Transcriptomics of a Giant Freshwater Prawn (Macrobrachium rosenbergii): De Novo Assembly, Annotation and Marker Discovery

Hyungtaek Jung1*, Russell E. Lyons2, Hung Dinh1, David A. Hurwood1, Sean McWilliam2, Peter B. Mather1

1 Biogeosciences, Queensland University of Technology, Brisbane, Queensland, Australia, 2 Commonwealth Scientific and Industrial Research Organisation Livestock Industries, St Lucia, Queensland, 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.

Citation: Jung H, Lyons RE, Dinh H, Hurwood DA, McWilliam S, et al. (2011) Transcriptomics of a Giant Freshwater Prawn (Macrobrachium rosenbergii): De Novo Assembly, Annotation and Marker Discovery. PLoS ONE 6(12): e27938. doi:10.1371/journal.pone.0027938

Editor: John Parkinson, Hospital for Sick Children, Canada

Received September 15, 2011; Accepted October 28, 2011; Published December 8, 2011

Copyright: © 2011 Jung et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This research forms part of H. Jung’s Ph.D 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 P. Mather (QUT VC NPSG award). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: h7.jung@qut.edu.au

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 [1–4]. Due to its high value, research is now focusing on improving the growth performance of farmed M. rosenbergii [2–6]. 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 [7–10]. Whilst their application to cultured fish species has produced significant production gains, such gains are only beginning to be realized in penaeid species [11–15], 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 [16–18] and mitochondrial DNA sequence comparisons [19] have been used to examine the phylogeography of M. rosenbergii [20,21] sampled from Asia and northern Australia and genes potentially associated with pathogen defence responses [22–24] and sexual maturation traits [25] have also been identified. However, more genome-wide or transcriptome-wide datasets have yet to be generated as a basis for functional genomics approaches [26–29] 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 [26,33,31] and
exceptionally accurate [26–31]. Here it was used to characterize the transcriptome of *M. rosenbergii* using cDNA prepared from mRNA isolated from muscle, ovary and testis tissues. Expessed 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 are also reported.

**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 high 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 from 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 RNA later 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; [26]), flooded or rose gum (245 nt; [32]) or shore pine (306 nt at 3.6 x coverage; [33]). 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. Singleton ranges 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 <e–5) with proteins in the GenBank non-redundant (nt) 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 [14,15]. 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 (*Macrobrachium japonicas*), white leg shrimp (*Liocarcinus depurator*), 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.

| Dataset name | All          | Muscle       | Ovary        | Testis       |
|--------------|--------------|--------------|--------------|--------------|
| Total number of bases (Mb) | 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        | 1,004        |
| 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        |

Table 1. Summary of 454 pyrosequencing, assembly and analysis of *M. rosenbergii* transcriptomic sequences.

doi:10.1371/journal.pone.0027938.t001
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 [14,15]. 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 [34–36], 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 [26,37,38]. 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 [39–42]. 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 [11,14,15,43].

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 [44], 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 [45]. A total of 66 transcripts were involved in oxidative phosphoryla-
tomic sequences among three libraries. Putative sequence

doi:10.1371/journal.pone.0027938.g003

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 [14,15,28],
domains that dominated occur in RNA-binding proteins, protein
kinases and transcription factors (zinc finger domains) that are
essential for cellular processing functions including signal trans-
duction and transcription regulation, regulation of RNA stability
and translation control (RNA recognition motifs), innate immu-
nity, cell division, proliferation, apoptosis and cell differentiation
[53,54].

The most common DNA-binding motifs present in eukaryotic
and prokaryotic transcription factors [55] 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 [56], and can also bind to RNA and protein targets
[57]. A total of 112 nucleotide-binding a-b plait domains found in
RNA-binding domains from various ribonucleoproteins or in viral
DNA-binding domains [50,59] were predicted to exist among the
M. rosenbergii EST coding sequences. In addition, 108 Armadillo-
type fold and 94 Armadillo-like helical domains which form
structural domains consisting of a multi-helical fold comprised of 2
curved layers of a-helices [60], were predicted (Table 2).

Among M. rosenbergii EST coding sequences, 104 domains
containing WD40/YVTV repeat-like sequences, 90 domains
containing WD40-repeat sequences and 88 domains containing
WD40 repeat-like sequences were predicted (Table 2). These
domains are involved in a variety of functions ranging from signal
transduction and transcription regulation to cell cycle control and
apoptosis [61,62]. A total of 86 Ran GTPase families which are
involved in regulating GTP hydroases [63], contain GTP-binding
domains [64] and regulate receptor-mediated transport between
the nucleus and the cytoplasm [65,66] were also predicted, as were
84 immunoglobulin (Ig)-like fold domains (Table 2). Ig-like fold
domains are involved in a variety of functions including cell-cell
recognition, cell-surface receptors, muscle structure and the
immune system [67], and are often involved with protein-protein
interactions mediated by their a-helices as in other Ig-like domains
[67,68]. Other domains identified abundantly included Serpin
(some proteinase inhibitor) domains (n = 79) and NAD(P)-binding
domains (n = 72) (Table 2). 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 [69–71] and the clear demonstration of effective
RNAi-mediated knockdown of gene expression in shrimp [70].

Similar transcriptome analyses of other tissues including haemo-
cytes from the lymphoid organs for example that are primary
mediators of pathogen defence responses [14,15,72] might be
useful for identifying 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 Macrobrachium aquacul-
ture, 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 [14,15,72]. Thus genes mediating growth performance and
potentially of value in selective breeding programs await discovery.

Putative genes affecting muscle development and/or

The M. rosenbergii EST sequence database was mined for coding
sequences with domains potentially with muscle develop-
ment 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 character-
istics 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 [73,74]. Different actin isoforms have been identified in various crustaceans [40], and are likely to be involved in playing important roles in cytoskeletal structure, cell division and mobility, and muscle contraction [40–42]. The large super-family of myosin proteins interact with actin filaments by hydrolysing adenosine triphosphate that combine to form thick muscle filaments [75]. 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 [76]. Multiple MHC isoforms are expressed ubiquitously in all eukaryotic cells and they are the most abundant contractile protein present in skeletal muscle [77,78]. 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* [79]. 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 [80–82]. According to Perry et al. [83], 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 [84]. In striated muscle, tropomyosin mediates interactions between the troponin complex and actin to mediate muscle contraction [85]. 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 [86] 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 [39,41,87]. 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 [88,89].

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 2. Summary of top 20 domains predicted in *M. rosenbergii* sequences.

| IPR accession | Domain name            | Domain description                 | Occurrence |
|---------------|------------------------|------------------------------------|------------|
| IPR0000504    | RRM_dom                | RNA recognition motif domain        | 188        |
| IPR017442     | Ser/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         |

doi:10.1371/journal.pone.0027938.t002
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; A004047; 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                    | A002527 | 1587 |
| O-methyltransferase               | A004362 | 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; A008388; 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; A0007719 | 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.

doi:10.1371/journal.pone.0027938.t003

(Table 3): Amongst these, a putative cyclophilin was identified. Although cyclophilins possess diverse functions and have been linked to innate immunity [90,91] and testicular development [48], expression levels of cyclophilin-like proteins have also been found to be highly correlated with body-weight in the shrimp *P. monodon* [92].

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 [93]. Few FABPs have been identified in invertebrates [93,94], 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 [95]. 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 [96].

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 [97]. Previously, one LIM domain gene (ISL1) has been identified as a positional candidate for obesity and for controlling leptin organisation, cell fate determination and organ development [97]. 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 [99]. 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 [99].

Physiologically, O-methyltransferase (OMT) plays an important regulatory role in plant and animal growth, development, reproduction and immune response [99,101]. 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 [102–104]. 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 [103,105]. In some crustaceans, circulating titer and biosynthesis of MF appear to be correlated positively with maturation of the ovary [105,106]. MF has also been suggested to play a role in delaying onset of molting in larval crustaceans [102,106]. This evidence implicates MF in both crustacean growth and reproduction. Farnesoic acid O-methyltransferase (FAME-T; 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 [107,108]. Studies of crustacean FAME-T indicate that it may directly or indirectly (through MF) modulate reproduction and growth in crustaceans [109–112] 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 [112]. 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 is critical for cytoskeletal dynamics [113,114]. 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 [115]. Profilins have diverse roles in cellular processes, including membrane trafficking, small-GTPase signalling and nuclear activities, neurological diseases, and tumor formation [116–118]. 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 [119–121].

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 [122]. 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 [30,37]. Moreover, a total 658 simple sequence repeats (SSRs) or microsatellites comprising 61.85% dinucleotide repeats, 35.87% trinucleotide repeats and 2.28% tetra/penta/hexanucleotide 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 [123–125]. 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 [125–127]. 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.

![Figure 5. Distribution of putative single nucleotide polymorphisms (SNP) and indels in *M. rosenbergii* sequences.](doi:10.1371/journal.pone.0027938.g005)
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.

Methods

Tissue samples

*M. rosenbergii* with variable growth phenotypes were sampled from cohorts that were reared in a GFP stock improvement program in Vietnam [128]. 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 [6]. 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) [129] 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 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™ Kit (Ambion) according to the manufacturer’s protocol.

Library construction and 454 pyrosequencing

mRNA purified from pooled muscle, ovary, tests 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 [31]. 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_G10H9PT01AF0I7). 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. All M. rosenbergii EST sequences obtained were submitted to NCBI Sequence Read Archive under Accession no. SRP007672.

Annotation of mRNAs

BLASTx searches [130] 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 [131]. Total EST numbers in the Venn diagram quadrants excluding abundant ESTs for ribosomal proteins counted redundant ESTs only once. The Blast2GO software suite [132,133] was used to predict functions of individual ESTs, assign Gene Ontology terms [134,135], and to predict metabolic pathways using Kyoto Encyclopaedia of Genes and Genome (KEGG) [136,137]. 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 [138]. The numbers of contigs annotated with each GO term for each library were quantified using WEGO [139].

Identification of EST-SSR motifs and EST-SNPs

All EST sequences were searched for SSR motifs using the QDD program [140]. 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 [141] were used to design PCR primers flanking 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 [33,127,142] 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 [26,127]. In addition, only an overall transition vs transversion (Ts/Tv) ratio was calculated across the dataset.

Supporting Information

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

Table S2 Gene Ontology of M. rosenbergii contig sequences.

Table S3 KEGG summary of M. rosenbergii contig sequences.

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

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

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

The authors would like to acknowledge the help provided by the Australian Genome Research Facility (Matthew Tinning, Tim Bruexner, Rachel Kliese, and Adam Skarszewski) 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.

Author Contributions

Conceived and designed the experiments: HJ REL DAH PBM. Performed the experiments: HJ REL. Analyzed the data: HJ SM. Contributed reagents/materials/analysis tools: HJ REL PBM. Wrote the paper: HJ REL DAH PBM. Established breeding line: HD. Collected samples: HD.

References

1. FAO (2009) Fisheries Statistical Database, Global Aquaculture Production (Fisheries Global Information System, online query).
2. Thanh NM, Nguyen NH, Ponsoni RW, Vu NT, Barnes AC, et al. (2010) 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.
3. Thanh NM, Barnes AC, Mathur PB, Yotaka, 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: 7–13.
4. Schwartzes VS, Diana JS, Yi Y (2009) Social, economic, and production characteristics of giant freshwater prawn Macrobrachium rosenbergii culture in Thailand. Aquaculture 287: 120–127.
5. 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.
6. Thanh NM, Ponsoni 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.
7. Staelens J, Rombaut D, Vercauteren I, Argue B, Benzie J, et al. (2008) High-throughput gene discovery project: The Penaeid shrimp. Mar Biotechnol 13: 75–83.
8. Wu P, Qiu D, Chen Z, Zhang H, Zhang X, et al. (2009) Gene discovery from an ovary cDNA library of oriental river prawn Macrobrachium nipponense by EST annotation. Comp Biochem Physiol Part D Genomics Proteomics 4: 111–120.
9. Tassaanakajorn A, Klinbungra S, Paunglarp N, Rimphanitchayakit V, Udomkiat A, et al. (2006) Peneaus monodon gene discovery project: the generation of an EST collection and establishment of a database. Gene 384: 104–112.
10. Lyons RE, Diersen LM, Tan SH, Preston NP, Li Y (2007) Characterization of AFLP markers associated with growth in the Kuruma prawn, Macrobrachium japonicus, and identification of a candidate gene. Mar Biotechnol 9: 712–721.
11. Hamasaki K, Kitada S (2008) Potential of stock enhancement for decapods crustaceans. Rev Fisheries Sci 16: 164–174.
12. Leekitchairoenphon P, Taweemuang U, Falatangporphirit P, Kotsong H, Supasi T, et al. (2010) Predicted sub-populations in a marine shrimp protostome as revealed by combined EST and cDNA data from multiple Panaeid species. BMC Res Notes 3: 295.
13. 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. Mar Biotechnol 13: 608–621.
14. Chand Y, de Bruyn M, Mathur PM (2005) Microsatellite loci in the eastern form of the giant freshwater prawn (Macrobrachium rosenbergii). Mol Ecol Notes 5: 308–310.
15. Divu D, Khushiramani R, Malathi S, Karunasagar I, Karunasagar I (2008) Isolation, characterization and evaluation of microsatellite DNA markers in the giant freshwater prawn, Macrobrachium rosenbergii, from South India. Aquaculture 284: 281–284.
16. See LM, Tan SG, Hassan R, Siraj SS, Bhashu S (2009) Development of microsatellite markers from an enriched genomic library for the genetic analysis of the Malaysian giant freshwater prawn, Macrobrachium rosenbergii. Biochem Genet 47: 722–729.
17. Miller AD, Murphy NP, Burridge CP, Austin CM (2005) Complete mitochondrial DNA sequences of the Decapod Crustaceans Penaeus japonicus (Menippidae) and Macrobrachium rosenbergii (Palmaeroniidae). Mar Biotechnol 7: 339–349.
18. de Bruyn M, Wilson JG, Mathur PB (2004) Hind–Y line demarcates extensive genetic divergence between eastern and western forms of the giant freshwater prawn, Macrobrachium rosenbergii. Mol Phylogenet Evol 30: 251–257.
68. Potapov V, Moinard C, Vilella F, Vasson MP, Cybober L (2004) Metabolic pathways implicated in the kinetic impairment of muscle glutamine homocysteas in adult and old glucocorticoid-treated rats. Am J Physiol Endocrinol Metab 287: E671-676.

69. Høyer K, Møløe JL, Norman AM, Thompson NM, Davison M, et al. (2009) Prenatal induced changes in muscle structure and metabolic function facilitate exercise-induced obesity prevention. Endocrinology 150: 4135-4144.

70. Leser MP (2000) Oxidative stress in marine environments: Biochemistry and physiological ecology. Annu Rev Physiol 61: 253-278.

71. Gnaiger E, Ménez G, Haend S (2000) High phosphorylation efficiency and depression of uncoupled respiration in mitochondria under hypoxia. PNAS 97: 11080-11085.

72. Soonthornchai W, Rungrassamee W, Karoonuthaisiri N, Jarayabhand P, et al. (2008) Characterization of argonaute 1 in the coral, Pocillopora damicornis. Biochem Biophys Res Commun 367: 768-774.

73. Kabsch W, Vandekerckhove J (1992) Structure and function of actin. Annu Rev Biophys Bioeng 21: 49-76.

74. Vogt T (2010) Prenylpropanoid biosynthesis. Mol Plant 3: 2–20.

75. Teichmann SA, Chothia C (2000) Immunoglobulin superfamily proteins in Caenorhabditis elegans. FEBS Lett 486: 7-12.

76. Smith TF, Gaitatzes C, Saxena K, Neer EJ (1999) The WD repeat: a common functional module in a emerging superfamily of proteins. J Mol Biol 289: 235-255.

77. Kraemer B, Crittenden S, Gallegos M, Moulder G, Barstead R, et al. (1999) The LIM domain: regulation by association. Mech Dev 91: 5–17.

78. Strasser P, Gimona M, Moessler H, Herzog M, Small JV (1993) Mammalian cyclophilin: Identification and functional characterization of Dicer-1. Fish Shellfish Immunol 24: 223-233.

79. Kamimura MT, Meier KM, Cavalli RO, Laurino J, Maggioni R, et al. (2008) Characterization of growth-related genes in the southwestern Atlantic pink shrimp Farfantepenaeus paulensis (Perezo-Farfante 1967) through a modified DRTT-PCR protocol. Aquatic Rev 39: 200-204.

80. 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.

81. Sutherland LC, Rintala-Maki ND, White RD, Morin CD (2005) RNA binding proteins: A review. RNA 11: 1606–1611.

82. Abdel Rahman AM, Kamath S, Lopata AL, Heilveil RJ (2010) Analysis of the allergenic proteins in black prawn (Penaeus monodon) and characterization of the major allergens tropomyosin using mass spectrometry. Rapid Commun Mass Spectrom 24: 2462-2470.

83. 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. J Exp Biol 212: 673-683.

84. Barat-Houari M, Clement K, Vatin V, Dina C, Bonhomme G, et al. (2002) Postnatal induction of the transcription factor TAFl in muscle and implication of its role in RNA interference. Comm Soil Sci Plant Anal 33: 2203-2214.

85. Barat-Houari M, Clement K, Vatin V, Dina C, Bonhomme G, et al. (2002) Postnatal induction of the transcription factor TAFl in muscle and implication of its role in RNA interference. J Biochem 110: 54-59.

86. Schiaffino S, Reggiani C (1996) Molecular diversity of myofibrillar proteins: Gene regulation and function. Physiol Rev 76: 371-423.

87. DeNardis C, Ausoni S, Moretti P, Gozza L, Velleca M, et al. (1993) Type 2X- myosin heavy chain is coded by a muscle fiber type-specific and developmentally regulated gene. J Cell Biol 123: 821-835.

88. Jang HH, Leeber RL, Ryan AF (1996) Quantification of myosin heavy chain mRNA in somatic and brachial arch muscles using competitive PCR. Am J Physiol 275: 68-74.

89. Kamimura MT, Meier KM, Cavalli RO, Laurino J, Maggioni R, et al. (2008) Characterization of growth-related genes in the southwestern Atlantic pink shrimp Farfantepenaeus paulensis (Perezo-Farfante 1967) through a modified DRTT-PCR protocol. Aquatic Rev 39: 200-204.

90. Ibrahim RK, Bruneau A, Bantignies B (1998) Plant o-methyltransferase: The smooth muscle connection. Microsec Res Tech 58: 504-513.

91. Towers GJ (2007) The control of viral infection by tripartite motif proteins and cyclophilin A. Retrovirology 4: 40-50.

92. Barat-Houari M, Clement K, Vatin V, Dina C, Bonhomme G, et al. (2002) Postnatal induction of the transcription factor TAFl in muscle and implication of its role in RNA interference. J Biochem 110: 54-59.

93. Barat-Houari M, Clement K, Vatin V, Dina C, Bonhomme G, et al. (2002) Postnatal induction of the transcription factor TAFl in muscle and implication of its role in RNA interference. J Biochem 110: 54-59.

94. Barat-Houari M, Clement K, Vatin V, Dina C, Bonhomme G, et al. (2002) Postnatal induction of the transcription factor TAFl in muscle and implication of its role in RNA interference. J Biochem 110: 54-59.

95. Barat-Houari M, Clement K, Vatin V, Dina C, Bonhomme G, et al. (2002) Postnatal induction of the transcription factor TAFl in muscle and implication of its role in RNA interference. J Biochem 110: 54-59.

96. Barat-Houari M, Clement K, Vatin V, Dina C, Bonhomme G, et al. (2002) Postnatal induction of the transcription factor TAFl in muscle and implication of its role in RNA interference. J Biochem 110: 54-59.

97. Barat-Houari M, Clement K, Vatin V, Dina C, Bonhomme G, et al. (2002) Postnatal induction of the transcription factor TAFl in muscle and implication of its role in RNA interference. J Biochem 110: 54-59.

98. Barat-Houari M, Clement K, Vatin V, Dina C, Bonhomme G, et al. (2002) Postnatal induction of the transcription factor TAFl in muscle and implication of its role in RNA interference. J Biochem 110: 54-59.

99. Barat-Houari M, Clement K, Vatin V, Dina C, Bonhomme G, et al. (2002) Postnatal induction of the transcription factor TAFl in muscle and implication of its role in RNA interference. J Biochem 110: 54-59.
