACTGGCGGCC
HSP90	94_T/C	ABI/454	Exon	S	PCR1: ACGTTGGATGCTGAAGATGAAGAACTCAAC PCR2: ACGTTGGATGTCCTCAGAGATGTCATCAGG EXT: CTACCTCAACAAAACCAAGCCACT
ING	84_A/G	ABI	Exon	S	PCR1: ACGTTGGATGTTGGGATCCACAGGCATGTC PCR2: ACGTTGGATGGAAACAGAATCTGCCATTGC EXT: TGTCCAGTACGTCTGA
METTL7A	63_T/C Failed[‡]	ABI	Exon	NS Phe → Leu	PCR1: ACGTTGGATGAACCACGCCTTCAGGAGG PCR2: ACGTTGGATGTTCAGGATGTCCCTGTTGAG EXT: TAGGCGCCTGCAGAGC
NHCP	230_T/C	ABI	Intron	NA	PCR1: ACGTTGGATGCTTCCCATTCCTGAATAAAG PCR2: ACGTTGGATGAAATGATGTTTACTGTAGC EXT: CTTCCCTTCATGGACAC
Peroxidasin	113_T/C	ABI	Exon	NS Leu → Pro	PCR1: ACGTTGGATGTTCAAGATTCCGAGCTCGAC PCR2: ACGTTGGATGCTTGCCAAGGGCTTAAGCTG EXT: CACGAGCTCGACTAATTCTC

PHLCFM2	68_A/C	ABI	Exon	S	PCR1: ACGTTGGATGTTGTCATCTTGGAAGCCTGC PCR2: ACGTTGGATGACAGCTACTCATGTACTCTG EXT: GATCGTGGAAGCCTGCTTATTTCAT
RP	192_T/C	ABI	Intron	NA	PCR1: ACGTTGGATGTATTTAGCAACTCAACTGC PCR2: ACGTTGGATGGGTAGTTTAGTTCATTAAC EXT: CCTCCACTCAACTGCAAGCACA
SERCA	58_T/C	ABI	Exon	NS Ser → Phe	PCR1: ACGTTGGATGTGATTATGGCTGTGGTCGTG PCR2: ACGTTGGATGCGAGATCCATATCTTCTGGC EXT: TCGTTGCCTTGCCCT
SMA3/5	136_C/G Failed[‡]	ABI/454	Exon	NS Ser → Cys	PCR1: ACGTTGGATGGTTGGCCTTTGGGTTAAGGG PCR2: ACGTTGGATGTCTGGTACCACACCTTCTAC EXT: CCCTCAGTAAGCATTGTAGGG
UGP2_A1	200_T/C	ABI/454	Exon	S	PCR1: ACGTTGGATGATCTGTTTCTCCTGCAGGTG PCR2: ACGTTGGATGCCACCACCAACTTATTTAAC EXT: GACATGCAGGTGAGGAAATATGC
UGP2_A2	84_A/C	ABI	Intron	NA	PCR1: ACGTTGGATGTCACAGGAATCACAACAATG PCR2: ACGTTGGATGGATATACAGAAGCTCCCTGG EXT: AATCACAACAATGAATAAACTCA
VAIPRCP	57_A/G	ABI/454	Exon	S	PCR1: ACGTTGGATGTTGACTTCGTCCACATTTGC PCR2: ACGTTGGATGCAGTTTGAGCCTGAGTTCAC EXT: CTTTAATTACCTCTACTCGTTC
Zasp52	38_A/T	ABI/454	Exon	S	PCR1: ACGTTGGATGATCAGGTAGTGCCGTGTTTG PCR2: ACGTTGGATGTTCTTGGTTGGGTGAATCGC EXT: GTGAGAGAGAGACGCT

Locus[a]: Locus symbol is abbreviated according to the gene name of the best BLAST match for each sequence. See Table 1 for description.

SNP position & type[b]: SNP position based on each contig length sequenced (Supporting information).

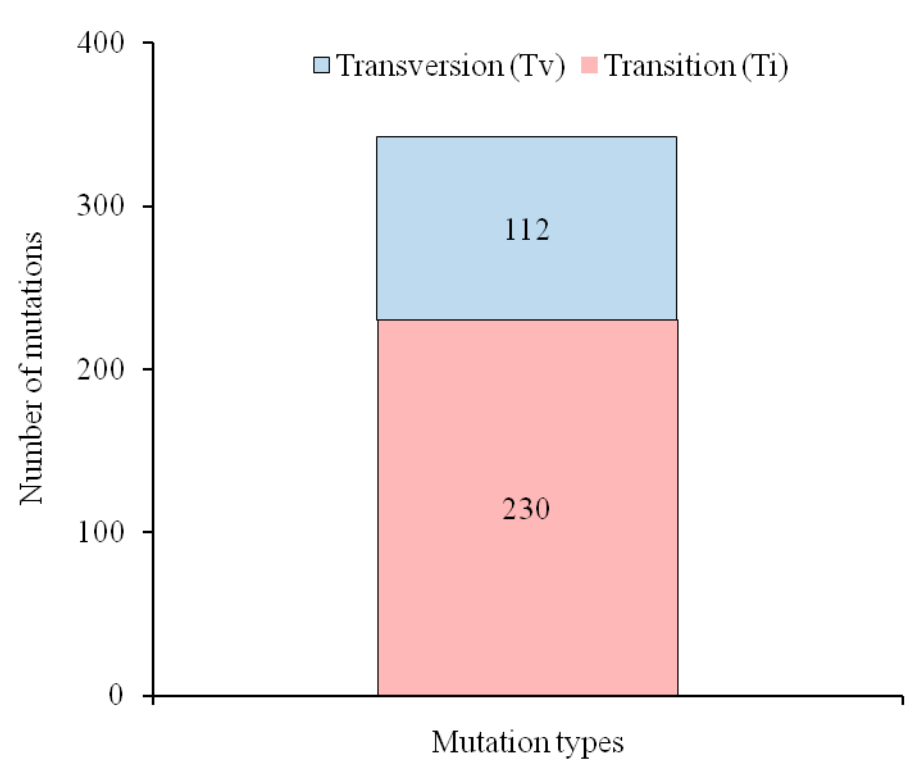
SNP source[c]: Source of initial SNP identification. ABI, Sanger sequencing; 454, 454 pyrosequencing.

S/NS/NA[d]: S, Synonymous; NS, Non-synonymous; NA, Not applicable.

All C[†]: All homozygosity C genotyped in SEQUENOM. Not considered any further analysis.

Failed[‡]: Genotyping failed in SEQUENOM. No further analysis performed.

A



B

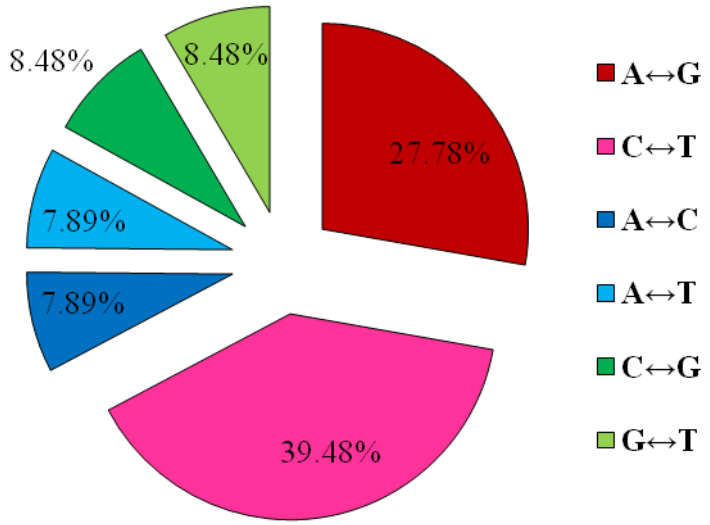


Figure 1. Classes of SNPs detected in 47 amplicons in *M. rosenbergii*. A) Total number of transversion and transition. B) Percentage of polymorphism variation.

Table 3. Summary of SNPs in growth and muscle development candidate genes in *M. rosenbergii*. H_o: observed heterozygosity, H_e: expected heterozygosity.

Locus	SNP type	Fast			Slow			Allele frequency		
		H_o	H_e	HWE (P-value)	H_o	H_e	HWE (P-value)	Fast	Slow	Overall
	TT	52	54.66	0.20	45	40.21	**0.05***	0.74	0.63	0.69
	T/C	44	38.68		37	46.59		-	-	-
ANK-like protein_A1	CC	4	6.66		18	13.20		0.26	0.37	0.31
	AA	48	47.50	0.21	40	35.28	0.16	0.69	0.60	0.64
	A/T	36	34.67		35	42.46		-	-	-
	TT	4	6.16		15	12.49		0.25	0.35	0.30
	A/G	6	8.32		4	5.98		-	-	-
	T/G	6	3.02		6	3.57		-	-	-
ANK-like protein_A2	GG	0	0.33		0	0.22		0.06	0.05	0.06
	AA	69	70.49	0.45	73	73.04	1.00	0.84	0.86	0.85
	A/G	30	27.02		25	24.92		-	-	-
ANK-like protein_B	GG	1	2.49		2	2.04		0.16	0.14	0.15
	TT	97	97.01	1.00	91	91.18	1.00	0.98	0.95	0.97
	T/C	3	2.97		9	8.64		-	-	-
Cathepsin L	CC	0	0.02		0	0.18		0.02	0.05	0.03
	AA	71	71.34	1.00	67	63.92	0.07	0.85	0.80	0.82
	A/G	27	26.32		26	32.16		-	-	-
Calreticulin	GG	2	2.34		7	3.92		0.15	0.20	0.18
	AA	77	76.51	0.64	67	68.82	0.29	0.88	0.83	0.85
	A/G	21	21.99		32	28.36		-	-	-
CHF	GG	2	1.5		1	2.82		0.12	0.17	0.15
	AA	16	13.94	0.40	18	13.57	0.09	0.38	0.37	0.37
	A/G	43	47.11		38	46.86		-	-	-
FAP	GG	41	38.95		44	39.57		0.62	0.63	0.63
	TT	6	5.43	0.78	10	7.46	0.22	0.24	0.28	0.26
	T/C	35	36.14		35	40.08		-	-	-
FAMeT	CC	59	58.43		55	52.46		0.76	0.72	0.74

	TT	0	0	NA	0	0	NA	0	0	0
	T/C	0	0		0	0		-	-	-
Ferritin	CC	100	100		100	100		100	100	100
	AA	18	17.52	1.00	20	15.09	0.06	0.42	0.39	0.41
	A/G	48	48.96		38	47.82		-	-	-
FLCS	GG	34	33.52		42	37.09		0.58	0.61	0.59
	TT	38	37.09	0.83	26	30.68	0.07	0.61	0.55	0.58
	T/G	46	47.82		59	49.64		-	-	-
GAPDH	GG	16	15.09		15	19.68		0.39	0.45	0.42
	TT	32	17.10	**0.00**‡	41	23.88	**0.00**‡	0.42	0.49	0.45
	T/C	19	48.80		16	50.23		-	-	-
GluCl_A1	CC	49	34.10		43	25.89		0.58	0.51	0.55
	AA	33	37.09	0.09	44	43.45	0.83	0.61	0.66	0.63
	A/T	56	47.82		44	45.10		-	-	-
GluCl_A2	TT	11	15.09		12	11.45		0.39	0.34	0.37
	TT	66	66.35	1.00	55	52.46	0.21	0.81	0.73	0.77
	T/C	31	30.30		35	40.08		-	-	-
GP	CC	3	3.35		10	70.46		0.19	0.27	0.23
	GG	0	0.14	1.00	0	0.23	1.00	0.04	0.05	0.04
	G/C	8	7.72		10	9.55		-	-	-
HMGBa	CC	92	92.14		90	90.22		0.96	0.95	0.96
	TT	0	1.63	0.36	0	0.60	1.00	0.13	0.08	0.11
	T/C	26	22.74		16	14.80		-	-	-
HSP90	CC	74	75.63		84	84.60		0.87	0.92	0.89
	AA	6	9.81	0.10	16	11.79	0.07	0.32	0.35	0.33
	A/G	51	43.37		37	45.42		-	-	-
ING	GG	43	46.82		47	42.79		0.68	0.65	0.67
	TT	0	0.14	1.00	0	0.46	1.00	0.04	0.07	0.05
	T/C	8	7.72		14	13.08		-	-	-
NHCP	CC	92	92.14		86	86.46		0.96	0.93	0.95
	TT	74	75.63	0.36	70	71.34	0.45	0.87	0.84	0.86
	T/C	26	22.74		29	26.32		-	-	-
Peroxidasin	CC	0	1.63		1	2.34		0.13	0.16	0.14
	AA	34	31.24	0.32	29	32.94	0.15	0.56	0.58	0.57
PHLCFM2	A/C	44	49.52		57	49.12		-	-	-

	CC	22	19.24		14	17.94		0.44	0.42	0.43
	TT	9	11.11	0.38	9	13.21	0.09	0.34	0.37	0.35
	T/C	49	44.78		55	46.59		-	-	-
RP	CC	42	44.11		36	40.20		0.66	0.63	0.65
	TT	18	15.48	0.30	21	22.44	0.68	0.40	0.47	0.43
	T/C	43	48.04		53	50.12		-	-	-
SERCA	CC	39	36.48		26	27.44		0.60	0.53	0.57
	TT	5	6.41	0.60	11	7.74	0.13	0.25	0.28	0.27
	T/C	41	38.18		34	40.52		-	-	-
TGF-βIP_A1	CC	54	55.41		55	51.74		0.75	0.72	0.73
	AA	33	27.96	**0.03**[*]	32	25.38	**0.00**[‡]	0.53	0.51	0.52
	A/T	23	22.90		18	24.87		-	-	-
	TT	5	4.54		11	5.91		0.22	0.24	0.23
	A/C	17	27.17		19	25.37		-	-	-
	T/C	10	11.02		9	12.31		-	-	-
TGF-βIP_A2	CC	12	6.41		11	6.16		0.25	0.25	0.25
	TT	9	6.41	0.19	9	8.02	0.63	0.26	0.29	0.27
	T/C	33	38.18		39	40.96		-	-	-
UGP2_A1	CC	58	55.41		52	51.02		0.74	0.71	0.73
	AA	15	15.09	1.00	8	9.19	0.64	0.39	0.31	0.35
	A/C	48	47.82		45	42.61		-	-	-
UGP2_A2	CC	37	37.09		47	48.20		0.61	0.69	0.65
	AA	18	13.57	0.09	21	17.94	0.23	0.37	0.42	0.40
	A/G	38	46.86		43	49.12		-	-	-
VAIPRCP	GG	44	39.57		36	32.94		0.63	0.58	0.60
	AA	71	66.35	**0.00**[‡]	72	73.04	0.68	0.82	0.85	0.83
	A/T	21	30.30		27	24.92		-	-	-
Zasp52	TT	8	3.35		1	2.04		0.18	0.15	0.17

[*]Significant P value after FDR

[‡]Significant P value after Bonferoni correction

Table 4. Estimated individual SNP effects on the genetic values of body weight estimated breeding values (EBVs) in *M. rosenbergii* under additive and dominance models.

	Additive model						Dominant model				
SNP in loci	**Geno-type[a]**	**Num-ber**	**Effect ± s.e[b]**	**P-value**	**R^2**	**SNP in loci**	**Geno-type[a]**	**Num-ber**	**Effect ± s.e[b]**	**P-value**	**R^2**
ANK-like protein_A1	1,2,3	200	-0.070 ± 0.030	**0.021**[*]	0.026	ANK-like protein_A1	TT T/C CC	97 81 22	0.158 ± 0.029 0.154 ± 0.032 -0.047 ± 0.061	**0.008**[*]	0.048
ANK-like protein_A_A2	1,2,3	200	0.054 ± 0.028	0.057[†]	0.018	ANK-like protein_T_A2	A/G/AG TG/TA TT	98 83 19	0.166 ± 0.029 0.136 ± 0.031 0.039 ± 0.066	**0.018**[*]	0.040
ANK-like protein_T_A2	1,2,3	200	-0.075 ± 0.031	**0.017**[*]	0.029	GP	CC C/T TT	13 66 121	0.005 ± 0.080 0.098 ± 0.035 0.168 ± 0.026	0.070[†]	0.026
GAPDH	1,2,3	200	-0.057 ± 0.030	0.062[†]	0.017	HSP90	CC C/T	158 42	0.102 ± 0.023 0.255 ± 0.044	**0.002**[*]	0.047
GP	1,2,3	200	0.076 ± 0.033	**0.022**[*]	0.026	ING	GG GA AA	90 88 22	0.117 ± 0.030 0.179 ± 0.031 0.023 ± 0.061	0.058[†]	0.028
HSP90	1,2,3	200	0.153 ± 0.049	**0.002**[*]	0.047	NHCP	CC C/T	178 22	0.147 ± 0.022 0.027 ± 0.062	0.067[†]	0.017
NCHP	1,2,3	200	-0.120 ± 0.065	0.067[†]	0.017	TGF-βIP_T_A2	A/C/AC TA/TC TT	124 60 16	0.140 ± 0.026 0.172 ± 0.037 -0.058 ± 0.072	**0.017**[*]	0.041

Genotype[a]: refers to the number of copies of a SNP allele for dominance model.

[b]Estimated SNP allele substitute effect under additive model and estimated individual genotype effect under dominance model. s.e – standard error. [†]0.05<P<0.7.

[*]Bold characters: P<0.05 after adjusting Tukey-Kramer Test.

CHAPTER 5. General Discussion and Conclusions

The importance of aquaculture as a production industry especially in developing nations is well recognised (Gjedrem 2012; Gjedrem *et al*. 2012) and in the future this industry will follow the lead of terrestrial agriculture where farming improved breeds becomes the industry standard. Traditional approaches to genetic improvement of livestock, crop and aquatic organisms has been very successful (Thornton 2010; McAndrew & Napier 2011; Morrell *et al*. 2012), but can be inexact, inefficient, and/or time-consuming, with many traits difficult to estimate and/or measurable only after the target organism has reached maturity (Bastiaansen *et al*. 2012; Börner & Reinsch 2012). Breeding values including relative inbreeding level and declines in genetic diversity can be predicted and estimated potentially more accurately using genomic information for target organisms, an approach known as marker-assisted selection (MAS) or genetic/genomic selection (GS) (Goddard & Hayes 2007; Bastiaansen *et al*. 2012; Hayes *et al*. 2013). While MAS is still a relatively new approach in aquatic sciences, this approach has already been used successfully to increase the rate of improvement in some farmed stocks of aquatic species (Miller 2010; Wheeler *et al*. 2010; Morrell *et al*. 2012; Varshney *et al*. 2012; Hayes *et al*. 2013). While a number of traits provide potential targets for improvement in modern breeding programs in aquaculture species, to date genomic studies have focused mainly on improving growth rate with some programs producing significant production gains in particular species including for Atlantic salmon, carp, catfish, rainbow trout, and tilapia (McAndrew & Napier 2011).

A fundamental requirement of any efficient modern breeding program will not only be to accumulate a large enough reference population of several phenotypes but also to identify critical genes and provide basic genomic information. In particular, a large amount of genomic resources can be indispensible for a successful GS. At the time the current project commenced (2010), a well established stock improvement program for GFP was on going in Vietnam applying family selection that could provide reference phenotype resources for the current study. Therefore, the aim of the current study was to provide genomic resources for a non-model aquaculture species to develop a foundation for future trialling of GS. Using this resource, we explored several putative growth-related candidate genes in a transcriptome dataset based on criteria addressed in Chapters 2 and 3, and assessed the correlations between

candidate gene SNP markers and individual growth performance in an improved GFP culture line in Chapter 4. The current study is the first large-scale analysis of CGA targeted at stock improvement of GFP. The main focus was on growth traits using data generated from the earlier comprehensive study of a GFP transcriptome. Validation of SNP markers within selected families in Vietnam demonstrates the potential for these markers to be used in the selective breeding program in the future. Traditional selective breeding of GFP in Vietnam has already generated a fast growth line that is now available for the culture industry, and this development should improve profitability for farmers that culture this species (Dinh *et al*. 2012, 2013). Knowledge obtained in the current study suggest that MAS/GS can be integrated into this program in the future to increase the rate of genetic gains for *M. rosenbergii*, and the approach can also potentially be applied to other economically important cultured crustacean species.

5.1. Genomic selection

GS is the selection of individuals for breeding on the basis of availability of a large number of genome-wide markers, without a requirement for measuring external phenotypes (Hamblin *et al*. 2011). DNA genotype variation and/or genome-wide molecular markers recorded from a target organism can be used to calculate EBVs by predicting potential performance early in life so that individuals with the best potential phenotype can be identified earlier than with traditional breeding approaches (based on pedigree and phenotype alone), and in doing so accelerate breeding for optimized product quality and commercial return (Rothschild & Plastow 2008; Hayes *et al*. 2013). MAS and/or GS can be used to increase the rate of genetic gains in modern breeding programs but a fundamental difference between MAS and GS is the scale at which they can be implemented. MAS is limited in its ability to predict EBVs as only a small number of QTLs are normally assessed that are tagged by markers with well-defined associations with production phenotypes. Furthermore, MAS has largely been practiced on simple traits that are difficult to score and the approach is not usually appropriate for complex polygenic, quantitative traits (Gupta *et al*. 2010; Calus *et al*. 2012; Nakaya & Isobe 2012). GS, in contrast, applies a dense set of molecular markers that are distributed widely across the entire genome. This approach virtually guarantees that all detectable QTLs are likely to be in LD with at least one SNP marker (Goddard & Hayes 2009; Eggen 2012). GS also offers breeders the opportunity to decrease the generation interval, and to reduce inbreeding levels

over time because organisms can be selected accurately early in life, based on genomic predictions (Börner & Reinsch 2012; Cleveland *et al*. 2012; Heslot *et al*. 2012; Hayes *et al*. 2013). This can then provide more accurate EBVs for complex polygenic, quantitative traits compared with MAS (Gupta *et al*. 2010; Calus *et al*. 2012; Nakaya & Isobe 2012) because GS relies on LD between the markers and the polymorphsisms that cause variation in important traits.

Simulations and actual empirical studies of GS in livestock and crop species have shown levels of accuracy sufficient to generate rapid genetic gains per unit of time and cost (Börner & Reinsch 2012; Cleveland *et al*. 2012; Heslot *et al*. 2012) because a linear prediction equation, based on the marker genotypes, can predict the cumulative effect of many causal variants on the additive genetic value or breeding value of the organism (Hayes *et al*. 2013; Taylor 2013). Considering that most economically important traits are potentially influenced by epistatic and/or pleiotropic interactions from gene-gene or gene-environment interactions, a large amount of genetic/genomic data are often required to accurately estimate individual gene effects on a target trait. A key feature of GS, therefore, is that dense molecular marker coverage is possible across the whole genome so that most of the genetic variation in a trait can be explained potentially by the linked markers (Goddard & Hayes 2007). The generality of genetic interactions affecting most quantitative traits is, however, not always transparent and defining the source of variation can be very challenging, even when a large amount of molecular markers available. Another key factor when detecting epistatic relationships between genes is not only the number of markers employed but also the training sample size used (Jiang *et al*. 2011; Park & Lehner 2013). In the case of selection in aquatic species, manageability of breeding programs can provide a biased vision of genetic interactions. Therefore, the focus should be on understanding the stability and heritability of epistatic variation to determine whether this variation can effectively be captured in breeding programs, or whether it may be too unstable for practical use due to unpredictable responses to environmental variation (Xu *et al*. 2011; Springer 2013).

5.2. Selection of candidate genes related to growth performance in crustacean species

Understanding gene function and genetic effects on expressed phenotypes will be fundamental to developing better aquaculture production systems in the future. The criteria adopted here for selecting potential candidate genes, as discussed previously in Chapter 2,

was based on knowledge of the direct physiological roles of key genes that are regarded as being important in regulating crustacean growth, moulting and muscle development. Therefore, expanding this to include genes that affect cellular energy states, metabolic flux, ecdysteroid biosynthesis, and sexual maturation could provide a more comprehensive set of growth-related gene candidates since most growth-related phenotypes are complex traits that result from multi-gene interactions. When a reference genome sequence is available for GFP, genome based prediction (GBP) and genome wide association studies (GWAS) (Villanueva *et al*. 2012; Ober *et al*. 2012), could offer new opportunities for investigating patterns of variation among potentially hundreds of genes that influence important economic traits. The recent availability of large panels of SNPs in agricultural and domestic species has given new momentum to the search for mutations that underlie variation in complex traits (Goddard & Hayes 2009; Rubin *et al*. 2010; vonHoldt *et al*. 2010; Jiao *et al*. 2012; Morrell *et al*. 2012). In particular, gene-associated high-density SNP arrays have been suggested as an efficient, powerful and cost-effective means for addressing genetic applications including improving the accuracy of estimating EBVs in catfish and Atlantic salmon (Wang *et al*. 2010b; Liu *et al*. 2011; Houston *et al*. 2012). While many genes involved in growth rate phenotypes are likely to be identified in the future, gene sequences identified here in GFP that include; 5-HT, Actin, α-AMY, CatL, CHH, Cyps, FABPs, GAPDH, HSPs, MF, MIH, MSTN, MYC, STAT, SPARC, and TRAX are primary targets as candidate genes that affect growth phenotypes more widely in crustacean species. Their specific roles in GFP growth and moulting remain to be determined, but their widespread effects on these important traits in other species, identifies them as potential major gene targets on which more focus should be applied.

5.3. Transcriptomic data and GFP growth

At the start of the current project, all GFP tissue samples were transported from Vietnam preserved via RNA-later treatment. While RNA extraction was conducted on heptatopancreas and eyestalk tissues which at the time were considered to be important organs affecting crustacean growth (Uawisetwathana *et al*. 2011; Asusena *et al*. 2012; Brady *et al*. 2012; Zara *et al*. 2013), these tissue libraries were not included in the current study due to relative quantity/quality and pigmentation issues. In addition, muscle was not sampled from males as their growth performance was considered to be confounded by social factors (Thanh *et al*. 2009). Considering that males grow faster then females, a variety of male tissue samples will

be required in the futureto develop a more comprehensive understanding of growth in GFP, as well as to characterise the whole GFP genome and to determine the complexity of gene-gene and gene-environment interactions. Until these data are available, progress can be made by developing transcriptome data at a finer level of scale to gain insight into the genetic effects on growth.

Although several high-throughput sequencing technologies have become available over recent years, longer gene sequences were considered important for GFP in 2010, because a genome reference is still unavailable making it difficult to generate high quality contigs for *de novo* assembly from short sequences. Therefore, 454 sequencing technology was employed in the current study in preference to other NGSTs as this has proven to be the platform of choice for similar studies in other non-model species (Hornett *et al*. 2012; Ma *et al*. 2012). Even with longer reads, however, a large number of singletons were generated and remain unassembled mainly due to the high number of repeat sequences, G/C content and alternative splicing sequences obtained. Availability of longer sequence reads or paired sequence reads with better read depth could simplify sequence assembly in a future GFP study because they can overcome difficulties associated with repeat sequences. In particular, the value of availability of long contigs, generated by pair-end shotgun sequencing using plasmid and fosmid libraries, in the water flea assisted full genome sequencing and allowed a better understanding of a number of biological processes including functional responses of genes to environmental conditions (Colbourne *et al*. 2011). This approach has also yielded important knowledge regarding the evolution of the adaptive immune system for vaccine development and disease management in Atlantic cod. A BAC library shotgun and pair-end library with 454 pyrosequencing and Illumina was applied (Star *et al*. 2011). In addition, genes that respond to environmental stress and shell formation in Pacific oyster have been examined via fosmid pooling strategies using short reads with Illumina technologies (Zhang *et al*. 2012). While the current dataset does not represent a complete transcriptome for GFP, because only three tissue libraries (taken mainly from adult females) were screened, it still provides the best resource for deciphering the putative function of novel genes in each tissue at this stage. A large transcriptome dataset characterised from diverse tissues could eventually benefit future generations of genetic maps and genome studies by identifying putative functions of novel genes in each tissue. While the majority of ESTs in the transcriptome matched well to archived crustacean and other arthropod proteins based on BLAST searches and GO annotations, many unmatched sequences still may constitute novel

genes unique to the target species. It is not easy at this stage, however, to develop more detailed genomic information for GFP because of the limited transcriptome data that are available currently. Therefore, a wide range of tissue samples including those from different developmental stages, different tissues and from both sexes will be required to understand the molecular functions of specific genes identified here and to construct a fine transcriptome dataset for GFP. Combining the current transcriptome dataset with that from EST libraries generated using different NGS methodologies could also assist refining the existing transcriptome by increasing average read length and depth of genome coverage in GFP because more singletons could be scaffolded into larger contigs that would be more informative for identifying more putative genes and structural variants in future studies. If differential gene expression studies can also be included from RNA-Seq analysis, this will increase the rate at which the molecular basis of growth and other traits are characterised in GFP.

Understanding physiological characteristics of proteins with known functions from GO, KEGG pathways and InterProScan searches is a powerful approach for deciphering the roles of individual genes and for mapping functional relationships in new taxa. RNAi-mediated knockdown has also emerged as a powerful tool for manipulating gene expression in shrimp to identify key genes, gene families and gene domains because presence of a dsRNA in eukaryotic cells can trigger the post-transcriptional gene-silencing mechanism. In turn, this can lead to sequence-specific degradation of target mRNA (Krishnan *et al*. 2009; Hirono *et al*. 2011). While we did not explore additional physiological roles for any of the genes identified in GFP, more extensive studies will be required to develop a better understanding of the genes, proteins and enzymes in GFP in particular for metabolic pathways that show strong associations with growth and muscle development. Since most studies of commercial crustaceans to date have focused on identification and characterisation of pathogen defence-related genes (Dechklar *et al*. 2008; Su *et al*. 2008; Wu *et al*. 2010), the current expression of ESTs for growth and muscle development can provide important baseline information for discovery of new functional genes. Thus, genes that mediate growth performance and that are potentially of value in MAS/GS breeding programs largely still await discovery. Despite the fact that only limited data were generated from reproductive tissues, several key gene sequences were identified in ovary and testis that provide a good resource for studying reproductive traits because a few recent studies of *M. rosenbergii* have indicated that individual reproductive quality is correlated with individual growth

performance (Nhan *et al*. 2009, 2010). Other groups, notably in Israel, are currently studying genomic control of reproduction and sex determination in GFP, so data on novel genes from reproductive tissues in GFP identified here could also benefit this work (Ventura *et al*. 2011a, b; Ventura *et al*. 2012; Ventura & Sagi 2012).

5.4. Development of candidate gene SNP markers

Regardless of whether MAS and/or GS is used to identify individuals of preferred phenotype to increase the rate of genetic gain in modern breeding programs, a large number of gene-based markers are required to determine EBVs in target organisms (based on a sufficiently large reference pedigree and phenotype information) because complex traits (e.g. growth) often result from multi-gene interactions. ESTs provide a primary resource for developing Type I markers in many species (Bouck & Vision 2007). Data generated from NGST in the current study identified a large number of EST-SSRs but not as many EST-SNPs in the GFP transcriptome sequences. The small number of EST-SNPs could result from the small number of samples which resulted in relatively low sequence coverage and poor SNP-calling performance in possible CpG islands, promoter and UTR regions of the transcriptome which is common when sequencing depth is low (Wang *et al*. 2011). Despite this, SNP variants identified here provide essential genomic resources for future study of GFP. Many of these putative SNPs could represent allelic variants but they would require validation before being considered for future studies.

Limited numbers of suitable markers has been a major problem in the past for genetic applications in *M. rosenbergii,* including for assessing genetic diversity and for screening variation in geographically dispersed populations. Therefore, EPIC and EST-SNP markers identified in this study potentially may provide better markers for several genetic applications. Previous studies have shown that markers designed in or near highly conserved exons (EPIC, EST-SNP and EST-SSR) tend to be more widely transferable among species (Bouck & Vision 2007; Kim *et al*. 2011) and offer potential utility for examining genetic variation at both the intra- and inter-specific levels simultaneously, a feature that is particularly useful for studying species complexes (Li *et al*. 2010). Some of the markers recently identified fit these criteria, and hence would likely overcome many problems currently encountered in molecular ecological studies of GFP globally.

5.5. Candidate gene association studies with individual growth performance

CGA characterises variation in genes of known biological function relevant to the trait(s) of interest to investigate genotype-phenotype associations. This approach is being adopted increasingly in livestock and crop breeding programs (Hu *et al*. 2009; Morrell *et al*. 2012). Markers developed in putative growth-related genes in the present study could provide an important starting point for conducting CGA analysis in GFP, an approach that has proven to be effective for detecting genes and regions that underlie phenotypic variation in adaptively significant traits (Nachman *et al*. 2003). Exonic SNPs have been a major target for estimating the genetic basis of phenotypic correlations among production traits, but all SNPs should be considered for future association studies regardless of where they occur in genomic regions (coding *vs* non-coding regions) and the importance of polymorphisms (synonymous *vs* non-synonymous) because of their potential influence on phenotypic variation, an issue that has been addressed earlier in Chapter 4. In addition, a possible epistatic effect on EBVs from stepwise selection analysis was observed for GFP but showed a low proportion of genetic variance (R^2) because only a limited number of SNPs were screened, and which did not cover the GFP genome-based influence on EBVs. While relatively low R^2 could compensate for advantages of Type I markers (e.g. EST-SNPs), genome wide abundance and a direct link to physiological functions of genes, employing a genome-based high-density SNP genotyping approach could be an effective way to estimate R^2 more accurately and to identify more major and/or minor effect loci and/or SNPs that respond to selection (high, moderate and low genetic heritabilities).

While the detection of statistically significant associations between EBVs and individual SNP marker genotypes may be useful for making breeding decisions in GFP, a cautious approach should be adopted because statistical associations do not provide direct evidence for a real contribution to a phenotype (i.e. EBVs) (Beuzen *et al*. 2000). In addition, the current estimations were made only for EBVs in adult female individuals using body weight so their effects could be different when applied to other traits, to EBVs estimated from other traits, or in males or other developmental stages because some of the same loci or linked loci affect growth traits across all life-stages while some loci may be expressed only at certain times. A good example of this has been reported in coho salmon, where pleiotropic effects on length differed among ages (McClelland & Naish 2010). In particular, if the

current GFP loci have pleiotropic effects across traits expressed at a specific age, periodical moulting stages must be considered in crustaceans because this is a unique physiological process essential for normal muscle development and/or muscle degradation in growth of crustaceans. While this remains a novel area for study in crustacean species, a recent study has provided a good example of the applicability of a high-throughput sequencing approach for studying complex traits, including crustacean metamorphosis (Ventura *et al*. 2013). Despite further validation being required with larger sample sizes and several generations for new training populations, markers identified here potentially will be informative for predicting BVs in GFP because of their important physiological roles as well as the data presented here provide the first direct evidence for genetic variation in these functional loci.

Most breed improvement programs for aquatic species commence by focussing on improving growth rates, and later progress to include additional traits that show favourable correlations with growth rate (Gjedrem 2000; Nguyen *et al*. 2010a, b). In general, body weight in animals shows high (either positive or negative) associations with other production traits (Calus *et al*. 2012) as were observed for GFP where a positive relationship was evident between body weight and carcass weight (Dinh *et al*. 2013). The detection of a QTL for multiple traits in the same linkage group can be explained by either the linkage of two QTLs (one for each trait), or the presence of a single QTL with pleiotropic effects (Baranski *et al*. 2010). In the case of Atlantic salmon, Reid *et al*. (2005) detected a QTL for both body weight and condition factor in multiple linkage groups, suggesting that they may represent different sets of genes due to low genetic correlations reported between the two traits previously. The nature of this interaction was not clear and remains unresolved. Several markers identified in the current study appear to be in LD ($P < 0.001$) which suggests they may be to be associated with linked QTLs for related phenotypes, with epistatic and/or pleiotropic interactions. To confirm this, it will be necessary to identify if the markers in LD are present on the same chromosome in GFP. While a whole genome sequence for GFP will provide the ultimate roadmap for showing these markers are in LD, as well as identifying more putative genes and structural variants in future studies, development of a fine linkage map can also be an efficient approach in the interim for identifying QTLs (including for mapping relative locations on chromosomes and for understanding effects of recombination events) and to determine new candidate genes that control quantitative trait variation. Identification of QTLs and/or candidate genes that influence multiple traits could increase the efficiency of MAS/GS and so lead to more rapid genetic gains (Upadyayula *et al*. 2006) as long as large reference

populations (pedigree and phenotype information) are available. This approach will be informative in creating marker panels for each QTL for use in GFP stock improvement in addition to benefiting GFP culture industries by shortening the desirable selection interval, reducing the overall labour time and saving running costs for future GS breeding programs in this species because selection decisions can be predicted potentially before phenotypes of selection candidates are measured.

5.6. Evolutionary pressures

Estimating allele frequencies and Ka/Ks and Ti/Tv ratios have been used widely to understand selection pressure during molecular evolution and to monitor long-term effective population sizes, but the current data were observed under growth selection pressure using only EBVs from G2 adult females (body weight). Since the quality of the estimates can be highly dependent on farmers' breeding practices and undetected false positive associations can also affect accuracy of their estimations, future studies should consider different forms of selection driven by reproduction and at different developmental stages using several generations because of potential differential impacts which often act unevenly among genes and genomic regions (McClelland & Naish 2010; Loukovitis *et al*. 2011; Chen *et al*. 2012b; Gruber *et al*. 2012).

Evolutionary response is determined by phenotypic variation, by the strength and direction of selection, by the underlying genetic variation within a population and by the genetic covariance between traits (Falconer & Mackay 1996). In particular, genetic correlations between traits also have implications for adaptive evolution of a species, a process that limits the potential number of trait combinations that are advantageous, and thus constrains evolution of correlated phenotypes (McGuigan 2006). In general, body weight of animals shows positive and/or negative correlations with other production traits (Calus *et al*. 2012). A few studies conducted in crustaceans have identified that some negatively correlated traits (pathogen resistance and growth) may have resulted from selection of certain allelic variants acting in opposing ways (Gitterle *et al*. 2005; Ciobanu *et al*. 2010; Bonami & Widada 2011; Panphut *et al*. 2011). The nature of this interaction is not clear but if one trait is segregating in the offspring but the other is not, pleiotropy may fail to be detected. In addition, alleles may vary in their pleiotropic effect (Phillips & McGuigan 2006) so the identity of alleles present in the parents may impact the pleiotropic effects at different loci. While the

current study was not mainly focused on understanding evolutionary pressures on EBVs, information and gene-based markers obtained here could benefit from a future comparison of allele frequency across generations and among phenotypic traits. Identification of combinations of functional molecular polymorphisms that differ in frequencies among populations inhabiting different environments and/or generations, or with a different phenotypic range could facilitate future GFP stock improvement programs by estimating major/minor allele frequency, in particular for germplasm protection of endangered or for the commercially improved lines.

5.7. Future directions for stock improvement in GFP

5.7.1. Traditional selection to advance GFP stock improvement

GFP has been domesticated for approximately 40 years but the culture industry still currently relies on farming essentially wild/unimproved stocks (Thanh *et al*. 2009; Dinh *et al*. 2012). In general, base populations used in selection programs require high levels of genetic diversity in order to secure long-term genetic gains (Falconer & Mackay 1996). To improve brood-stock resources, family selection combined with pedigree mating is now used to improve the relative productivity of culture lines for expanding aquaculture industries (Thanh *et al*. 2009; Nhan *et al*. 2010; Aflalo *et al*. 2012; Dinh *et al*. 2012). The stock improvement program for Atlantic salmon initiated by AKVAFORSK in Norway and the GIFT stock improvement program for Nile tilapia initiated by ICLARM (now World Fish Centre) provide excellent examples of successful traditional genetic improvement programs for farmed aquatic species, and highlight the importance of basing any stock improvement program on an initial broad genetic base that includes genetic diversity sourced from different genetic populations (Ponzoni *et al*. 2011; Gjedrem 2012). In the GFP stock improvement program in Vietnam, a similar approach was applied but with fewer founding strains (two local Vietnamese and a single exotic Malaysian strain) to develop the Vietnamese base population (Dinh *et al*. 2012, 2013). In order to secure high genetic variation levels and to reduce the inbreeding impact of selection over the long-term in the Vietnamese program for GFP it may be necessary to freshen up the gene pool periodically from external stocks as is practiced for the Norwegian Atlantic salmon breed improvement program.

While the major freshwater prawn producing countries are all in Asia (e.g. Bangladesh, China, India, Myanmar, Taiwan, Thailand and Vietnam), selective breeding programs for GFP initiated recently in India and Vietnam mainly constitute an initial step towards long-term more sustainable production of GFP. Since GFP is an economically important species, a few recent studies have demonstrated that family selection combined with pedigree mating can be used successfully to improve the relative productivity of culture lines and to expand aquaculture industries with genetic gains of 10% to 26% per generation possible in the selected population compared with wild or base populations (Thanh *et al*. 2009; Nhan *et al*. 2010; Aflalo *et al*. 2012; Kitcharoen *et al*. 2012; Luan *et al*. 2012). An average positive response, 7% per generation to selection (genetic gain) for growth rate was reported in the Vietnamese stock improvement program (Dinh *et al*. 2012, 2013). Strains developed this way will, however, require rigorous testing under real farming conditions and in different production environments. While the traditional phenotype-based selection typically used to select for desirable production traits requires several generations to optimize genetic improvement, a well-designed stock improvement program in GFP must be in place before MAS/GS is applied to increase the selection accuracy based on genomic predictions over a short period.

5.7.2. Combination of genomic and traditional selection to advance GFP stock improvement

More than just genomic tools, however, need to be put into place. In order for the potential of genomics to be fulfilled, there needs to be parallel effort on collection of detailed trait information from diverse phenotypes and this should also include the use of NGSTs to help describe the environments where organisms are reared (Rothschild & Plastow 2008). In order to achieve the ultimate goal of genome-assisted breeding, phenotypic values will need to be measured in the training population and also in successive generations for new training to avoid the loss of significant associations between phenotypes and markers due to unpredictable effects from recombination. While traditional approaches to stock improvement that employ family selection have been trialled widely and applied successfully to a wide range of cultured aquatic species (McAndrew & Napier 2011), some recent studies have reported that traditional family selection combined with pedigree matings utilising genetic and/or genomic information can be more effective when multiple phenotypic traits are

considered (Goddard & Hayes 2007; Varshney *et al*. 2009; McAndrew & Napier 2011; Morrell *et al*. 2012). GS has been highlighted as a new approach for MAS in recent years because GS is a form of MAS where selection of individuals is based on genomic EBVs (Morrell *et al*. 2012) offering breeders the opportunity to determine EBVs more accurately and rapidly than using time-consuming traditional methods.

Currently, GS has not been optimised for many farmed species, but it is a potent, attractive and valuable approach for breed improvement in the future. New GS tools have been trialled, however, in a wide range of farm species mainly in terrestrial and crop breeding, although there have also been some recent applications in aquatic species including in Atlantic cod, Atlantic salmon, carp, catfish, Pacific oyster, Penaeid shrimp and tilapia (McAndrew & Napier 2011; Quinn *et al*. 2012: Robalino *et al*. 2012). Most breed improvement programs in aquatic species, however, usually commence by focussing on improving growth rate and only later progress to include additional production traits (e.g. environmental tolerance, food conversion efficiency, body shape, sexual maturation, disease resistance, flesh quality, carcass weight, and survival rate etc.) because these traits tend to show favourable correlations with growth rate (Gjedrem 2000; Nguyen *et al*. 2010a, b). For GFP, the breed improvement program in Vietnam has focused only on improving growth rate. Additional traits including reproductive quality, sex ratio, proportion of edible to non-edible meat (carcass ratio) and meat quality (e.g. protein, fat and moisture contests) are all potential future targets for improvement and will in time be added to the breeding objectives. Results in the second generation Vietnamese lines in the current study showed that the selection response for both body weight and abdominal weight were approximately 7% (Dinh *et al*. 2013), indicating that the increase in both weight and carcass weight were linear over time.

While the benefits of applying genomic approaches to improving growth-related traits is still openly debated, there is more general agreement about applying MAS/GS tools for traits with low heritability traits, including disease resistance, or reproduction-related traits because growth-related traits with moderate to moderate-high heritabilities often respond well to selection in aquaculture organisms following traditional approaches (Wringe *et al*. 2010; Sánchez-Molano *et al*. 2011; Sauvage *et al*. 2012b; Chiasson *et al*. 2013). In order to integrate any new trait into a selection program, genetic parameters including heritability estimates and correlations of each trait with body weight need to be investigated using both approaches. It is important, however, that the number of traits to be added to new breeding

objectives for GFP need to be restricted, to keep the breeding program focussed and the cost of the program in balance with increases in productivity achieved.

The ultimate aim of genomic-assisted breeding is to increase the rate of genetic gains in a variety of target environments, in less time and at lower cost compared with conventional selection, based exclusively on expressed phenotypes. Nonetheless, it is critical to remember that even with the availability of the best genotyping resources, genomic-assisted breeding may not be successful in the absence of high quality phenotypic data (Varshney *et al*. 2012). Taking advantage of genome information and new technologies combined with availability of quality phenotypic data generated from conventional selections can increase the rate of progress of genetic improvement of GFP brood stocks.

5.7.3. Sequencing resources for genomic selection to advance GFP stock improvement

While limited genomic or transcriptomic sequences generated from the current study can provide base information for MAS in GFP, it will not be sufficient to develop and apply GS methodology. In order to hasten the application of GS multiple resources are required including a large number of markers, linkage/physical maps, BAC libraries, and bioinformatics skills, which requires cost, time and effort. Completion of a whole reference genome sequence would provide the ultimate roadmap for future breeding programs but full genome sequencing of a large eukaryotic genome is not an easy task and is still a costly exercise so currently it is not feasible for GFP. As an example in penaeid shrimps, a large number of ESTs, differential gene expression experiments, microarray studies, and RNAi-based experimental data have been compiled to develop shrimp genetic and genomic resources that can further contribute to advancing biotechnological applications in aquaculture (Robalino *et al*. 2012). While genomic data for marine shrimps has been expanding rapidly, the same is not true for freshwater prawns used in culture. New technological developments, however, including NGSTs and automated high-throughput SNP genotyping technologies have largely eliminated the economic hurdles associated with obtaining genomic information in a cost-effective way for aquatic non-model species (Ekblom & Galindo 2011) and these developments will help to create the opportunity for genomics to make more significant impacts on productivity of aquaculture species in general in the future.

While the current study suggested a new paradigm of NGST application in aquatic species generating partial transcriptome data, there are several major obstacles to overcome in order to achieve the ultimate goal of GS application. As a first perspective for GFP, comprehensive transcriptome or EST sequences from diverse tissue samples using NGSTs can provide a platform for identification and characterisation of full-length cDNAs that will be important for ongoing studies of crustacean gene duplication and gene family structure. Second, when these sequences are available, future studies should focus on identifying expressed genes (i.e. characterisation and microarray) involved in various biological processes because to date this approach has not been trialled in GFP, mainly due to limited availability of ESTs. They can aid not only in shedding light on various fundamental processes but also in annotation of GFP genome sequences via comparisons of GO composition between full-length cDNAs and those from other *Macrobrachium* and/or crustacean species. In addition, identifying the genetic and environmental factors that control variation and elucidating the biochemical and physiological pathways that connect these factors to the phenotype can provide a detailed understanding of trait biology (i.e. quantitative traits) and would aid future genomics-assisted breeding approaches in GFP (i.e. favourable correlations with growth rate). Eventually, data generated from these approaches will provide a complete suite of gene-based markers (i.e. Type I markers) for identification of genes and alleles responsible for production traits including growth, disease resistance, feed efficiency and age at sexual maturation. The approach is likely to be integrated into many practical breeding programs in the near future with further advances and the maturation of relevant theory, which have already been successful in Atlantic cod (Star *et al*. 2011) and Pacific oyster (Zhang *et al*. 2012) or are on-going projects in Atlantic salmon (Davidson *et al*. 2010) and catfish (Liu 2011).

5.7.4. Candidate gene approach

Considering most economic phenotypes in stock improvement programs are quantitative traits influenced by multi-gene interactions, identifying key genes is an essential step. This is not easy, however, when only limited genetic information is available. This is why CGA was employed here based on *a priori* hypotheses about each gene's potential physiological roles. This approach has been widely acknowledged to take advantage of both the increased statistical efficiency of association analysis with quantitative traits and the biological

understanding of the phenotype, tissues, genes and proteins that are likely to be involved in the traits (Tabor *et al*. 2002). In spite of their promise, CGAs have been subject to two important criticisms, lack of replication and thoroughness (Tabor *et al*. 2002). In order to overcome these issues, more diverse samples and breeding lines will be required to be screened under a variety of environmental conditions to provide robust evidence for the physiological roles of particular genes for each phenotypic trait. In the case of GFP, only adult female GFP samples were screened here because growth performance is confounded by social factors in male GFP (Thanh *et al*. 2009, 2010) and by developmental stage (i.e. moulting). Therefore, testing the SNPs and genes identified here in male and in various GFP developmental stage samples will be important to understand selection signatures between genders and across ages. Furthermore, the current study was screened only for growth (EBV) so expanding to other phenotypic traits will be important to provide better biological understanding between genes and quantitative traits. Gene interactions, however, estimated from widely spaced genetic markers that do not cover the target organisms' genome adequately can produce misleading results.

5.7.5. Linkage map and reproductive traits

To my knowledge, no attempt has been made yet to construct a linkage map for GFP. Time and cost are the main reasons why this has not yet occurred. Restriction-site associated DNA (RAD) genotyping using NGSTs, however, has recently become a feasible alternative approach for identifying thousands of informative SNPs even in non-model species where little or no previous genome data are available (Davey *et al*. 2011; Hohenhohe *et al*. 2011; Wang *et al*. 2012a). RAD has also provided a powerful system capable of rapidly generating chromosome specific data even for detection of putative sex chromosomes and reproductive QTLs in aquatic species (Baxter *et al*. 2011; Anderson *et al*. 2012; Sauvage *et al*. 2012a). Interest in the genetics of sex determination, in particular sexual dimorphism for growth rate in aquaculture, is motivated by practical and commercial implications for production of monosex progeny for use in stock improvement in addition to fundamental scientific curiosity. Indeed, the enhanced yield achievable from monosex culture has been shown in case studies in shrimps where females were shown to grow faster and reach higher weights than males of equivalent age (Argue *et al*. 2002; Pérez-Rostro & Ibaraa 2003; Gopal *et al*. 2010). In GFP, in contrast, males grow faster and reach larger body sizes, so culturing all-male GFP cohorts

can result in significantly higher yields over a shorter culture period and provide improved financial returns than that achieved from mixed-sex culture or all-female culture (Sagi *et al*. 1986; Cohen *et al*. 1988).

Another case study of *M. rosenbergii* reported about a 60% increase in net income to growers when they cultured a hand-segregated, all-male population and discarded females (Nair *et al*. 2006). For the Vietnamese GFP culture stocks used here, male GFP generally reached significantly larger size and heavier body weight (approximately 93%) compared with females (Dinh *et al*. 2013). Due to the potential for better financial returns to farmers, a number of attempts have been made to sex juvenile GFP manually and to use biotechnology (e.g. manipulating insulin-like androgenic gland hormone) at early developmental stages to produce all-male GFP culture lines (Sagi & Aflalo 2005; Aflalo *et al*. 2006; Aflalo *et al*. 2012; Ventura *et al*. 2012; Ventura & Sagi 2012). Aflalo *et al*. (2012) recently suggested a combined approach, with selective breeding used to produce a fast growing line combined with neo-female technology to produce fast-growth, all-male cohorts for culture. In the current study, growth rate in the Vietnamese GFP strain showed a cumulative improvement of approximately 21% (Dinh *et al*. 2013), so potentially if neo-female technology was applied to the improved strain there, the all-male stock that would result should have an even better growth advantage than the improved line resulting from both selection and sex-related (all male) phenotypes.

Sex determination can result from either genetic and/or environmental factors and to date, farming single sex cohorts has been achieved either through manual sexing, direct hormonal sex reversal or hybridization some of which often have significant disadvantages in their application including high inbreeding rates that could result from using neo-females and males (or vice versa) from a single strain (Liu *et al*. 2010; Aflalo *et al*. 2012). Thus, utilising genome-based scale data including a high density linkage map developed by RAD sequencing provides an option for addressing this issue in GFP, because RADs could be used to reveal the underlying mechanism for sex determination or allow discrimination of loci responsible. Once sex reversal has been perfected, all-male progeny could result in significant increases in commercial returns with high production that will eventually accelerate future GFP GS approaches to stock improvement.

5.7.6. Biological pathway mapping in non-model species

Dissecting the genetic control of complex trait variation (i.e. growth) remains very challenging, despite many advances in technologies. One approach has been to study intermediate –omics data (e.g. transcriptome) as it can be useful for understanding the genetic architecture of complex traits and for bridging the genotype-phenotype gap (Ehsani *et al*. 2012). Growth in crustaceans is a complex process where interactions occur between genes and key developmental processes (e.g. moulting) at each level of biological organization from cell to organism. While the current transcriptomic dataset has provided promising baseline data for GFP, it is essentially fragmented and provides only scattered evidence for loci that affect target traits. This can be resolved via functional mapping, systems mapping and network mapping, which allow a comprehensive picture to be developed of the genetic landscape of complex phenotypes that underlie economically important traits like growth (Wang *et al*. 2012b), from a genome-wide perspective. This approach has proven very powerful for addressing questions about the specific roles and functions of genes in livestock species (Kemper & Goddard 2012).

Unlike livestock, this advance is more complicated in crustaceans because moulting, sex determination, and some environmental factors and their interaction combine to affect growth traits. This is a highly complex process that requires appropriate coordination and integration. Loci involved in moulting and sexual dimorphism can affect the genetic architecture of growth traits in crustaceans though most of these loci may have small effects. Occasionally, however, loci may be present with moderate-to-large individual effects segregating due to recent selection for the mutant allele through a biological signalling pathway (Kemper & Goddard 2012). Epistatic and pleiotropic effects have long been recognized as fundamentally important in understanding the structure, function, and evolutionary dynamics of biological systems (Xu *et al*. 2011). Yet, little is known about how their impacts affect specific traits even in model species (e.g. cattle). Despite limited information, a number of attempts have been made in crustaceans to elucidate the hormones and glands responsible for the process of enzyme cascades during sexual differentiation (Ventura *et al*. 2012; Ventura & Sagi 2012) and moulting (Nakatsuji *et al*. 2009; Chang & Mykles 2011) by filling in the gaps for genetic interactions in the key pathways. More recently, availability of detailed transcriptomic data from crustacean studies exemplifies the applicability of a high-throughput sequencing approach for studying complex traits, for example metamorphosis, spermatozoa conservation, transport, maturation and capacitation

(He *et al*. 2013; Ventura *et al*. 2013). It is unknown, however, whether these interactions are linked in a single signalling pathway or are components of independent pathways activated by different factors in crustaceans.

In reality, developing a comprehensive picture for interactive signalling pathways for growth in GFP (i.e. metabolic processes, biosynthesis of secondary metabolites and regulatory pathways) is not an easy task. Filling in the gaps, however, via mapping and understanding of gene interaction networks among moulting, sex reversal, growth, environmental factors and their interaction with growth traits is feasible as has been done for sex determination and moulting pathways elsewhere (Chang & Mykles 2011; Ventura & Sagi 2012; He *et al*. 2013; Ventura *et al*. 2013). For this, researchers have to 1) generate vast genomic and RNA-Seq data from a variety of tissue and developmental stage samples using NGST, 2) conduct a large number of differential gene expression analyses (e.g. microarray and SNPchips) to unravel the biological functions of each peptide under variable experimental conditions and tissues, 3) silence target gene expression using RNA-i techniques in different tissues and conditions as functional assays, 4) analyse accurate GO annotations, 5) start to sort out priority pathways to map depending on research objectives (e.g. apoptosis, cardiac muscle contraction, cell cycle, ecdysteroid synthesis, ErbB signalling, glutamatergic synapse, glycolysis/gluconeogenesis, GnRH signalling, insect hormone biosynthesis, insulin signalling, MAPK signalling, TGF-beta signalling, starch and sucrose metabolism, and vascular smooth muscle contraction), 6) determine the co-occurrence and co-absence of genes/proteins in animated KEGG pathway maps, 7) compare functional compositions within and among the dataset, 8) analyse the statistically enriched pathways across datasets, 9) build the pathway-to-pathway networks for each dataset, 10) explore potential interaction/reaction paths between pathways, 11) identify common pathway-pathway networks across the datasets, and 12) apply the identified genes (i.e. RNA-i) to diverse species to understand key roles.

A gene set-based approach is an important step towards reconstruction of signalling pathway structures from molecular profiling data. Since pathways are functional units that result from the interplay of interacting genes, RNAs, proteins and small molecules, mapping genes or proteins into pathway models can provide deep insights into their functions and interactions in an organism. Given the strategy, a major breakthrough would be to predict and interpret the effect of mutational and biochemical changes during crustacean growth and understand how this signal is transmitted spatially (among tissues) and temporally (spanning

development). Pathway-based approaches using prior biological knowledge on gene function have facilitated more powerful analysis of GWAS datasets (Wang *et al*. 2010a) so combining GWAS methodology with current strategies may be useful for understanding growth in crustacean species.

5.7.7. Next-generation sequencing, bioinformatics and genomic databases

Effective management of wild and captive GFP populations is in general impaired by limited knowledge about different aspects of this species' biology. Recently, research on aquatic species, has increased significantly, and 'omics' technologies have been applied to better characterise reproduction, development, nutrition, immunity and toxicology in these species (Cerdà *et al*. 2010; Jiang *et al*. 2011; Liu 2011; McAndrew & Napier 2011). As a result of recent significant reduction in sequencing costs, NGST approaches have become a powerful method for production of extensive genomic and EST resources for finding desirable traits in the genome of target species. These technologies are also powerful tools that allow efficient gene discovery and gene identification and permit resolution of several genetic and molecular regulation of biological processes in stock improvement studies that can pinpoint the signals that control gene function for unique characteristics as well as assisting the resolution of evolutionary lineages at the molecular level (Rothschild & Plastow 2008; Dheilly *et al*. 2011; Houston *et al*. 2012; Morrell *et al*. 2012), and therefore, they have great relevance for GFP. The advancement of NGSTs, however, has resulted in the generation of vast amounts of biological data which has fuelled the need for computational systems and techniques to manage large datasets as well as for improved data analysis methods (Quinn *et al*. 2012). On-going, bioinformatics should include the development of computational tools for data management, data manipulation and data mining for aquaculture species that mainly involve non-model species.

The rapid evolution of analysis and processing methods is indicative of the community effort to provide effective tools for understanding genomic data (Clarke *et al*. 2012). Large-scale data generation and analysis projects can benefit from an organised and centralised data-management process by providing necessary support and infrastructure to the public (Washington *et al*. 2011; Clark *et al*. 2012; Rosenbllom *et al*. 2012). The concept of 'open data', or public accessibility should be adopted online to prevent redundant work, to enable research groups to build from the work of others, and to allow the potential for cross-

species comparisons (Quinn *et al*. 2012). Eventually, availability of large EST resources for most model species as well as a number of agriculturally important species could allow efficient gene discovery and gene identification in GFP simply from online public databases. Rapid progress can be made via comparative genome analysis in understanding structural, organizational, and functional properties of the genome of this species that can still expedite stock improvement objectives in GFP. Five great challenges lie ahead in GFP studies before this approach can be replicated; 1) limited breeding objectives with few phenotypic traits exposed to selection, 2) difficulties with dissemination of the improved stocks, 3) little genomic information and the analytical challenges of genomics, 4) effective data management and distribution to users, and 5) limited research budget issues. International collaboration, however, can be an effective way to overcome these problems. In conjunction with traditional breeding methods, sequencing the complete genome of GFP could have a huge impact on the culture industry and also explore more fundamental scientific questions about growth in crustaceans.

5.8. Conclusions

The current study has provided the first transcriptome dataset for GFP and has the potential to promote development of MAS/GS in future giant freshwater prawn breeding programs. While samples used in this study came from only a single improved culture line from Vietnam, the overall results were promising in identifying genes with potential roles in growth and growth-related traits in GFP. In particular, NGST was very effective at generating a transcriptomic dataset for GFP where there was no prior genomic information. While the current study focused on finding and evaluating growth-related genes in individual growth performance in Vietnamese GFP stocks, the *M. rosenbergii* transcriptome data generated from muscle, ovary and testis tissue samples should provide an invaluable resource in the future for gene discovery in this species. These data will eventually instruct future functional studies to manipulate or select for genes that influence growth and reproduction that should find practical application in future GFP stock improvement programs. It is obvious that the amount and quality of data available will grow rapidly for genomic analysis of GFP using NGSTs, the only question is from where and which genes we should look at first. As an effort to stimulate increased research on the identification of growth-related genes in crustacean species, three strategies are suggested here. The information in concert can provide a

foundation for increasing the rate at which knowledge about key genes affecting growth traits in crustacean species is gained. While different SNP allele distributions, violation of HWE and significant LD in CGSM pairs were observed between the two growth performance groups in Vietnamese GFP, these results will require detailed confirmation as to whether the source of these polymorphisms are selection signatures for a particular trait or are simply random effects as a result of chance.

The current project was also the first attempt to develop a large set of CGSMs for assessing correlations between polymorphisms in growth-related candidate genes with individual growth performance phenotype in female GFP. A number of SNPs in several candidate genes showed significant and/or marginal associations with EBVs showing measurable heritability. Multi-SNPs involvement in multi-genes identified here could be involved in influencing growth traits in Vietnamese female GFP populations suggesting a possible epistatic and/or pleiotropic effect. A whole genome sequence can be viewed as the ultimate map of the genome, allowing researchers to mine the genes and genetic variation associated with important production traits to improve stocks used in culture. Until this is available for GFP, the current dataset potentially provides a fundamental resource for application in future GFP stock improvement programs and may also provide a model for how to adapt a MAS/GS approach to the GFP culture industry and even potentially for crustaceans in general. As has been done for other aquaculture species, organising an international collaboration to characterise the complete GFP genome could be the best solution to allow rapid identification of candidate genes responsible for production and performance traits in the future.

REFERENCES

Aflalo ED, Hoang TTT, Nguyen VH, Lam Q, Nguyen DM, Trinh QS *et al*. (2006) A novel two-step procedure for mass production of all-male populations of the giant freshwater prawn *Macrobrachium rosenbergii*. *Aquaculture* **256**: 468–478.

Aflalo ED, Raju DVSN, Bommi NA, Verghese JT, Samraj TYC, Hulata G *et al*. (2012) Toward a sustainable production of genetically improved all-male prawn (*Macrobrachium rosenbergii*): Evaluation of production traits and obtaining neo-females in three Indian strains. *Aquaculture* **338-341**: 197–207.

Anderson JL, Marí AR, Braasch I, Amores A, Hohenlohe P, Batzel P *et al*. (2012) Multiple sex-associated regions and a putative sex chromosome in zebrafish revealed by RAD mapping and population genomics. *PLoS One* **7**: e40701.

Argue BJ, Arce SM, Lotz JM, Moss SM (2002) Selective breeding of Pacific white shrimp (*Litopenaeus vannamei*) for growth and resistance to Taura Syndrome Virus. *Aquaculture* **204**: 447–460.

Asusens ACJ, Carlos SHJ, Arturo FCJ, Genaro DP (2012) The effects of eyestalk ablation on the reproductive and immune function of female *Macrobrachium americanum*. *Aquaculture Research and Development* **3**: 1–7.

Baranski M, Moen T, Våge DI (2010) Mapping of quantitative trait loci for flesh colour and growth traits in Atlantic salmon (*Salmo salar*). *Genetics Selection Evolution* **42**: 17.

Bastiaansen JWM, Coster A, Calus MPL, van Arendonk JAM, Bovenhuis H (2012) Long-term response to genomic selection: effects of estimation method and reference population structure for different genetic architectures. *Genetics Selection Evolution* **44**: 3.

Baxter SW, Davey JW, Johnston JS, Shelton AM, Heckel DG, Jiggins CD *et al*. (2011) Linkage mapping and comparative genomics using next-generation RAD sequencing of a non-model organism. *PLoS One* **6**: e19315.

Beuzen ND, Stear JL, Chang KC (2000) Molecular markers and their use in animal breeding. *Veterinary Journal* **160**: 42–52.

Bonami JB, Widada JS (2011) Viral diseases of the giant fresh water prawn *Macrobrachium rosenbergii*: A review. *Journal of Invertebrate Pathology* **106**: 131–142.

Börner V, Reinsch N (2012) Optimising multistage dairy cattle breeding schemes including genomic selection using decorrelated or optimum selection indices. *Genetics Selection Evolution* **44**: 1.

Bouck A, Vision T (2007) The molecular ecologist's guide to expressed sequence tags. *Molecular Ecology* **16**: 907–924.

Brady P, Elizur A, Williams R, Cummins SF, Knibb W (2012) Gene expression profiling of the cephalothorax and eyestalk in *Penaeus monodo* during ovarian maturation. *International Journal of Biological Sciences* **8**: 328–343.

Calus MPL, de Haas Y, Pszczola M, Veerkamp RF (2012) Predicted accuracy of an response to genomic selection for new traits in dairy cattle. *Animal* **Online Version**.

Cerdà J, Douglas S, Reith M (2010) Genomics resources for flatfish research and their application. *Journal of Fish Biology* **77**: 1045–1070.

Chang ES, Mykles DL (2011) Regulation of crustacean molting: A review and our perspectives. *General and Comparative Endocrinology* **172**: 323–330.

Chen R, Irwin DM, Zhang YP (2012) Differences in selection drive olfactory receptor genes in different direction in dogs and wolf. *Molecular Biology and Evolution* **Online version**.

Chiasson MA, Quinton CD, Danzmann RG, Ferguson MM (2013) Comparative analysis of genetic parameters and quantitative trait loci for growth traits in Fraser strain Arctic charr (*Salvelinus alpinus*) reared in freshwater and brackish water environments. *Journal of Animal Science* **91**: 2047–2056.

Ciobanu DC, Bastiaansen JWM, Magrin J, Jiang DH, Yu N, Geiger B *et al*. (2010) A major SNP resource for dissection of phenotypic and genetic variation in Pacific white shrimp (*Litopenaeus vannamei*). *Animal Genetics* **41**: 39–47.

Clarke L, Zheng-Bradley X, Smith R, Kulesha E, Xiao C, Toneva I *et al*. (2012) The 1000 genome project: data management and community access. *Nature Methods* **9**: 459–462.

Cleveland MA, Hickey JM, Forni S (2012) A common dataset for genomic analysis of livestock populations. *Genes Genomes Genetics* **2**: 429–435.

Cohen D, Sagi A, Ráanan Z, Zohar G (1988) The production of *Macrobrachium rosenbergii* in monosex population: iii. Yield characteristics under intensive monoculture conditions in earthen ponds. *Israelei Journal of Aquaculture* **40**: 57–63.

Colbourne JK, Pfrender ME, Gilbert D, Thomas WK, Tucker A, Oakley TH *et al*. (2011) The ecoresponsive genome of *Daphnia pulex*. *Science* **331**: 555–561.

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaster ML (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics* **12**: 499–510.

Davidson WS, Koop BF, Jones SJM, Iturra P, Vidal R, Maass A *et al*. (2011) Sequencing the genome of the Atlantic salmon (*Salmo salar*). *Genome Biology* **11**: 403.

Dechklar M, Udomkit A, Panyim S (2008) Characterization of argonaute cDNA from *Penaeus monodon* and implication of its role in RNA interference. *Biochemical and Biophysical Research Communications* **367**: 768–774.

Dheilly NM, Lelong C, Huvet A, Favrel P (2011) Development of a Pacific oyster (*Crassostrea gigas*) 31,918-feature microarray: identification of reference genes and tissue-enriched expression patterns. *BMC Genomics* **12**: 468.

Dinh H, Coman G, Hurwood DA, Mather PB (2012) Experimental assessment of the utility of visible implant elastomer tags in a stock improvement programme for giant freshwater prawn (*Macrobrachium rosenbergii*) in Vietnam. *Aquaculture Research* **43**: 1471–1479.

Dinh H, Vu NT, Nguyen NH, Ponzoni RW, Hurwood DA, Mather PB (2013) Quantitative genetic parameter estimates for body and carcass weight traits in a cultured stock of giant freshwater prawn (*Macrobrachium rosenbergii*) selected for harvest weight in Vietnam. *Aquaculture*, **404-405**: 122–129.

Eggen A (2012) The development and application of genomic selection as a new breeding paradigm. *Animal Frontiers* **2**: 10–15.

Ehsani A, Sørensen P, Pomp D, Allan M, Janss L (2012) Inferring genetic architecture of complex traits using Bayesian integrative analysis of genome and transcriptome data. *BMC Genomics* **13**: 456

Ekblom R, Galindo J (2011) Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity* **107**: 1–15.

Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics (Fourth ed.): Longman, Essex CM20 2JE, England.

Gitterle T, Salte R, Gjerde B, Cock J, Johansen H, Salazar M *et al*. (2005) Genetic (co)variation in resistance to White Spot Syndrome Virus (WSSV) and harvest weight in *Penaeus* (*Litopenaeus*) *vannamei*. *Aquaculture* **246**: 139–149.

Gjedrem T (2000) Genetic improvement of cold-water fish species. *Aquaculture Research* **31**: 25–33.

Gjedrem T (2012) Genetic improvement for the development of efficient global aquaculture: A personal opinion review. *Aquaculture* **344-349**: 12–22.

Gjedrem T, Robinson N, Rye M (2012) The importance of selective breeding in aquaculture to meet future demands for animal protein: A review. *Aquaculture* **350-353**: 117–129.

Goddard ME, Hayes BJ (2007) Genomic selection. *Journal of Animal Breeding and Genetics* **124**: 323–330.

Goddard ME, Hayes BJ (2009) Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reveiws Genetics* **10**: 381–391.

Gopal C, Gopikrishna G, Krishna G, Jahageerdar SS, Rye M, Hayes BJ, *et al*. (2010) Weight and time of onset of female-superior sexual dimorphism in pond reared *Penaeus monodon*. *Aquaculture* **300**: 237–239.

Gruber JD, Vogel K, Kalay G, Wittkopp PJ (2012) Contrasting properties of gene-specific regulatory, coding, and copy number mutations in Saccharomyces cerevisiae: Frequency, effects, and dominance. *PLoS Genetics* **8**: e1002497.

Gupta PK, Langridge P, Mir RR (2010) Marker-assisted wheat breeding: present status and future possibilities. *Molecular Breeding* **26**: 145–161.

Hamblin MT, Buckler ES, Jannink JL (2011) Population genetics of genomics-based crop improvement methods. *Trends in Genetics* **27**: 98–106.

Hayes BJ, Lewin HA, Goddard ME (2013) The future of livestock breeding: genomic selection for efficiency, reduced emissions intensity, and adaptation. *Trends in Genetics* **29**: 206–214.

He L, Jiang H, Cao D, Liu L, Hu S, Wang Q (2013) Comparative transcriptome analysis of the accessory sex gland and testis from the Chinese mitten crab (Eriocheir sinensis). *PLoS One* **8**: e53915.

Heslot N, Yang HP, Sorrells ME, Jannink JL (2012) Genomic selection in plant breeding: a comparison of models. *Crop Science* **52**: 146-160.

Hirono I, Fagutao FF, Kondo H, Aoki T (2011) Uncovering the mechanisms of shrimp innate immune response by RNA interference. *Marine Biotechnology* **13**: 622–628.

Hohenlohe PA, Amish SJ, Catchen JM, Allendorf FW, Luikart G (2011) Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Molecular Ecology Resources* **11S1**: 117–122.

Hornett EA, Wheat CW (2012) Quantitative RNA-Seq analysis in non-model species: assessing transcriptome assemblies as a scaffold and the utility of evolutionary divergent genomic reference species. *BMC Genomics* **13**: 361.

Houston RD, Davey JW, Bishop SC, Lowe NR, Mota-Velasco JC, Hamilton A *et al*. (2012) Characterisation of QTL-linked and genome-wide restriction site-associated DNA (RAD) markers in farmed Atlantic salmon. *BMC Genomics* **13**: 244.

Hu X, Gao Y, Feng C, Liu Q, Wang X, Du Z *et al*. (2009) Advanced technologies for genomic analysis in farm animals and its application for QTL mapping. *Genetica* **136**: 371–386.

Jiang Y, Lu J, Peatman E, Kucuktas H, Liu S, Wang S *et al*. (2011) A pilot study for channel catfish whole genome sequencing and de novo assembly. *BMC Genomics* **12**: 629.

Jiang H, Xu L, Gu Z (2011) Growth of novel epistatic interactions by gene duplication. *Genome Biology and Evolution* **3**: 295–301.

Jiao Y, Zhao H, Ren L, Song W, Zeng B, Buo J *et al*. (2012) Genome-wide genetic changes during modern breeding of maize. *Nature Genetics* **44**: 812-815.

Kemper KE, Goddard ME (2012) Understanding and predicting complex traits: knowledge from cattle. *Human Molecular Genetics* **21**: R45–51.

Kim WJ, Jung HT, Gaffney PM (2011) Development of Type I genetic markers from expressed sequence tags in highly polymorphic species. *Marine Biotechnology* **86**: 263–272.

Kitcharoen N, Rungsin W, Koonawootrittriron S, Na-Nakorn U (2012) Heritability for growth traits in giant freshwater prawn, *Macrobrachium rosenbergii* (de Mann 1879) based on best linear unbiased prediction methodology. *Aquaculture Research* **43**: 19–25.

Krishnan P, Gireesh-Babu P, Rajendran KV, Chaudhari A (2009) RNA interference-based therapeutics for shrimp viral diseases. *Disease Aquatic Organisms* **13**: 127–132.

Li C, Riethoven JM, Ma L (2010) Exon-primed intron-crossing (EPIC) markers for non-model teleosts fishes. *BMC Evolutionary Biology* **10:** 90.

Liu ZH, Zhang YG, Wang DS (2010) Studies on feminization, sex determination, and differentiation of the Southern catfish, Silurus meridionalis: A review. *Fish Physiology and Biochemistry* **36**: 223-235.

Liu Z (2011) Development of genomic resources in support of sequeicng, assembly, and annotation of the catfish genome. *Comparative Biochemistry and Physiology D* **6**: 11–17.

Liu S, Zhou Z, Lu J, Sun F, Wang S, Liu H *et al*. (2011) Generation of genome-scale gene-associated SNPs in catfish for the construction of a high-density SNP array. *BMC Genomics* **12**: 53.

Loukovitis D, Sarropoulou E, Tsigenopoulos CS, Batargias C, Magoulas A, Apostolidis AP *et al*. (2011) Quantitative trait loci involved sex determination and body growth in the gilthead sea bream (*Sparus aurata* L.) through targeted genome scan. *PLoS One* **6**: e16599.

Luan S, Yang G, Wang J, Luo K, Zhang Y, Gao Q *et al*. (2012) Genetic parameters and response to selection for harvest body weight of the giant freshwater prawn *Macrobrachium rosenbergii*. *Aquaculture* **362-363**: 88–96.

Ma K, Qiu G, Feng J, Li J (2012) Transcriptome analysis of the oriental river prawn, Macrobrachium nipponense using 454 pyrosequencing for discovery of genes and markers. *PLoS One* **7**: e39727

McAndrew B, Napier J (2011) Application of genetics and genomics to aquaculture development: current and future directions. *Journal of Agricultural Science* **149**: 143–151.

McClelland EK, Naish KA (2010) Quantitative trait locus analysis of hatch timing, weight, length and growth rate in coho salmon, *Oncorhynchus kisutch*. *Heredity* **105**: 562–573.

McGuigan K (2006) Studying phenotypic evolution using multivariate quantitative genetics. *Molecular Ecology* **15**: 883–896.

Miller S (2010) Genetic improvement of beef cattle through opportunities in genomics. *Revista Brasileira de Zootecnia* **39S**: 247–255.

Morrell PL, Buckler ES, Ross-Ibarra J (2012) Crop genomics: advances and applications. *Nature Review Genetics* **13**: 85–96.

Nachman MW, Hoekstra HE, D'Agostino SL (2003) The genetic basis of adaptive melanism in pocket mice. *Proceedings of the National Academy of Sciences of the USA* **100**: 5286–5273.

Nair CM, Salin KR, Raju MS, Sebastian M (2006) Economic analysis of monosex culture of giant freshwater prawn (*Macrobrachium rosenbergii* De Man): a case study. *Aquaculture Research* **37**: 949–954.

Nakatsuji T, Lee CY, Watson RD (2009) Crustacean molting-inhibiting hormone: Structure, function, and cellular mode of action. *Comparative Biochemistry and Physiology* **152A**: 139–148.

Nakaya A, Isobe SC (2012) Will genomic selection be a practical method for plant breeding? *Annals of Botany* **Online Version**.

Nguyen NH, Ponzoni RW, Hamzah A, Yee HY, Abu-Bakar KR, Khaw HL (2010a) Genetics of flesh quality in fish. In the proceedings of the 8th world congress in genetics applied to livestock production. *1-6 August 2010 in Germany*.

Nguyen NH, Ponzoni RW, Yee HY, Abu-Bakar KR, Hamzah A, Khaw HL (2010b) Quantitative genetic basis of fatty acid composition in the GIFT strain of Nile tilapia (*Oreochromis niloticus*) selected for high growth. *Aquaculture* **309**: 66–74.

Nhan DT, Wille M, Hung LT, Sorgeloos P (2009) Comparison of reproductive performance and offspring quality of giant freshwater prawn (*Macrobrachium rosenbergii*) broodstock from different regions. *Aquaculture* **298**: 36–42.

Nhan DT, Wille M, Hung LT, Sorgeloos P (2010) Effects of larval stocking density and feeding regime on larval rearing of giant freshwater prawn (*Macrobrachium rosenbergii*). *Aquaculture* **300**: 80–86.

Ober U, Ayroles JF, Stone EA, Richards S, Zhu D, Gibbs RA *et al*. (2012) Using whole-genome sequence data to predict quantitative trait phenotypes in Drosophila melanogaster. *PLoS Genetics* **8**: e1002685.

Panphut W, Senapin S, Sriurairatana S, Withyachumanarnkul B, Flegel TW (2011) A novel integrase-containing element may interact with Laem-Singh virus (LSNV) to cause slow growth in giant tiger shrimp. *BMC Veterinary Research* **7**: 18.

Park S, Lehner B (2013) Epigenetic epistatic interactions constrain the evolution of gene expression. *Molecular Systems Biology* **9**: 645.

Pérez-Rostro CI, Ibarra AM (2003) Heritabilities and genetic correlations of size traits at harvest size in sexually dimorphic Pacific white shrimp (*Litopenaeus vannamei*) growth in two environments. *Aquaculture Research* **34**: 1079–1085.

Phillips PC, McGuigan KL (2006) Evolution of genetic variance-covariance structure, pp. 310–325 in Evolutionary Genetics: Concepts and case studies, edited by C. W. Fox and J. B. Wolf. Oxford University Press, Oxford.

Ponzoni RW, Nguyen NH, Khaw HL, Hamzah A, Bakar KRA, Yee HY (2011) Genetic improvement of Nile tilapia (*Oreochromis niloticus*) with special reference to the work conducted by the WorldFish Center with the GIFT strain. *Reviews in Aquaculture* **3**: 27–41.

Quinn NL, Gutierrez AP, Koop BF, Davidson WS (2012) Genomics and genome sequencing: benefits for finfish aquaculture. *Aquaculture*. Dr, Zainal Muchlisin (Ed.).

Reid DP, Szanto A, Glebe B, Danzmann RG, Ferguson MM (2005) QTL for body weight and condition factor in Atlantic salmon (*Salmo salar*): comparative analysis with rainbow trout (*Oncorhynchus mykiss*) and Arctic charr (*Salvelinus alpines*). *Heredity* **94**: 166–172.

Robalino J, Chapman RW, de la Vega E, O'Leary NA, Gorbach DM, Du ZQ *et al*. (2012) Advances in genomics and genetics of penaeid shrimp. *Aquaculture Biotechnology* (John Wiley & Sons, Ltd.).

Rosenbloom KR, Dreszer TR, Long JC, Malladi VS, Sloan CA, Raney BJ *et al*. (2012) ENCODE whole-genome data in the UCSC genome browser: update 2012. *Nucleic Acids Research* **40**: D912–917.

Rothschild MF, Plastow GS (2008) Impact of genomics on animal agriculture and opportunities for animal health. *Trends in Biotechnology* **26**: 21-25.

Rubin CJ, Zody MC, Eriksson J, Meadows JRS, Sherwood E, Webster MT *et al*. (2010) Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature* **464**: 587-591

Sagi A, Ráanan Z, Cohen D, Wax Y (1986) Production of *Macrobrachium rosenbergii* in monosex populations: Yield characteristics under intensive monoculture conditions in cages. *Aquaculture* **51**: 265–275.

Sagi A, Aflalo ED (2005) The androgenic gland and monosex culture of freshwater prawn *Macrobrachium rosenbergii* (De Man): a biotechnological perspective. *Aquaculture Research* **36**: 231–237.

Sánchez-Molano E, Cerna A, Toro MA, Bouza C, Hermida M, Pardo BG, Cabaleiro S *et al*. (2011) Detection of growth-related QTL in turbot (*Scophthalmus maximus*). *BMC Genomics* **12**: 473.

Sauvage C, Vagner M, Derôme, Audet C, Bernatchez L (2012a) Coding gene single nucleotide polymorphism mapping and quantitative trait loci detection for physiological reproductive traits in brook. *Genes Genomes Genetics* **2**: 379–392.

Sauvage C, Vagner M, Derôme, Audet C, Bernatchez L (2012b) Coding gene SNP mapping reveals QTL linked to growth and stress response in brook charr (*Salvelinus fontinalis*). *Genes Genomes Genetics* **2**: 707–720.

Springer NM (2013) Epigenetics and crop improvement. *Trends in Genetics* **29**: 241–247.

Star B, Nederbragti AJ, Jentoft S, Grimholt U, Malmstrøm M, Gregers TF *et al*. (2011) The genome sequence of Atlantic cod reveals a unique immune system. *Nature* **477**: 207–210.

Su J, Oanh DTH, Lyons RE, Leeton L, van Hulten MCW *et al*. (2008) A key gene of RNA interference pathway in the black tiger shrimp, *Penaeus monodo*: Identification and fuctional characterization of Dicer-1. *Fish and Shellfish Immunology* **24**: 223–233.

Tabor HK, Risch NJ, Myers RM (2002) Candidate-gene approach for studying complex genetic traits: practical considerations. *Nature Reviews Genetics* **3**: 391–397.

Taylor JF (2013) Implementation and accuracy of genomic selection. *Aquaculture* **Online version**.

Thanh NM, Ponzoni RW, Nguyen NH, Vu NT, Barnes A *et al*. (2009) Evaluation of growh performance in a diallel cross of three strains of giant freshwater prawn (*Macrobrachium rosenbergii*) in Vietnam. *Aquaculture* **287**: 75–83.

Thanh NM, Barnes AC, Mather PB, Yutao L, Lyons RE (2010) Single nucleotide polymorphisms in the actin and crustacean hyperglycemic hormone genes and their correlation with individual growth performance in giant freshwater prawn *Macrobrachium rosenbergii*. *Aquaculture* **301**: 7–15.

Thornton PK (2010) Livestock production: recent trends, future prospects. *Philosophical Transctions of the Royal Society B* **365**: 2853–2867.

Uawisetwathana U, Leelatanawit R, Klanchui A, Prommoon J, Klinbunga S, Karoonuthaisiri N (2011) Insights into eyestalk ablation mechanism to induce ovarian maturation in the black tiger shrimp. *PLoS One* **6**: e24427.

Upadyayula N, da Silva HS, Bohn MO, Rocheford TR (2006) Genetic and QTL analysis of maize tassel and ear inflorescence architecture. *Theoretical and Applied Genetics* **112**: 592–606.

Varshney RK, Nayak SN, May GD, Jackson SA (2009) Next-generatin sequencing technologies and their implications for crop genetics and breeding. *Trends in Biotechnology* **27**: 522–530.

Varshney RK, Ribaut JM, Buckler ES, Tuberosa R, Rafalski JA, Langridge P (2012) Can genomics boost productivity of orphan crops? *Nature Biotechnology* **30**: 1172–1176.

Ventura T, Rosen O, Sagi A (2011a) From the discovery of the crustacean androgenic gland to the insulin-like hormone in six decades. *General Comparative Endocrinology* **173**: 381–388.

Ventura T, Aflalo ED, Weil S, Kashkush K, Sagi A (2011b) Isolation and characterization of a female-sepcific DNA marker in the giant freshwater prawn *Macrobrachium rosenbergii*. *Heredity* **107**: 456–466.

Ventura T, Manor R, Aflalo ED, Weil S, Rosen O, Sagi A (2012) Timing sexual differentiation: full functional sex reversal achieved through silencing of a single insulin-like gene in the prawn, *Macrobrachium rosenbergii*. *Biology of Reproduction* **86**:90, 1–6.

Ventura T, Sagi A (2012) The insulin-like androgenic gland hormone in crustaceans: from a single gene silencing to a wide array of sexual manipulation-based biotechnologies. *Biotechnology Advances* **30**: 1543–1550.

Ventura T, Manor R, Aflalo ED, Chalifa-caspi V, Weil S, Sharabi O, Sagi A (2013) Post-embryonic transcriptomes of the prawn Macrobrachium rosenbergii: Multigenic succession through metamorphosis. *PLoS One* **8**: e55322.

Villanueva B, Fernández J, García-Cortés LA, Varona L, Daetwyler HD, Toro MA (2011) Accuracy of genome-wide evaluation for disease resistance in aquaculture breeding programs. *Journal of Animal Sciecne* **89**: 3433-3442.

vonHoldt BM, Pollinger JP, Lohmuller KE, Han E, Parker HG, Quignon P *et al*. (2010) Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. *Nature* **464**: 898-902.

Wang K, Li M, Hakonarson H (2010a) Analysing biological pathways in genome-wide association studies. *Nature Reviews Genetics* **11**: 843–854.

Wang S, Peatman E, Abernathy J, Waldbieser G, Lindquist E, Richardson P *et al*. (2010b). Assembly of 500,000 inter-specific catfish expressed sequence tags and large scale gene-associated marker development for whole genome association studies. *Genome Biology* **11**: R8.

Wang W, Wei Z, Lam TW, Wang J (2011) Next generation sequencing has lower sequence coverage and poorer SNP-detection capability in the regulatory regions. *Scientific Reports* **1**: 55.

Wang S, Meyer E, McKay JK, Matz MV (2012a) 2b-RAD: a simple and flexible method for genome-wide genotyping. *Nature Methods* **9**: 808–810.

Wang Z, Wang Y, Wang N, Wang J, Wang Z, Vallejos CE *et al*. (2012b) Towards a comprehensive picture of the genetic landscape of complex traits. *Briefings in Bioinformatics Advance* **Online Version**.

Washington NL, Stinson EO, Perry MD, Ruzanov P, Contrino S, Smith R *et al*. (2011) The modENCODE data coordination center: lessons in harvesting comprehensive experimental details. *Database* **Bar023**.

Wheeler MB, Monaco E, Bionaz M, Tanaka T (2010) The role of existing and emerging biotechnologies for livestock production: toward holism. *Acta Scientiae Veterinariae* **38S2**: 463–484.

Wringe BF, Devlin RH, Ferguson MM, Moghadam HK, Sakhrani D, Danzmann RG (2010) Growth-related quantitative trait loci in domestic and wild rainbow trout (*Oncorhynchus mykiss*). *BMC Genetics* **138**: 433–451.

Wu Q, Luo Y, Lu R, Lau N, Lai EC *et al*. (2010) Virus discovery by deep sequencing and assembly of virus-derived small silencing RNAs. *Proceedings of the National Academy of Sciences of the USA* **107**: 1606–1611.

Xu L, Jiang H, Chen H, Gu Z (2011) Genetic architecture of growth traits revealed by global epistatic interactions. *Genome Biology Evolution* **3**: 909–914.

Zara FJ, Gaeta HH, Costa TM, Toyama MH, Caetano FH (2013) The ovarian cycle histochemistry and its relationship with hepatopancreas weight in the blue crab *Callinectes danae* (Crustacea: Portunidae) *Acta Zoologica* **94**: 134–146.

Zhang G, Fang X, Guo X, Li L, Luo R, Xu F *et al*. (2012) The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* **490**: 49–54.