

Universidade Federal de São Carlos

Centro de Ciências Biológicas e da Saúde

Programa de Pós Graduação em Genética Evolutiva e Biologia Molecular

Tese de Doutorado

**Identificação de genes candidatos relacionados a traços de desempenho em transcriptomas do
camarão marinho *Litopenaeus vannamei* (Penaeidae, Decapoda)**

Doutoranda: Camilla Alves Santos

Orientadora: Prof. Dra. Patrícia Domingues de Freitas

Co-orientadora: Prof. Dra. Sônia Cristina da Silva Andrade

São Carlos

2016

Universidade Federal de São Carlos
Centro de Ciências Biológicas e da Saúde
Programa de Pós Graduação em Genética Evolutiva e Biologia Molecular

Tese de Doutorado

**Identificação de genes candidatos relacionados a traços de desempenho em transcriptomas do
camarão marinho *Litopenaeus vannamei* (Penaeidae, Decapoda)**

Doutoranda: Camilla Alves Santos

Orientadora: Prof. Dra. Patrícia Domingues de Freitas

Co-orientadora: Prof. Dra. Sônia Cristina da Silva Andrade

Tese de doutorado apresentada ao Programa de Pós-Graduação em Genética Evolutiva e Biologia Molecular do Centro de Ciências Biológicas e da Saúde da Universidade Federal de São Carlos, como parte dos requisitos para obtenção do título de doutor em Genética Evolutiva e Biologia Molecular.

São Carlos

2016

Ficha catalográfica elaborada pelo DePT da Biblioteca Comunitária UFSCar
Processamento Técnico
com os dados fornecidos pelo(a) autor(a)

S237i Santos, Camilla Alves
Identificação de genes candidatos relacionados a traços de desempenho em transcriptomas do camarão marinho *Litopenaeus vannamei* (Penaeidae, Decapoda) / Camilla Alves Santos. -- São Carlos : UFSCar, 2017. 156 p.

Tese (Doutorado) -- Universidade Federal de São Carlos, 2016.

1. Genética. 2. Expressão Diferencial. 3. Polimorfismos. 4. Transcriptomas. I. Título.



UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Biológicas e da Saúde
Programa de Pós-Graduação em Genética Evolutiva e Biologia Molecular

Folha de Aprovação

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a Defesa de Tese de Doutorado da candidata Camilla Alves Santos, realizada em 28/11/2016:

Prof. Dra. Patrícia Domingues de Freitas
UFSCar

Prof. Dra. Sandra Regina Costa Maruyama
UFSCar

Prof. Dra. Luciane Maria Perazzolo
UFSC

Prof. Dr. Diogo Teruo Hashimoto
UNESP

Prof. Dra. Polyana Cristine Tizioto
EMBRAPA

*“O pessimista reclama do vento, o otimista
espera que ele mude, o realista ajusta as velas.”*

(Provérbio chinês)

“A persistência é o caminho do êxito.”

Charles Chaplin

Agradecimentos

Primeiramente, agradeço a Deus por tornar possível a realização deste trabalho e por todos os colaboradores que encontrei ao longo do caminho.

À minha família, por sempre acreditaram em mim e serem o meu alicerce.

À Universidade Federal de São Carlos (UFSCar), ao Programa de Pós Graduação em Genética Evolutiva e Biologia Molecular (PPGGEv) e ao Laboratório de Biodiversidade Molecular e Conservação (LABBMC) pelos meios que tornaram possível a realização desse trabalho.

À minha orientadora, Patrícia, por aceitar me orientar e tornar possível o meu doutoramento em uma área da genética em que sempre tive grande interesse. Agradeço pelo trabalho desenvolvido e por todas as oportunidades e desafios que vieram com ele. Agradeço também ao Prof. Pedro Galetti, pela colaboração sempre pertinente na elaboração de artigos científicos e no meu trabalho.

À minha co-orientadora, Sônia, cujo auxílio e conhecimentos foram essenciais para a conclusão dos resultados da parte de sequenciamento de próxima geração desse trabalho. Obrigada pela parceria e dedicação a este trabalho.

À Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP – Projeto 2012/13069-6), pelo período de bolsa e verba investida em meu aprimoramento profissional. Agradeço também pelo investimento concedido pelo Processo 2012/17322-8 que possibilitou, principalmente, o sequenciamento das amostras e construção das bibliotecas de cDNA.

Agradeço a todos os (muitos) colaboradores pelo auxílio durante o desenvolvimento do trabalho: Patrícia Tholon e John Rocha pelo apoio estatístico; a equipe do Centro de Genômica Funcional do Laboratório de Biotecnologia Animal da ESALQ, por me auxiliar no delineamento experimental e pela realização de um serviço de excelência; ao pessoal das empresas de

melhoramento genético Genearch e larvicultura Aquatec, de onde foram obtidos os camarões utilizados aqui. Obrigada Ana Carolina Guerrelhas, Flávio Farias e Ana Karina Teixeira pela confiança em nos ceder os camarões e pela parceria e auxílio. Agradeço também ao Prof. Reinaldo Brito pelo uso de equipamentos de seu laboratório durante o período das extrações de RNA. Sou grata também a todos os colegas de laboratório pelas conversas e sugestões que vieram a acrescentar neste trabalho.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro ao laboratório que contribuíram para a realização desse trabalho.

Agradeço por todos os desafios enfrentados durante o desenvolvimento desse trabalho que fizeram com que eu me aprimorasse e me tornasse uma pessoa diferente da que deu início a este projeto. Agradeço aos amigos que estiveram comigo em todos os momentos, agradáveis ou difíceis, pelos ensinamentos, palavras, carinho, atenção e amizade.

A todos que não foram citados aqui, mas que de alguma forma contribuíram para este trabalho.

Muito obrigada.

Sumário	
Resumo	- 1 -
Abstract.....	- 2 -
2. Objetivos e Hipóteses de Trabalho.....	- 5 -
3. Material e Métodos.....	- 7 -
3.1. Composição e Manutenção dos <i>Batches</i> na Genearch	- 7 -
3.2. Seleção das Famílias	- 9 -
4. Resultados e Discussão.....	- 12 -
Capítulo 1	- 13 -
RNA-seq as a powerful tool for penaeid shrimp genetic progress (Publicado)	- 13 -
Capítulo 2	- 14 -
Manuscript 1 - MINING GENES, SSRs AND SNPs WITHIN PROTEIN MAIN CHAIN FROM THE TRANSCRIPTOME OF <i>LITOPENAEUS VANNAMEI</i> MARINE SHRIMP	- 16 -
Abstract.....	- 16 -
1. Introduction.....	- 17 -
2. Material and Methods.....	- 18 -
2.1. cDNA libraries and sequencing.....	- 18 -
2.2. Trimming and assembly.....	- 18 -
2.3. Data mining for SSR and SNP	- 18 -
2.4. Transcriptome annotation.....	- 19 -
3. Results and Discussion.....	- 19 -
3.1. Transcriptome assembly and annotation	- 19 -
3.2. Mining SSR within coding regions	- 25 -
3.3. Mining SNP within coding regions	- 27 -
Final considerations	- 29 -
Manuscript 2 - IDENTIFICATION OF A RICH GENE SET RELATED TO STRESS AND ENERGY METABOLISM BY ANALYSIS OF EXPRESSED SEQUENCE TAGS (EST) FROM <i>LITOPENAEUS VANNAMEI</i> SHRIMP.....	- 34 -
Abstract.....	- 34 -
1. Introduction.....	- 35 -
2. Material and Methods.....	- 37 -
2.1. EST Database.....	- 37 -
2.2. Functional annotation and pathways establishment.....	- 37 -
3. Results and Discussion.....	- 38 -
3.1. Functional annotation and pathway identification.....	- 38 -
3.3. Stress protection.....	- 41 -

3.4. Energy production	- 44 -
3.5. Most expressed genes	- 45 -
Final considerations	- 46 -
Manuscript 3 – MAIN CHAIN PROTEIN POLYMORPHISMS STUDIES IN GENES FROM <i>LITOPENAEUS VANNAMEI</i> AND CROSS-SPECIES AMPLIFICATION EVALUATION IN OTHER PRAWN SPECIES	- 52 -
Abstract.....	- 52 -
Capítulo 3	- 62 -
Manuscrito 4 - ANÁLISE DO TRANSCRIPTOMA PARA IDENTIFICAÇÃO DE GENES DE PROTEÍNAS COM FUNÇÃO RELACIONADA A CRESCIMENTO E IMUNIDADE E DETECÇÃO DE POLIMORFISMOS DE BASE ÚNICA (SNPs) NO CAMARÃO <i>LITOPENAEUS VANNAMEI</i>.....	- 63 -
Resumo	- 63 -
1. Introdução	- 64 -
2. Material e Métodos.....	- 66 -
2.1. Obtenção das amostras	- 66 -
2.2. Extração de RNA, controle de qualidade e sequenciamento	- 67 -
2.3. Limpeza das sequências e montagem <i>de novo</i>	- 68 -
2.4. Detecção dos Polimorfismos de Base Única (SNPs)	- 69 -
3. Resultados e Discussão.....	- 70 -
3.1. Sequenciamento do transcriptoma e montagem <i>de novo</i>	- 70 -
3.2. Detecção dos SNPs.....	- 71 -
3.3. BlastX <i>hits</i> em espécies de artrópodes encontrados nas bases do SwissProt.....	- 73 -
3.4. Anotação funcional no GO e análise de vias metabólicas.....	- 79 -
Considerações finais.....	- 83 -
Capítulo 4	- 89 -
Manuscrito 5 - ANÁLISES DE RNA-SEQ REVELAM DIFERENÇAS NO PERFIL DE EXPRESSÃO DE GENES INDUZIDOS EM FAMÍLIAS DE ALTO E BAIXO CRESCIMENTO DO CAMARÃO <i>LITOPENAEUS VANNAMEI</i>. .	- 90 -
Resumo	- 90 -
1. Introdução	- 91 -
2. Material e Métodos.....	- 92 -
2.1. Coleta de amostras e preparação das bibliotecas	- 92 -
2.2. Genes Diferencialmente Expressos (DE)	- 92 -
2.2.1. Comparação entre as condições de alto e baixo crescimento	- 94 -
2.3. Enriquecimento funcional no Gene Ontology (GO).....	- 95 -
3. Resultados	- 95 -
3.1. Genes induzidos na comparação entre animais de Baixo X Alto crescimento	- 95 -

3.2. Enriquecimento no GO	- 99 -
4. Discussão	- 100 -
Considerações finais.....	- 105 -
Capítulo 5	- 108 -
Manuscrito 6 - ANÁLISE DO TRANSCRIPTOMA DE HEPATOPÂNCREAS E MÚSCULO DE CAMARÕES LITOPENAEUS VANNAMEI EXPOSTOS AO VÍRUS DA SÍNDROME DA MANCHA BRANCA (WSSV) ATRAVÉS DE RNA-SEQ.....	- 109 -
Resumo	- 109 -
1. Introdução	- 110 -
2. Material e Métodos.....	- 113 -
2.1. Coleta de amostras e preparação das bibliotecas	- 113 -
2.2. Genes Diferencialmente Expressos (DE)	- 113 -
2.2.1. Experimento 1: Comparação pareada entre tecidos	- 114 -
2.2.2. Experimento 2: Comparação entre tratamentos considerando o efeito do tecido	- 115 -
2.3. Enriquecimento funcional no Gene Ontology (GO).....	- 116 -
3. Resultados	- 116 -
3.1. Experimento 1: comparação entre hepatopâncreas e músculo	- 116 -
3.2. Experimento 2: comparação entre tratamentos considerando o efeito do tecido.....	- 122 -
4. Discussão	- 127 -
4.1. Principais proteínas imunes encontradas em crustáceos	- 128 -
4.1.1. Peptídeos antimicrobianos.....	- 128 -
4.1.2. Proteínas antimicrobianas.....	- 131 -
4.2. Cascatas de proteases	- 132 -
4.3. Lectinas e Proteínas de Reconhecimento Padrão	- 132 -
4.4. Outras proteínas relacionadas à imunidade.....	- 134 -
4.5. Conclusões sobre as comparações entre tecidos e tratamentos.....	- 135 -
4.6. Enriquecimento do GO	- 137 -
Considerações finais.....	- 139 -
5. Conclusões.....	- 146 -
6. Referências citadas na Introdução	- 146 -
Material Suplementar	- 148 -
Manuscrito 4	- 149 -
Figura S1.....	- 149 -
Manuscrito 5	- 149 -

Tabela S1	- 149 -
Tabela S2	- 150 -
Manuscrito 6	- 156 -
Tabela S1	- 156 -

Lista de Figuras

	Página
Figura 1: O camarão <i>Litopenaeus vannamei</i>	3
Figura 2: Esquema de abertura dos <i>batches</i> pela Genearch.	8
Figura 3: Sistema de seleção aplicado na Genearch.	11

Capítulo 2

Manuscript 1

Figure 1: Reads number observed in contigs assembly.....	20
Figure 2: Sequence similarity distribution related to similarity (green) and e-value (red) values found in blastX hits.	21
Figure 3: BlastX results from all-hits species distribution.	22
Figure 4: Blast results from top-hits species distribution.	22
Figure 5: Biological processes identified in Gene Ontology (GO) database for the 2.984 contigs annotated.	23
Figure 6: Molecular functions identified in Gene Ontology (GO) database for the 2.984 contigs annotated.	24
Figure 7: Cellular components identified in Gene Ontology (GO) database for the 2.984 contigs annotated.	24

Manuscript 2

Figure 1: Identity (%) (blue) and E-value (red) distribution for the 293 enzyme products.	38
Figure 2: Some of the most representative hits from GO functional annotation for the 293 enzymatic products.	40
Figure 3: Total EST number (4.184) used in the assembly of the 293 contigs with enzymatic products (gray bars).	46

Manuscript 3

Figure 1: The schema used in shrimp population sampling.	55
---	----

Capítulo 3

Manuscrito 4

Figura 1: Distribuição dos termos do GO encontrados do nível 2 para as categorias Processo Biológico, Função Molecular e Componente Celular.	80
Figura 2: Distribuição do top 15 das vias mapeadas no KEGG.	82

Capítulo 4

Manuscrito 5

Figura 1: Esquema do desenho amostral dos 48 animais utilizados na análise de expressão diferencial no edgeR, para identificação dos genes induzidos na comparação entre Alto e Baixo crescimento.	93
Figura 2: MA plot resultante da comparação entre alto (log2FC negativos) e baixo (log2FC positivos) crescimento.....	96
Figura 3: Representação dos 40 genes com menor valor de p-ajustado para a comparação entre os animais de alto (<i>High</i>) e baixo (<i>Low</i>) crescimento.	98

Capítulo 5

Manuscrito 6

Figura 1: Esquema amostral dos oito animais utilizados no experimento de exposição ao WSSV.....	114
Figura 2: MAplot resultante da comparação entre hepatopâncreas e músculo.....	117
Figura 3: <i>Heatmap</i> das amostras de hepatopâncreas (HP) e músculo (MU).....	121
Figura 4: MAplot resultante da comparação entre as condições doente e saudável.....	123
Figura 5: Distância entre as amostras dos animais saudáveis e doentes considerando-se a informação do tecido.	125
Figura 6: União dos genes DE <i>up-regulated</i> nos tecidos hepatopâncreas (HP) e músculo (MU) e nos tratamentos saudável (<i>Healthy</i>) e doente (<i>Ill</i>).	137

Lista de Tabelas

Capítulo 2

Manuscript 1

Table 1: Assembly summary of EST-transcriptome analyses from <i>L. vannamei</i>	20
Table 2: Characterization of the identified non-redundant SSR and SNP loci through data mining analysis performed on transcriptome of <i>L. vannamei</i>	25

Manuscript 2

Table 1: Main enzymes present in most common pathways. Note molecular function and main participating pathways, based on KEGG database.....	41
---	----

Manuscript 3

Table 1: Information about the four loci tested in association tests.....	54
Table 2: Weight mean and standart deviation (SD) values observed in families, groups and sex, followed by the whole animal body weight for the 200 animals collected.....	55

Table 3: The allelic and genotypic frequencies for the two polymorphic loci used in genotypes x phenotypes tests.....	57
Table 4: Variance Analysis (ANOVA) results for the effects tested.....	57
Table 5: Cross amplification results for three <i>L. vannamei</i> loci tested in seven marine and two fresh water shrimp species.....	59

Capítulo 3

Manuscrito 4

Tabela 1: Estatísticas das 12 amostras utilizadas para a construção da referência após a normalização e filtragem no SeqyClean.	71
Tabela 2: Taxas de transições e transversões observadas nos SNPs identificados no transcriptoma de <i>L. vannamei</i>	73
Tabela 3: Locos com maior número de SNPs exclusivos para as amostras de rápido crescimento, exposição ao WSSV e comuns a ambas as condições com <i>hits</i> para espécies de artrópodes.....	78

Capítulo 4

Manuscrito 5

Tabela 1: Peso médio, em gramas, das famílias avaliadas para alto e baixo crescimento no momento da coleta na fase de pré-engorda.....	94
Tabela 2: Genes candidatos envolvidos no ganho de peso e imunidade inata nos artrópodes na comparação alto (log2FC negativos) e baixo (log2FC positivos) crescimento.....	97
Tabela 3: Resultado dos termos do GO dos genes diferencialmente expressos entre as condições de alto e baixo crescimento e as vias enriquecidas.....	100

Capítulo 5

Manuscrito 6

Tabela 1: Genes candidatos envolvidos na resposta imune contra patógenos nos decápodes para a comparação entre os tecidos hepatopâncreas (HP) (log2FC < 0.5) e músculo (MU) (log2FC > 0.5) no experimento 1.....	118
Tabela 2: Resumo do resultado dos termos Processo Biológico (BP), Função Molecular (MF) e Componente Celular (CC) do GO dos genes DE da comparação entre hepatopâncreas e músculo e as vias enriquecidas.....	122
Tabela 3: Genes candidatos envolvidos na resposta imune encontrados na comparação entre animais doentes e saudáveis no Experimento 2.....	124
Tabela 4: Resultado dos termos Biológico (BP), Função Molecular (MF) e Componente Celular (CC) do GO dos genes diferencialmente expressos entre os animais saudáveis e doentes e as vias enriquecidas.....	127

Resumo

O presente trabalho teve como objetivo geral realizar a anotação genômica de sequências expressas (ESTs) de *Litopenaeus vannamei*, disponíveis no banco de dados do Projeto ShEST e avaliar o polimorfismo de marcas SSR e SNP mineradas. Esses marcadores estavam localizados na cadeia principal de genes de proteínas com função relacionada a traços de desempenho e foram validados em famílias de camarões SPF (*Specific Pathogen Free*) submetidas à seleção para rápido crescimento e sobrevivência. Adicionalmente aos locos SSR-EST e SNP-EST, obtidos por sequenciamento Sanger, análises de Sequenciamento de Próxima Geração (NGS) foram incluídas na proposta inicial de trabalho com o objetivo de ampliar o conjunto de SNPs disponíveis e verificar a expressão gênica diferencial. A montagem *de novo* das ESTs foi realizada e produziu um conjunto de 2.984 unigenes com produtos proteicos para 41% destes, sendo identificados 1.983 SSRs e 3.472 SNPs. Dentre os locos com produto gênico identificado, 231 eram enzimas com 127 *EC numbers* únicos inseridos em 94 vias metabólicas do KEGG. A validação dos locos SSR-EST e SNP-EST mostrou que os locos das proteínas 60S ribossomal (SSR-EST) e crustacianina (SNP-EST) apresentaram-se polimórficos nos animais amostrados da Genearch. Análises estatísticas foram conduzidas para verificação da existência de uma possível associação entre os genótipos e os fenótipos de peso analisados, embora não tenha sido observada associação. Além disso, testes de amplificação heteróloga foram realizados em sete espécies de camarões marinhos e duas de água doce, demonstrando sucesso na transferabilidade para estas espécies. A abordagem de RNA-seq foi incluída no presente trabalho com o propósito de ampliar o número de SNPs detectados em genes candidatos com função relacionada a traços de desempenho e identificar genes diferentemente expressos (DE) em animais sob condições experimentais. Foi realizada a montagem *de novo* de um segundo transcriptoma, dos tecidos músculo e hepatopâncreas de indivíduos *de L. vannamei* (i) avaliados para rápido crescimento e sobrevivência e (ii) expostos ao vírus da Síndrome da Mancha Branca ou *White Spot Syndrome Virus* (WSSV). Foram gerados 63.105 transcritos, com tamanho médio de 2.511 pb e N50 de 3.464 pb. Foram identificados mais de 15.500 SNPs (frequência > 50%). Também foi realizada a anotação funcional nas bases do SwissProt, Gene Ontology (GO) e KEGG. Análises de expressão diferencial gênica foram realizadas nas amostras dos animais avaliados para crescimento e resposta a infecção pelo WSSV. Os dados gerados demonstraram diferenças no perfil de expressão entre os genes (i) de animais de alto e baixo crescimento, (ii) do hepatopâncreas e músculo e (iii) dos animais não-infectados (saudáveis) e infectados (doentes) pelo WSSV, considerando-se o efeito do tecido. Foram identificados 207 genes DE para crescimento, 5.816 para a comparação entre os tecidos e 1.017 para os animais doentes e saudáveis.

Palavras-chave: ESTs, Polimorfismos, Transcriptomas, RNA-seq, Genes Diferencialmente Expressos (DE).

Abstract

The present work had as general objective to perform the genomic annotation of Expressed Sequences (ESTs) of *Litopenaeus vannamei* shrimp, available in the database of Project ShEST and to evaluate the polymorphism of mined SSR and SNP tags. These markers were located in the main chain of protein genes with function related to performance traits and were validated in SPF (Specific Pathogen Free) shrimp families submitted to selection for rapid growth and survival. In addition to the EST-SSR and EST-SNP loci, obtained by Sanger sequencing, Next Generation Sequencing (NGS) analyzes were included in the initial proposal of work with the objective of expanding the set of SNPs available and verifying the differential gene expression. The new assembly of ESTs was performed and produced a set of 2.984 unigenes with protein products for 41% of them, with 1.983 SSRs and 3.472 SNPs being identified. Among the loci with gene product identified, 231 were enzymes with 127 unique EC numbers inserted in 94 KEGG metabolic pathways. Loci validation showed that the loci of the 60S ribosomal (SSR-EST) and crustacyanin (SNP-EST) proteins were polymorphic in the animals sampled from Genearch. Statistical analyzes were conducted to verify the existence of a possible association between the genotypes and the analyzed weight phenotypes, although no association was observed. In addition, cross-species amplification tests were performed on seven species of marine and two freshwater prawns, demonstrating successful transferability for these species. The RNA-seq approach was included in the present work with the purpose of increasing the number of SNPs detected in candidate genes with performance-related function and identifying differentially expressed (DE) genes in animals under experimental conditions. A second transcriptome was assembled from the muscle and hepatopancreas tissues of *L. vannamei* individuals (i) evaluated for rapid growth and survival and (ii) exposed to the White Spot Syndrome Virus (WSSV). A total of 63.105 transcripts were generated, with an average size of 2.511 bp and N50 of 3.464 bp. More than 15.500 SNPs were identified (frequency > 50%). Functional annotation was also performed on the bases of SwissProt, Gene Ontology (GO) and KEGG. Differential gene expression analyzes were performed on the animal samples evaluated for growth and response to WSSV infection. The data generated showed differences in the expression profile between the genes of (i) high and low growth animals, (ii) the hepatopancreas and muscle and (iii) the uninfected (healthy) and infected (ill) animals by WSSV, considering the effect of the tissue. Two-hundred and seven DE genes were identified for growth, 5.816 for hepatopancreas and muscle and 1.017 for ill and healthy animals.

Keywords: ESTs, Polymorphisms, Transcriptomes, RNA-seq, Differentially Expressed (DE) Genes.

1. Introdução do Problema

A espécie *Litopenaeus vannamei* (Figura 1) trata-se de um camarão peneídeo classificado taxonomicamente da seguinte forma:

Filo *Arthropoda*

Sub-filo *Crustacea*

Classe *Malacostraca*

Ordem *Decapoda*

Família *Penaeidae*

Gênero *Litopenaeus*

Espécie *Litopenaeus vannamei* (Boone, 1932)

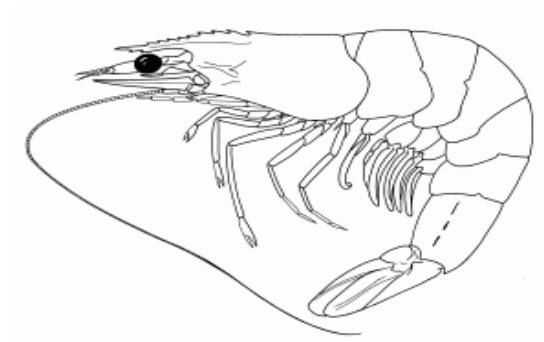


Figura 1: O camarão *Litopenaeus vannamei*

Popularmente, *L. vannamei* é conhecida como camarão branco do Pacífico e representa a principal espécie de camarão mundialmente comercializada. Sua distribuição original vai desde a costa oriental do oceano Pacífico nas proximidades do México até o Peru (Ma *et al.*, 2009). Apesar de sua ampla área de ocorrência a degradação dos mares e o aumento do consumo de camarão na alimentação humana, tornaram impossível o suprimento da demanda mundial através apenas da

pesca. Como alternativa à atividade extrativista, tecnologias de manejo foram desenvolvidas para possibilitar a criação desta espécie em cativeiro (Perez-Henriquez *et al.*, 2009).

Embora a espécie apresente uma boa adaptação ao cativeiro, alguns fatores ambientais podem frear o seu desempenho. Condições estressantes, como altas densidades populacionais, contatos com patógenos, dentre outros, podem resultar em crescimento lento e baixa sobrevivência (Jung *et al.*, 2013; Lv *et al.*, 2014), causando prejuízo significativo a carcinicultura (Chen *et al.*, 2015). Dessa forma, torna-se importante o conhecimento da fisiologia dos tecidos responsáveis por características como crescimento e imunidade em crustáceos, tais como músculo, hemócitos e hepatopâncreas (Baldwin *et al.*, 2009; Jung *et al.*, 2013).

A quantidade de informação conhecida sobre traços de desempenho em crustáceos ainda é limitada, sabendo-se muito pouco sobre o modo de ação de genes e proteínas determinantes (Lv *et al.*, 2014; Zhang *et al.*, 2014). Neste contexto, marcadores moleculares contidos em regiões expressas do genoma, como os microssatélites ou SSR (*Simple Sequence Repeats*) e os SNP (*Single Nucleotide Polymorphisms*), podem ser úteis para acessar a variabilidade genética dos plantéis, identificar locos relacionados a traços de interesse e auxiliar no estabelecimento de programas de monitoramento e melhoramento genético (Perez *et al.*, 2005; Jin *et al.*, 2013). Adicionalmente, a produção de transcriptomas através de Sequenciamento de Próxima Geração (NGS) de diversos tecidos de animais sob condições experimentais, além de análises de expressão diferencial (Oshlack *et al.*, 2010; McCarthy *et al.*, 2012), podem contribuir para a descrição de genes candidatos que possibilitem a seleção assistida por marcadores em *L. vannamei*.

A anotação funcional de transcritos contendo polimorfismos em diversos bancos de dados permite a atribuição de um significado biológico a uma sequência de nucleotídeos/aminoácidos e consiste em uma série de *blasts* (*Basic Local Alignment Search Tool*) ou alinhamentos diversos que buscam por similaridades entre as sequências estudadas (*query*) e as depositadas nos bancos de

dados (BDs) públicos (*subject*) com a finalidade de elucidar os possíveis produtos gênicos. Nesta abordagem, inúmeros bancos de dados são acessados, como por exemplo, o *SwissProt* (<http://www.ebi.ac.uk/uniprot/>) (UniProt Consortium, 2014) e o *Gene Ontology (GO)* (<http://www.geneontology.org/>) (Ashburner *et al.*, 2000), fornecendo informações sobre o produto e função gênicos, componente celular, processo biológico e função molecular, baseando-se em proteínas já bem estudadas.

Um dos principais BDs acessados é o KEGG (*Kyoto Encyclopedia of Genes and Genomes*) que é caracterizado por unir informações genômicas às de ordem funcional. A união destes dois conjuntos é vista pela forma de *networks* proteicas ou enzimáticas que possuem *EC numbers* conectando-os. Quando os genes são identificados, estes são rearranjados de acordo com a similaridade das sequências, os *EC numbers* são determinados e as vias metabólicas são construídas com o auxílio de ferramentas computacionais, correlacionando genes e enzimas (Kanehisa & Goto, 2000). As informações obtidas consistem em uma rica fonte de informação para estudos na espécie foco e em outras taxonomicamente próximas, nas quais poucos dados genéticos são conhecidos.

2. Objetivos e Hipóteses de Trabalho

Considerando a problemática apresentada, o presente trabalho teve como objetivo geral estabelecer um conjunto relevante de informações sobre genes e polimorfismos para a espécie *L. vannamei* e para outras espécies de camarão, tendo as seguintes hipóteses de trabalho testadas:

- **Hipótese I.** Regiões expressas (EST) do genoma contem polimorfismos de base única (SNP) e de sequências simples repetidas (SSR) inseridos na cadeia principal de proteínas.

- **Hipótese II.** Polimorfismos nos genes codificadores das proteínas 60S ribossomal, serina arginina, troponina e crustacianina podem estar associados a características de desempenho nos animais avaliados para crescimento e sobrevivência amostrados.
- **Hipótese III.** Existe diferença no perfil de expressão dos genes de indivíduos de *L. vannamei* de famílias com baixo e alto crescimento.
- **Hipótese IV.** Existe diferença no perfil de expressão dos genes de indivíduos de *L. vannamei* não infectados e infectados pelo vírus da Mancha Branca.

Para testar as hipóteses apresentadas, os seguintes objetivos específicos de trabalho foram estabelecidos:

Em relação à mineração das ESTs:

1. Realizar a anotação genômica nos transcriptomas de *L. vannamei* analisados para determinação dos respectivos produtos gênicos, caracterização de *EC numbers* e respectivas enzimas e identificação das possíveis vias metabólicas;
2. Identificar locos SSR e SNP presentes em genes de interesse que estejam relacionados a características de desempenho em peneídeos;
3. Validar os marcadores contidos nestes genes e verificar uma possível associação entre os polimorfismos identificados e os fenótipos avaliados em populações de *L. vannamei* submetidas a programas de seleção para crescimento e sobrevivência;
4. Avaliar a taxa de transferabilidade dos marcadores validados para *L. vannamei* em espécies de camarões marinhos e de água doce, as quais possuem poucos locos descritos;

Em relação ao transcriptoma obtido através de RNA-seq:

5. Avaliar o transcriptoma de indivíduos de *L. vannamei* SPF (i) de alto e baixo desempenho de crescimento e (ii) infectados e não infectados pelo vírus da Mancha Branca ou *White Spot Syndrome Virus* (WSSV);
6. Identificar polimorfismos (SNP) no transcriptoma construído de *L. vannamei*;
7. Identificar genes diferencialmente expressos (DE) entre as condições de alto e baixo crescimento;
8. Identificar genes DE nas comparações entre hepatopâncreas (Hp) e músculo (Mu) e entre animais não infectados (saudáveis) e infectados (doentes) pelo WSSV, levando-se em consideração o tecido amostrado.

3. Material e Métodos

As amostras biológicas das famílias utilizadas no presente trabalho consistiram de animais avaliados para rápido crescimento e sobrevivência da empresa Genearch, localizada no estado do Rio Grande do Norte. Os métodos de abertura de novos *batches* e do programa de seleção realizados na empresa Genearch estão detalhados abaixo.

3.1. Composição e Manutenção dos *Batches* na Genearch

O programa de melhoramento genético da Genearch foi desenvolvido com base no modelo familiar e se iniciou no ano de 2007, a partir de quatro linhagens de *L. vannamei* provenientes do Panamá, México, Havaí e Florida. Trata-se de animais SPF (*Specific Pathogen Free*) ou livres de patógenos específicos. Ao serem importadas para o Brasil, estas linhagens continuaram a ser mantidas como SPF, em instalações de sanidade desenvolvidas especificamente para essa

finalidade, dando origem aos *batches* iniciais que compuseram a base genética utilizada no programa de melhoramento desta empresa.

As linhagens SPF importadas dos EUA foram misturadas entre si para compor os diferentes *batches* da Genearch. Um *batch* é formado da desova de 20 casais, um de cada família. A cada três meses um novo *batch* é aberto, totalizando quatro novos *batches* por ano. Estes quatro *batches* abertos anualmente constituem uma geração. Para minimizar a perda de variação genética, um novo *batch* é aberto a partir do cruzamento entre indivíduos dos terceiro e quarto *batches* anteriores distintos (o *batch* atual X é originado do cruzamento entre indivíduos de famílias pertencentes aos *batches* X-4 e X-3). Por exemplo, o *batch* 5 foi formado pelo cruzamento de 20 machos e 20 fêmeas provenientes dos *batches* 1 e 2 (Figura 2).

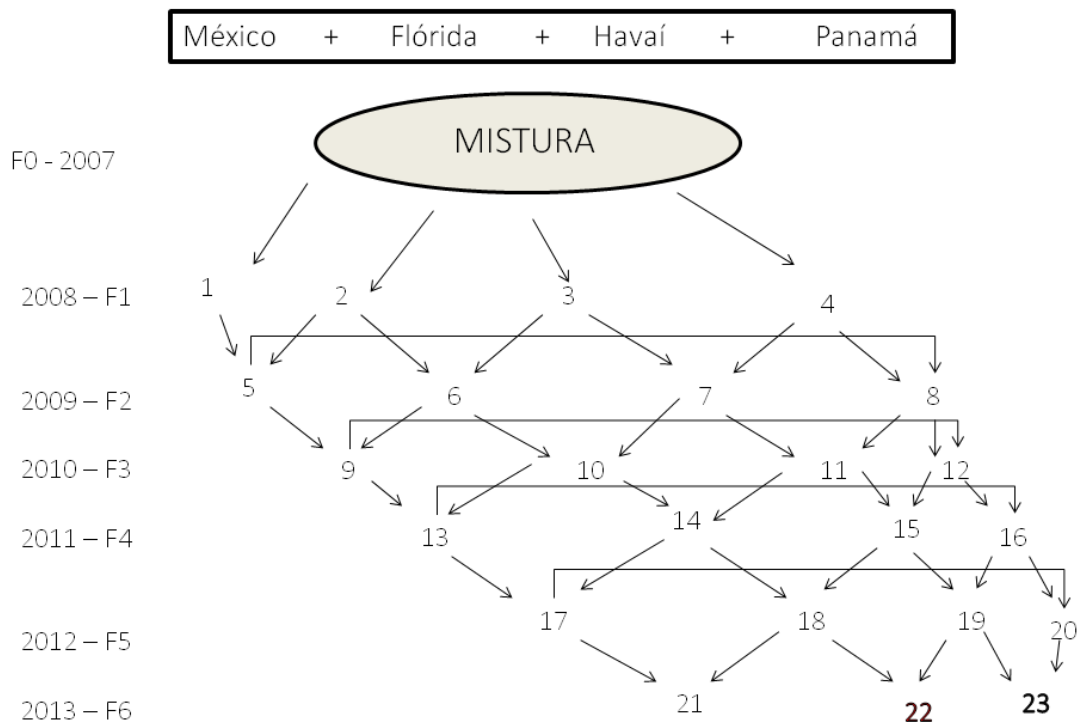


Figura 2: Esquema de abertura de novos *batches*, exemplificando a origem dos animais reprodutores que fundam os *batches* consecutivos.

3.2. Seleção das Famílias

A escolha dos casais para a formação de cada família é feita com base em critérios de seleção para sobrevivência e crescimento. Esta seleção inclui a avaliação das 40 famílias pertencentes aos dois *batches* que serão utilizados para abertura de um novo. Cada família é composta pela desova de um único casal, constituindo-se, portanto, em uma linhagem de irmãos completos. A desova ocorre em tanques individuais, sendo montadas trélicas de cada uma das desovas. Os indivíduos de uma mesma família são marcados com combinações de cores de elastômeros diferentes para haver um controle da origem dos mesmos. A seleção foi realizada de modo distinto, sendo primeiramente entre as famílias, nos períodos de pré-engorda e engorda, e dentro dos animais da mesma família, na etapa de maturação (Figura 3).

- Seleção Pré-Engorda (Interfamiliar)

A seleção na pré-engorda incluiu avaliações de performance para taxa de sobrevivência e crescimento realizadas durante o desenvolvimento dos estágios de náuplio à juvenil. Os animais de alto e baixo crescimento utilizados na construção das bibliotecas de cDNA empregadas no RNA-seq foram originados dessa etapa. Cada família foi mantida em triplicata para minimização do efeito do meio, sendo que as mensurações foram realizadas nos 60 grupos do mesmo *batch*, os quais foram avaliados nas respectivas fases:

- o **Fase 1:** cerca de 6 mil náuplios foram avaliados até atingirem a fase de PL3 (pós-larvas com 3 dias);
- o **Fase 2:** cerca de 2 mil PL4 (pós-larvas com 4 dias) foram avaliadas até atingirem a fase de PL20;
- o **Fase 3:** cerca de mil indivíduos juvenis (pós PL20, camarões com 2 gramas) foram avaliados.

Essas três fases duraram cerca de dois meses e durante todo o processo de seleção da pré-engorda foram realizadas análises quantitativas para estabelecimento das melhores famílias. Dentre as famílias avaliadas nesta primeira etapa de seleção, foram escolhidas as 12 que apresentaram as melhores taxas médias de sobrevivência e crescimento. Posteriormente a esta seleção, as réplicas de cada família selecionada foram reunidas em uma só. Tem início a fase de engorda, a qual incluiu a análise interfamiliar das 12 linhagens formadas, agora, sem suas respectivas réplicas.

- Seleção Engorda (Interfamiliar)

A partir das mensurações realizadas nas 12 famílias escolhidas, as sete melhores foram selecionadas com base na análise de critérios quantitativos das performances médias de cada uma delas.

- Seleção Maturação (Intrafamiliar)

Após a escolha das sete melhores famílias foi realizada a seleção intrafamiliar a partir de mensurações individuais para crescimento nos animais com cerca de três meses e meio de idade. Os camarões foram separados por sexo e classificados de acordo com critérios de corte estabelecidos para ganho de peso e taxa de crescimento dentro de cada família. Os pesos de corte foram calculados pela empresa através de modelos mistos e métodos de componentes de variância com base na estrutura dos pedigrees das famílias (*Best Linear Unbiased Prediction - BLUP*). Estes parâmetros definiram três classes, sendo elas:

o Classe Descarte - Não Reprodutores: Composta por animais que são descartados por apresentarem peso muito abaixo da média de corte e características morfológicas indesejáveis, como cutícula mole. Constitui cerca de 70% dos animais de uma família.

o Classe Não *Speed* – Reprodutores que seguirão para a empresa Aquatec: Composta por camarões com peso médio baseado no valor de corte. Constitui cerca de 26% dos animais de uma família. Os animais utilizados na validação dos SSRs e SNPs e testes de associação neste trabalho foram amostrados nessa etapa.

o Classe *Speed* – Reprodutores Superiores (Genearch): Composta por camarões com peso superior ao valor de corte. Constitui cerca de 4% dos animais de uma família.

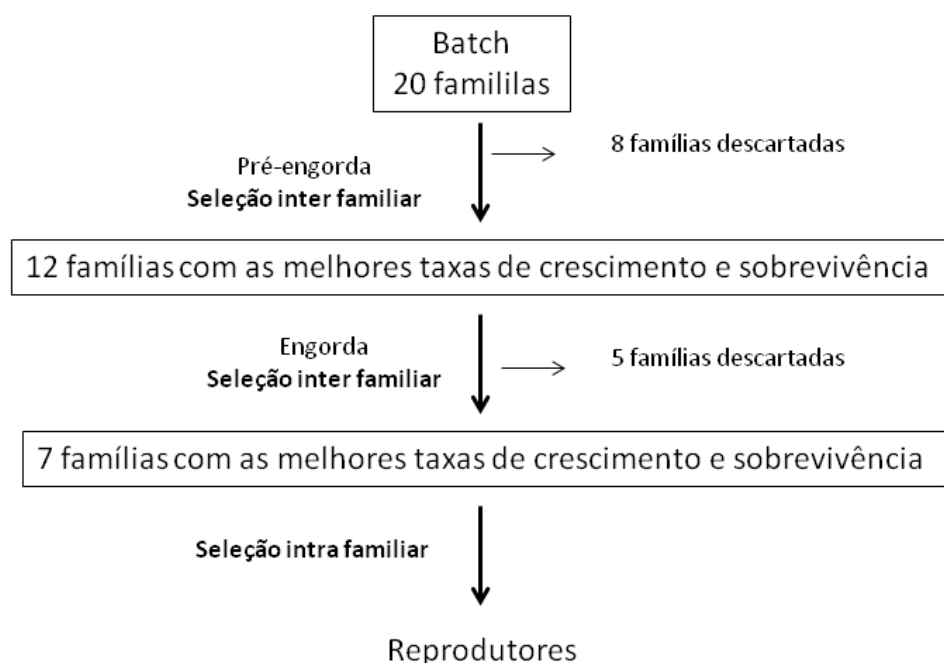


Figura 3: Sistema de seleção aplicado na Genearch nas fases de pré-engorda, engorda e maturação.

Dentro de cada uma das sete famílias de melhor desempenho, os animais com pesos acima da média de corte foram selecionados para comporem os plantéis de reprodutores superiores utilizados na abertura dos novos batches da Genearch. Nesta etapa de maturação, foram

selecionados também os indivíduos com pesos dentro da média de corte. Estes animais são os reprodutores comerciais da Genearch. Eles são fornecidos à empresa de larvicultura Aquatec como reprodutores SPFs para produzirem as larvas comerciais que abastecerão as inúmeras fazendas de engorda para produção comercial brasileira de camarões desta espécie. Os animais expostos ao Vírus da Mancha Branca utilizados no experimento de RNA-seq deste trabalho foram originados dos reprodutores comerciais da Genearch que foram para a Aquatec. Esses animais integraram a linhagem e SpeedLine SPF. Nessa última empresa, as pós-larvas produzidas por esses reprodutores foram compradas por uma fazenda de engorda com alta incidência da doença.

Na Aquatec, além dos animais fornecidos pela Genearch há outras linhagens geneticamente melhoradas que juntamente com a linha da Genearch são utilizadas para formar as pós-larvas comerciais. Os camarões pertencem a linhagem SpeedLine BR (BRA) e SpeedLine SPF (SPFA). A linhagem BRA é formada por animais que passaram por um processo de seleção familiar iniciado em 1999 e finalizado em 2005, quando a seleção na Aquatec passou a ser individual. Já a linhagem SPFA é composta por animais provenientes de diferentes famílias da Genearch, composta por animais SPF selecionados para altas taxas de crescimento.

Detalhes sobre as metodologias utilizadas para coleta do material biológico e procedimentos realizados para a produção e análise dos dados moleculares encontram-se descritos nos respectivos capítulos que compõem esta tese.

4. Resultados e Discussão

Os principais resultados obtidos neste trabalho são apresentados e discutidos em capítulos organizados sob a forma de manuscritos científicos que foram publicados, submetidos ou encontram-se em vias de submissão à publicação.

Capítulo 1

RNA-seq as a powerful tool for penaeid shrimp genetic progress (Publicado)



RNA-seq as a powerful tool for penaeid shrimp genetic progress

Camilla A. Santos*, Danielly V. Blanck and Patrícia D. de Freitas

Laboratory of Molecular Biodiversity and Conservation, Department of Genetics and Evolution, Federal University of São Carlos, São Carlos, Brazil

Edited by:

José Manuel Yáñez, University of Chile, Chile

Reviewed by:

Hsiao-Ching Liu, North Carolina State University, USA

Gonzalo Rincon, University of California, Davis, USA

*Correspondence:

Camilla A. Santos, Laboratory of Molecular Biodiversity and Conservation, Department of Genetics and Evolution, Federal University of São Carlos, Washington Luis Road (SP 310), Km 235, Caixa Postal 676, CEP: 13565-905, São Carlos, São Paulo, Brazil
e-mail: camilla.alves@yahoo.com.br

The sequences of all different RNA transcripts present in a cell or tissue that are related to the gene expression and its functional control represent what it is called a transcriptome. The transcripts vary between cells, tissues, ontogenetic and environmental conditions, and the knowledge that can be gained through them is of a solid relevance for genetic applications in aquaculture. Some of the techniques used in transcriptome studies, such as microarrays, are being replaced for next-generation sequencing approaches. RNA-seq emerges as a new possibility for the transcriptome complexity analysis as well as for the candidate genes and polymorphisms identification of penaeid species. Thus, it may also help to understand the determination of complex traits mechanisms and genetic improvement of stocks. In this review, it is first introduced an overview of transcriptome analysis by RNA-seq, followed by a discussion of how this approach may be applied in genetic progress within penaeid stocks.

Keywords: differential expression analysis, differential phenotype, high-throughput transcriptome, penaeidae transcriptome, candidate genes

INTRODUCTION

The term RNA-seq has been used to make reference to a transcriptome produced by methods of next generation sequencing (NGS), which ensure a good coverage of transcripts detection, due to the sequencing of millions of reads ranging from 25 to 300 bp, depending on the platform used (Wang et al., 2009; Oshlack et al., 2010). The full set of transcripts in a cell is known as transcriptome. They involve all types of ribonucleic acids (RNAs), including the protein coding messenger ribonucleic acid (mRNA) and the non-coding ribonucleic acid (ncRNA) such as ribosomal RNAs (rRNA), transfer RNAs (tRNA), and the small nuclear RNAs (snRNA). These RNAs may be differentially expressed according to the tissue, the stage of development and the physiological condition being accessed (Wang et al., 2009; Anders and Huber, 2010).

Transcriptome studies have been widely conducted in order to identify new genes, prospect simple sequence repeats (SSR) and single nucleotide polymorphisms (SNP) markers and to analyze differentially expressed genes. Such approaches have been helping to understand different mechanisms related to cellular control and describe important metabolic pathways, what enables a better understanding of the genotype–phenotype relationship (Marguerat and Bähler, 2010; Khatri et al., 2012; Qian et al., 2014).

Small and large scale transcriptome analyses and differential expression studies, such as Expressed Sequence Tags (ESTs) and microarrays, have been carried out in some penaeid shrimp species (Rojtinnakorn et al., 2002; La Vega et al., 2007; James et al., 2010; Brady et al., 2013). However, RNA-seq approaches are still incipient in shrimp (Table 1; Li et al., 2012, 2013; Guo et al., 2013; Sookruksawong et al., 2013; Xue et al., 2013; Zeng et al., 2013; Baranski et al., 2014; Yu et al., 2014). Therefore, the method herein has emerged as a new possibility for the transcriptome complexity

analyses in face of varied production and/or experimental conditions. Consequently, such approach aims to develop genetically improved strains, focusing mainly on the resistance factor.

In order to obtain a transcriptome via RNA-seq, some steps should be followed: (i) selection of tissue of interest and isolation of RNA molecules; (ii) construction of cDNA libraries; (iii) utilization of a NGS platform; and (iv) the reads analysis in order to establish unigenes and the transcriptome assemble through bioinformatics tools.

The tissue choice should be based on the study aim and/or the genes nature to be analyzed. As a parallel, a transcriptome consists in taking a photograph from a specific time in a cell, highlighting only the condition at that short period of time. In this manner, tissue selection and the suitable time to perform a transcriptome requires preparation, otherwise the experiment as a whole may be biased (Wang et al., 2009). Libraries establishment is crucial for the final result in face of the many laboratorial procedures that are conducted, leading to some biases in the obtained results (Wang et al., 2009). Bioinformatics analysis is also an important step and includes the use of computational tools that guarantee the processing of large volumes of data generated by next-gen (Gavery and Roberts, 2012; Guo et al., 2013).

Within this review, it is presented a brief overview of the RNA-seq method, including its main advantages and limitations. Following that, it will be discussed how such technique may be applied to obtain genetic progress in penaeid shrimp farming.

RNA-seq: ADVANTAGES AND LIMITATIONS

The transcriptome assembly may be based on a reference genome available (Wang et al., 2009), which allows to quickly locate similar

Table 1 | Summary of studies on penaeid shrimp using the RNA-seq approach.

Reference/species	NGS platform	Application	Annotation of genomic information
Baranski et al. (2014)/ <i>Penaeus monodon</i>	Illumina	SNPs and microsatellite prospecting	136,223 contigs; 473,620 SNPs
Li et al. (2013)/ <i>Fenneropenaeus chinensis</i>	Illumina	Immune response to WSSV	46,676 unigenes; 805 differentially expressed genes
Yu et al. (2014)/ <i>Litopenaeus vannamei</i>	Illumina	SNPs prospecting	96,040 SNPs
Guo et al. (2013)/ <i>L. vannamei</i>	Illumina	Stress; immune response; apoptosis	42,336 unigenes; 2,445 differentially expressed genes
Sookruksawong et al. (2013)/ <i>L. vannamei</i>	Illumina	Immune response to TSV	61,937 contigs; 1,374 differentially expressed contigs
Zeng et al. (2013)/ <i>L. vannamei</i>	454	Immune response to TSV; microsatellites prospecting	15,004 unigenes; 770 microsatellites; 1,311 differentially expressed genes
Xue et al. (2013)/ <i>L. vannamei</i>	Illumina	Immune response to WSSV	52,073 unigenes; 1,179 differentially expressed genes
Li et al. (2012)/ <i>L. vannamei</i>	Illumina	Genomic annotation	109,169 unigenes

regions using local alignment algorithms and presents higher reliability due to the large volume of small sized reads coming from alternative splicing. It equally provides a more even coverage of the genome (Anders and Huber, 2010; Qian et al., 2014). On the other hand, even when there is no reference genome available, *de novo* transcriptome assembly may be carried out using specific algorithms, which stands as a solid advantage for some species that have not been widely studied yet (Howe et al., 2013; O'Neil and Emrich, 2013).

When performing transcriptome via RNA-seq, a high coverage is obtained, which allows the discovery of new genes and polymorphisms (Marguerat and Bähler, 2010; Yu et al., 2014). Li et al. (2012) evaluated the abundance and coverage of transcriptomes obtained by RNA-seq in *Litopenaeus vannamei*. By comparing such data to the ESTs available on GenBank, it was found that only 14.2% (15,519 out of 109,169) of unigenes obtained by RNA-seq were also found in the EST libraries, generating a lot of new informative data. In addition to that, the wide coverage associated with high resolution provided by this technique ensured high accuracy in SNP discovery in coding genes (Yu et al., 2014).

That taken, RNA-seq allows the detection of variations in a single nucleotide, enabling the detection of the expression of protein isoforms and their respective allelic variants, characterizing SNPs (Baranski et al., 2014; Yu et al., 2014). Polymorphic microsatellites or SSRs have equally been identified through RNA-seq analysis (Mohd-Shamsudin et al., 2013; Zeng et al., 2013; Baranski et al., 2014). On those cases, though, a wider coverage of the reference genome is suggested (Qian et al., 2014) once the presence of highly repetitive regions could stand as a limiting factor by compromising the transcriptome assembly.

As could be observed, RNA-seq has been considered a solid method for the large-scale gene expression analysis due to the fact it does not require prior genome knowledge (Wang et al., 2009) and enables the detection of isoforms arising from alternative splicing (Ghosh and Qin, 2010). Even when involving several samples, such technique is accessible with moderate costs. In this case multiplex

runs containing up to 10 samples per sequencing lane can be performed in some platforms and the costs are no longer a limiting factor.

Another RNA-seq advantage is its wide dynamic range (ratio between the minimum and maximum expression level). This feature makes it suitable for measuring low, medium and high expression levels of the genes, not requiring very sophisticated normalization. By contrast, DNA microarrays show reliable results only for medium expression levels and therefore have a much smaller dynamic range. Thus, RNA-seq provides much more informative data, requiring less biological material and lower costs, becoming this technique popular for measuring gene expression on a large scale (Sharov et al., 2004; Wang et al., 2009).

RNA-seq APPLICATION WITHIN PENAID SHRIMP AQUACULTURE

The use of RNA-seq in species of penaeid shrimp can be focused on transcriptome characterization, functional annotation, gene expression profiles analysis, and gene-associated markers identification. In this section, the emphasis is given to the analysis of differential expression, identification of molecular markers, and its potential to promote genetic gain and development of improved penaeid strains. Similar studies have allowed the identification of candidate genes or quantitative trait loci (QTLs), which could be related to traits of interest for aquaculture, such as reproduction, sex determination, growth, immunity, and tolerance against environmental stress. Data involving pathways are also relevant in order to obtain more details about the interaction mechanism between the expressed products and their importance and applicability.

IDENTIFYING CANDIDATE GENES THROUGH DIFFERENTIAL EXPRESSION ANALYSIS

Although the applicability of RNA-seq in transcriptome and differential expression in aquatic organisms have increased in the past 3 years, the results found in literature and in the Sequence

Read Archive Database of the National Center for Biotechnology Information (SRA-NCBI) indicate that such approach still is incipient for penaeid shrimp. SRA databank, per instance, presents only 28 deposits of data generated by NGS in what regards the species *L. vannamei*, *L. stylirostris*, and *Penaeus monodon* (<http://www.ncbi.nlm.nih.gov/sra/?term=penaeidae>). In literature, it has been found next-gen data only for *P. monodon* (Baranski et al., 2014) *Fenneropenaeus chinensis* (Li et al., 2013) and *L. vannamei* (Li et al., 2012; Guo et al., 2013; Sookruksawong et al., 2013; Xue et al., 2013; Zeng et al., 2013; Yu et al., 2014). Mostly, the research in this field has been covering the identification of genes connected to immunity, mainly concerning the white spot syndrome virus (WSSV) and the taura syndrome virus (TSV; Li et al., 2013; Sookruksawong et al., 2013; Xue et al., 2013; Zeng et al., 2013; Baranski et al., 2014). Both syndromes have caused great economic losses for the shrimp industry throughout the past few decades.

Despite the fact that crustaceans do not own an immune system, some candidate genes have been obtained from hemolymph and hepatopancreas tissues. It is clearly seen in some differential expression studies concerning *L. vannamei* species, which represents the biggest portion of the marine shrimp worldwide production (Gucic et al., 2013). Among the main genes studied, it is possible to find those related to toll-like and signaling receptors, apoptosis, *Vibrio cholerae* infection and other immune proteins (e.g., phagosome, hemocyanin, crustacyanin, antiviral), antioxidant enzymes (the peroxidases and glutathione ones), and lectins (Figure 1; Li et al., 2012, 2013; Sookruksawong et al., 2013; Xue et al., 2013; Zeng et al., 2013; Baranski et al., 2014; Yu et al., 2014).

Data related to toll-like and lectin proteins demonstrate that those may act as signaling molecules, what causes the increase of peptides expression responsible for controlling the immune response (Wang et al., 2014). On the other hand, genes associated with apoptosis may indicate an attempt to prevent proliferation of viruses and possible damages to genetic material, through death of infected cells. Specifically in what regards large number of proteins related to infection by *V. cholerae* response is due to the recurring presence of this group of bacteria within shrimp farming tanks (Banerjee et al., 2012).

Information regarding the main metabolic pathways and the quantity of most frequent genes in each pathway were also collected, as part of the data obtained via functional annotation for RNA-seq. In penaeid, the most commonly described pathways were those involving the general metabolism, spliceosome, RNA transport, *V. cholerae* infection, phagosome and the antioxidant ones, which include peroxidase enzymes (Li et al., 2012, 2013; Sookruksawong et al., 2013; Xue et al., 2013; Zeng et al., 2013; Yu et al., 2014). Spliceosome and RNA transport pathway supposedly act in new transcripts formation, providing genetic variants that may contribute to resistance (Yang et al., 2007).

Regarding Gene Ontology (GO) categories, all studies in penaeid have mainly reported the same data. Considering the biological processes, per instance, the most frequent were metabolism and biological regulation. In what regards cellular components, genes are mostly expressed at the cell and some unspecific organelles. Finally, concerning the molecular function, the most

common ones were catabolic activity and binding (Li et al., 2012, 2013; Sookruksawong et al., 2013; Xue et al., 2013; Zeng et al., 2013; Baranski et al., 2014; Yu et al., 2014). Overall results such as these were expected, since the penaeid species previously mentioned herein have too little information about their genomes available. In the case of *L. vannamei*, only approximately 12,000 gene products were described, which may be useful in a comparative approach concerning a *de novo* assembly (<http://www.ncbi.nlm.nih.gov/protein/?term=Litopenaeus+vannamei>).

IDENTIFYING GENE ASSOCIATED MARKERS

In what regards RNA-seq technology, it has also been proving to be an extremely useful tool for identifying SNPs, which may be also used to develop high density SNPs chips for studies concerning the genome wide association (GWAS) and to build high density linkage maps (Baranski et al., 2014; Yu et al., 2014). Furthermore, SNPs can be used as markers in order to distinguish allelic transcripts whilst studying the allele-specific expression (Bell and Beck, 2009).

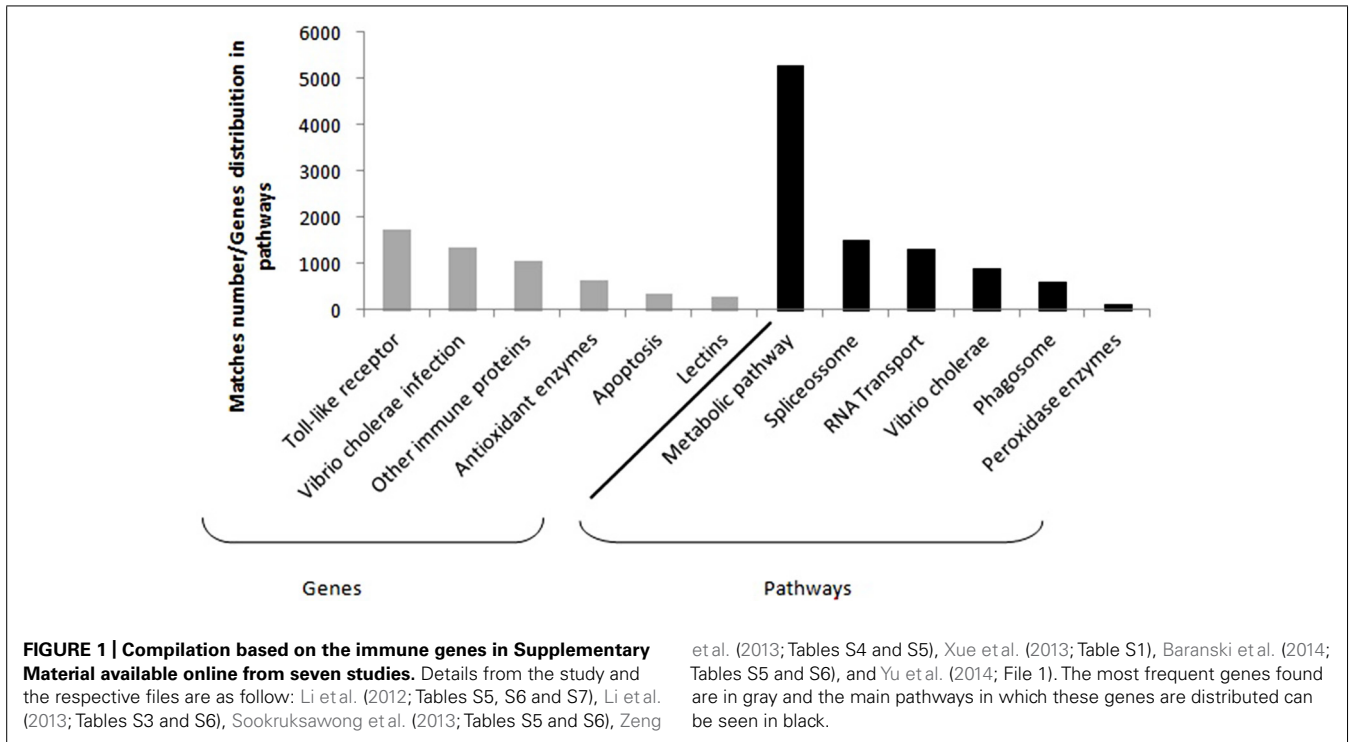
In a recent study, Yu et al. (2014) prospected SNPs in *L. vannamei*. A total of 58,717 unigenes and 36,277 high quality SNPs were predicted by transcriptomes “M” (produced by the authors themselves) and “P” (downloaded from SRA database, session number SRR346404, which was published by Li et al., 2012), respectively. Those SNPs were spread out among 25,071 unigenes and allocated to 254 pathways at the KEGG (Kyoto Encyclopedia of Genes and Genomes) database. The main pathways containing high number of SNPs were metabolic pathways, amoebiasis, *V. cholerae* infection, RNA transport, and actin cytoskeleton regulation.

Baranski et al. (2014) used the approach to build a high density linkage map in *P. monodon*. A total of 6,000 out of 473,620 SNPs/indels putative were genotyped by using the Illumina iSelectCerca genotyping matrix. Out of those SNPs, 3,959 were mapped in 44 linking groups and out of those 2,340 were functionally annotated according to the GO database (see dataset S5 and S6 from Baranski et al., 2014). According to the authors, these polymorphisms may be causal or closely related to other mutations that affect important traits, such as resistance to diseases and reproductive performance.

The identification and functional annotation of SNPs identified by Baranski et al. (2014) and Yu et al. (2014) studies represent a useful resource to comprehend mechanisms determining complex traits and, consequently, to develop programs aiming the genetic improvement of these characters in penaeid shrimp strains. That taken, those SNPs can be applied both in marker assisted selection (MAS), using SNPs closely associated with QTL, and in genomic selection, through complete set of identified SNPs. That increases the rate of genetic gain per generation in traits of great interest to the shrimp industry, such as growth and resistance to disease.

FINAL CONSIDERATIONS

One of the possible challenges that arise within genetic gain is the development of penaeid strains that may simultaneously present high growth development and pathogens resistance.



Genetic correlation studies have shown that there is a negative phenotypic correlation between the resistance to diseases and the weight gained by the animals (Argue et al., 2002; Gitterle et al., 2005; Cock et al., 2009). Cock et al. (2009), reinforce the fact specimen potentially resistant to WSSV also present low reproduction efficiency. Such remarks suggest that genes with pleiotropic effects may be responsible for the trade-off observed between these traits in penaeid shrimp. From this perspective, the RNA-seq technique can be used for discovering such genes, since the overlapping of differentially expressed genes in both strains resistant to pathogens and in large weight gain strains can also be verified. Therefore, up-regulated or down-regulated genes expressed in these two strains could indicate a possible pleiotropic effect. Besides, mRNA studies allied with RNA-seq method could also be used for micro RNAs (miRNA) analyses. This approach was applied in aquaculture species, such as freshwater prawn *Macrobrachium rosenbergii* (Tan et al., 2013), and tilapia (Huang et al., 2012). As a result, it has been shown that miRNAs are critical regulators of generalized cellular functions such as differentiation, proliferation, and cell growth.

Another challenge within aquaculture is the difficulty in achieving sexual maturity and spawning of penaeid species (except for *L. vannamei*), under the farming conditions (Lo et al., 2007; Brady et al., 2013). As an attempt to overcome such problem, ablation of the eyestalk has been conducted for many years. Nevertheless, such practice is associated to high mortality rates, and low spawning and survival rates (Huberman, 2000). Considering this, the transcriptome analysis obtained from reproductive organs of native and captive specimens of penaeid shrimp via RNA-seq may significantly contribute to the

identification of the underlying causes of reproductive dysfunction observed in farmed animals. Furthermore, the discovery of genes involved in gonadal maturation and reproductive performance may assist in gametogenesis, handling studies involving these species.

Finally, transcriptome and differential expression analysis by RNA-seq may be a powerful approach to optimize the penaeid diet composition (nutrigenomics), especially for those species that do not count on a specific availability of diet. The approach may be used to identify specific changes in molecular level (Chávez-Calvillo et al., 2010), which in turn also cause metabolic and physiological changes in shrimp treated with different diets (e.g., levels of crude protein, levels of plant protein inclusion and of antioxidants, vitamins, and polyunsaturated fatty acids). Thus, nutrigenomics can be used to produce healthy animals and safe and high quality products for the consumer, emerging as a promising area of research for sustainability and profitability in aquaculture (Cerdà and Machado, 2013).

Although NGS technologies are showing their efficiency in works related to gene expression, other methodologies such as third-generation sequencing, also referred to as single-molecule sequencing (Single-Molecule Real-Time, SMRT), are being developed, but already showing limitations. More advanced techniques of sequencing are also on the way, such as “next-next-generation,” which it is capable of handling millions of DNA molecules simultaneously, including cDNAs from the RNAs.

Considering the many technologies that are already available or emerging, researchers can only venture in this world of possible and promising technologies. Various research groups

should seek to unite efforts in order to overcome the difficult and challenging task of applying the enormous potential of these new methods to advance and progress in penaeid shrimp aquaculture.

ACKNOWLEDGMENTS

The authors would like to thank the Fundação de Apoio a Pesquisa do Estado de São Paulo (FAPESP) for the financial support (2012/17322-8).

REFERENCES

- Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* 11, R106. doi: 10.1186/gb-2010-11-10-r106
- Argue, B. J., Arce, S. M., Lotz, J. M., and Moss, S. M. (2002). Selective breeding of Pacific White Shrimp (*Litopenaeus vannamei*) for growth and resistance to Taura Syndrome Virus. *Aquaculture* 204, 447–460. doi: 10.1016/S0044-8486(01)00830-4
- Banerjee, S., Ooi, M. C., Shariff, M., and Khatoon, H. (2012). Antibiotic resistant Salmonella and Vibrio associated with farmed *Litopenaeus vannamei*. *Sci. World J.* 2012, 6. doi: 10.1100/2012/130136
- Baranski, M., Gopikrishna, G., Robinson, N. A., Katneni, V. K., Shekhar, M. S., Shanmugakarthik, J., et al. (2014). The development of a high density linkage map for Black Tiger Shrimp (*Penaeus monodon*) based on cSNPs. *PLoS ONE* 9:e85413. doi: 10.1371/journal.pone.0085413
- Bell, C. G., and Beck, S. (2009). Advances in the identification and analysis of allele-specific expression. *Genome Med.* 1, 56. doi: 10.1186/gm56
- Brady, P., Elizur, A., Cummins, S. F., Ngyuen, N. H., Williams, R., and Knibb, W. (2013). Differential expression microarrays reveal candidate genes potentially associated with reproductive dysfunction of captive-reared prawn *Penaeus monodon*. *Aquaculture* 400–401, 14–28. doi: 10.1016/j.aquaculture.2013.02.038
- Cerdà, J., and Machado, M. (2013). Advances in genomics for flatfish aquaculture. *Genes Nutrition* 8, 5–17. doi: 10.1007/s12263-012-0312-318
- Chávez-Calvillo, G., Perez-Rueda, E., Lizama, G., Zúñiga Aguilar, J. J., Gaxiola, G., Cuzon, G., et al. (2010). Differential gene expression in *Litopenaeus vannamei* shrimp in response to diet changes. *Aquaculture* 300, 137–141. doi: 10.1016/j.aquaculture.2009.11.027
- Cock, J., Gitterle, T., Salazar, M., and Rye, M. (2009). Breeding for disease resistance of penaeid shrimps. *Aquaculture* 286, 1–11. doi: 10.1016/j.aquaculture.2008.09.011
- Gavery, M. R., and Roberts, S. B. (2012). Characterizing short read sequencing for gene discovery and RNA-Seq analysis in *Crassostrea gigas*. *Comp. Biochem. Physiol. Part D* 7, 94–99. doi: 10.1016/j.cbcd.2011.12.003
- Ghosh, D., and Qin, Z. S. (2010). Statistical issues in the analysis of ChIP-Seq and RNA-Seq data. *Genes (Basel)* 1, 317–334. doi: 10.3390/genes1020317
- 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. doi: 10.1016/j.aquaculture.2005.02.011
- Gucic, M., Cortés-Jacinto, E., Civera-Cerecedo, R., Ricque-Marie, D., and Martínez-Córdova, L. R. (2013). Apparent carbohydrate and lipid digestibility of feeds for whiteleg shrimp, *Litopenaeus vannamei* (Decapoda: Penaeidae), cultivated at different salinities. *Rev. Biol. Trop.* 61, 1201–1213. doi: 10.15517/rbt.v61i3.11935
- Guo, H., Ye, C.-X., Wang, A.-L., Xian, J.-A., Liao, S.-A., Miao, Y.-T., et al. (2013). Transcriptome analysis of the Pacific White Shrimp *Litopenaeus vannamei* exposed to nitrite by RNA-Seq. *Fish Shellfish Immunol.* 35, 2008–2016. doi: 10.1016/j.fsi.2013.09.019
- Howe, G. T., Yu, J., Knaus, B., Cronn, R., Kolpak, S., Dolan, P., et al. (2013). A SNP resource for douglas-fir: de novo transcriptome assembly and SNP detection and validation. *BMC Genomics* 14:137. doi: 10.1186/1471-2164-14-137
- Huang, C. W., Li, Y. H., Hu, S. Y., Chi, J. R., Lin, G. H., Lin, C. C., et al. (2012). Differential expression patterns of growth-related microRNAs in the skeletal muscle of Nile tilapia (*Oreochromis niloticus*). *J. Anim. Sci.* 90, 4266–4279. doi: 10.2527/jas.2012-5142
- Huberman, A. (2000). Shrimp endocrinology. *Rev. Aquac.* 191, 191–208. doi: 10.1016/S0044-8486(00)00428-2
- James, R., Thampuran, N., Lalitha, K. V., Rajan, L. A., and Joseph, T. C. (2010). Differential gene expression profile of the hepatopancreas of White Spot Syndrome Virus infected Fenneropenaeus indicus by Suppression Subtractive Hybridization. *Fish Shellfish Immunol.* 29, 884–889. doi: 10.1016/j.fsi.2010.07.029
- Khatri, P., Sirota, M., and Butte, A. J. (2012). Ten years of pathway analysis: current approaches and outstanding challenges. *PLoS Comp. Biol.* 8:e1002375. doi: 10.1371/journal.pcbi.1002375
- La Vega, E., Degnan, B. M., Hall, M. R., and Wilson, K. J. (2007). Differential expression of immune-related genes and transposable elements in Black Tiger Shrimp (*Penaeus monodon*) exposed to a range of environmental stressors. *Fish Shellfish Immunol.* 23, 1072–1088. doi: 10.1016/j.fsi.2007.05.001
- Li, C., Weng, S., Chen, Y., Yu, X., Lü, L., Zhang, H., et al. (2012). Analysis of *Litopenaeus vannamei* transcriptome using the next-generation DNA sequencing technique. *PLoS ONE* 7:e47442. doi: 10.1371/journal.pone.0047442
- Li, S., Zhang, X., Sun, Z., Li, F., and Xiang, J. (2013). Transcriptome analysis on Chinese Shrimp Fenneropenaeus chinensis during WSSV acute infection. *PLoS ONE* 8:e58627. doi: 10.1371/journal.pone.0058627
- Lo, T. S., Cui, Z., Mong, J. L. Y., Wong, Q. W. L., Chan, S. -M., Kwan, H. S., et al. (2007). Molecular coordinated regulation of gene expression during ovarian development in the penaeid shrimp. *Mar. Biotechnol.* 9, 459–468. doi: 10.1007/s10126-007-9006-4
- Marguerat, S., and Bähler, J. (2010). RNA-Seq: from technology to biology. *Cell. Mol. Life Sci.* 67, 569–579. doi: 10.1007/s00018-009-0180-6
- Mohd-Shamsudin, M. I., Kang, Y., Lili, Z., Tan, T. T., Kwong, Q. B., Liu, H., et al. (2013). In-depth transcriptomic analysis on giant freshwater prawns. *PLoS ONE* 8:e60839. doi: 10.1371/journal.pone.0060839
- O’Neil, S. T., and Emrich, S. J. (2013). Assessing de novo transcriptome assembly metrics for consistency and utility. *BMC Genomics* 14:465. doi: 10.1186/1471-2164-14-465
- Oshlack, A., Robinson, M. D., and Young, M. D. (2010). From RNA-Seq reads to differential expression results. *Genome Biol.* 11, 220. doi: 10.1186/gb-2010-11-12-220
- Qian, X., Ba, Y., Zhuang, Q., and Zhong, G. (2014). RNA-Seq technology and its application in fish transcriptomics. *OMICS* 18, 1–13. doi: 10.1089/omi.2013.0110
- Rojtinnakorn, J., Hirono, I., Itami, T., Takahashi, Y., and Aoki, T. (2002). Gene expression in haemocytes of Kuruma Prawn, *Penaeus japonicus*, in response to infection with WSSV by EST approach. *Fish Shellfish Immunol.* 13, 69–83. doi: 10.1006/fsim.2001.0382
- Sharov, V., Kwong, K. Y., Frank, B., Chen, E., Hasseman, J., Gaspard, R., et al. (2004). The Limits of Log-Ratios. *BMC Biotechnol.* 4:3. doi: 10.1186/1472-6750-4-3
- Sookruksawong, S., Sun, F., Liu, Z., and Tassanakajon, A. (2013). RNA-Seq analysis reveals genes associated with resistance to Taura Syndrome Virus (TSV) in the Pacific White Shrimp *Litopenaeus vannamei*. *Dev. Comp. Immunol.* 41, 523–533. doi: 10.1016/j.dci.2013.07.020
- Tan, T. T., Chen, M., Harikrishna, J. A., Khairuddin, N., Shamsudin, M., Maizatul, I., et al. (2013). Deep parallel sequencing reveals conserved and novel miRNAs in gill and hepatopancreas of Giant Freshwater Prawn. *Fish Shellfish Immunol.* 35, 1061–1069. doi: 10.1016/j.fsi.2013.06.017
- Wang, X.-W., Xu, J.-D., Zhao, X.-E., Vasta, G. R., and Wang, J.-X. (2014). A shrimp C-type lectin inhibits proliferation of the hemolymph microbiota by maintaining the expression of antimicrobial peptides. *J. Biol. Chem.* 289, 11779–11790. doi: 10.1074/jbc.M114.552307
- Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10, 57–63. doi: 10.1038/nrg2484. RNA-Seq
- Xue, S., Liu, Y., Zhang, Y., Sun, Y., Geng, X., and Sun, J. (2013). Sequencing and de novo analysis of the hemocytes transcriptome in *Litopenaeus vannamei* response to White Spot Syndrome Virus infection. *PLoS ONE* 8:e76718. doi: 10.1371/journal.pone.0076718
- Yang, L., Park, J., and Graveley, B. R. (2007). Splicing from the outside in. *Mol. Cell.* 27, 861–862. doi: 10.1016/j.molcel.2007.09.003

- Yu, Y., Wei, J., Zhang, X., Liu, J., Liu, C., Li, F., et al. (2014). SNP discovery in the transcriptome of White Pacific Shrimp *Litopenaeus vannamei* by next generation sequencing. *PLoS ONE* 9:e87218. doi: 10.1371/journal.pone.0087218
- Zeng, D., Chen, X., Xie, D., Zhao, Y., Yang, C., Li, Y., et al. (2013). Transcriptome analysis of Pacific White Shrimp (*Litopenaeus vannamei*) hepatopancreas in response to Taura Syndrome Virus (TSV) experimental infection. *PLoS ONE* 8:e57515. doi: 10.1371/journal.pone.0057515

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 15 June 2014; accepted: 11 August 2014; published online: 28 August 2014.

Citation: Santos CA, Blanck DV and de Freitas PD (2014) RNA-seq as a powerful tool for penaeid shrimp genetic progress. *Front. Genet.* 5:298. doi: 10.3389/fgene.2014.00298

This article was submitted to *Livestock Genomics*, a section of the journal *Frontiers in Genetics*.

Copyright © 2014 Santos, Blanck and de Freitas. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Capítulo 2

(Mineração das ESTs)

Manuscrito 1 - MINERAÇÃO DE GENES, SSRs E SNPs LOCALIZADOS NA CADEIA PRINCIPAL DE GENES DO TRANSCRIPTOMA DO CAMARÃO MARINHO *LITOPENAEUS VANNAMEI*

Resumo

Para a caracterização de marcadores moleculares em regiões codificadoras e identificação de genes candidatos potencialmente úteis para aplicação em estudos genéticos de *Litopenaeus vannamei* foram realizadas análises no transcriptoma desta importante espécie marinha. Sequências de mRNA de diferentes tecidos e estádios larvais de indivíduos cultivados de *L. vannamei* foram isoladas. Um total de 44.881 ESTs foi sequenciado, o que resultou em 2.984 *contigs* após a montagem *de novo*. A anotação automática realizada nas bases de dados nr-NCBI, SwissProt, InterPro e GO mostrou que 1.219 *contigs* aqui identificados apresentaram alta similaridade com sequências de genes já estabelecidas para várias espécies, incluindo *L. vannamei*. Desse total, 248 *contigs* foram descritos pela primeira vez para *L. vannamei*, de acordo com o banco de dados TSA (*Transcriptome Shotgun Assembly Sequence Database*). As análises de mineração de dados identificaram 1.983 SSR e 3.472 SNP. Um total de 83 *contigs* contendo SSR e 299 contendo SNP foi identificado dentro de regiões codificadoras para proteínas conhecidas. Estes dados estão agora disponíveis em <http://www.ncbi.nlm.nih.gov/genbank/tsa>.

Palavras-chave: *L. vannamei*, Região codificadora, microsatélites, Polimorfismos de Base Única.

Manuscript 1 - MINING GENES, SSRs AND SNPs WITHIN PROTEIN MAIN CHAIN FROM THE TRANSCRIPTOME OF *LITOPENAEUS VANNAMEI* MARINE SHRIMP

Camilla A. Santos, Pedro M. Galetti Jr and Patrícia D. Freitas

¹Departamento de Genética e Evolução, Universidade Federal de São Carlos, São Carlos, SP, Brazil.

Correspondence to: Via Washington Luiz km 235, Phone +55-16-3351 8309, Fax: +55-16-3351 8377.

Caixa Postal 676, CEP 13565-905

E-mail: camilla.alves@yahoo.com.br

Abstract

Litopenaeus vannamei marine shrimp transcriptome *in silico* analyses was performed to characterize genes, SSR and SNP within coding regions and identify candidate loci potentially useful for application in genetic studies of this important marine species. mRNA sequences were isolated from different tissues and larval stages of cultured *L. vannamei* specimens. A total of 44.881 EST were sequenced, which resulted in 2.984 contigs after *de novo assemble*. Automatic annotation performed in *nr-NCBI*, *SwissProt*, *InterPro* and *GO* databases showed that 1.219 contigs identified herein presented high similarity to gene sequences already established for several species, including *L. vannamei*. Out of this total, 248 contigs were described for the first time for *L. vannamei*, according to TSA (Transcriptome Shotgun Assembly Sequence Database). Data mining analyses identified 1.983 SSR and 3.472 SNP. A total of 83 contigs containing SSR and 299 containing SNP was detected within coding regions for known proteins. These data are now available at <http://www.ncbi.nlm.nih.gov/genbank/tsa>.

Key words: *L. vannamei*, Coding region, microsatellites, Single Nucleotide Polymorphisms.

1. Introduction

The development of transcriptome projects and *in silico* analysis have been accompanied by a progressive increase in the description of genes, microsatellites or Simple Sequence Repeats (SSR) and Single Nucleotide Polymorphisms (SNP) for several aquatic organisms, including fish (Siemon *et al.*, 2005), clam (Yu and Li, 2008; Yu *et al.*, 2009, Li *et al.*, 2010), crab (Jiang *et al.*, 2008) and shrimp (Wang *et al.*, 2005; Zhao and Li, 2007; Santos *et al.*, 2012). SSR and SNP markers within interest genes may be applicable for construction of genetic maps (Jones *et al.*, 2013; Baranski *et al.*, 2014), identification of QTL (Quantitative Trait Loci) (Ma *et al.*, 2012) and population polymorphism studies (Wang *et al.*, 2008; Mohanty *et al.*, 2012), in *L. vannamei* and taxonomically related species (Marques *et al.*, 2013).

Litopenaeus vannamei is one of the most important marine shrimp species worldwide mainly because of its relevance for the aquaculture activity. During the last two decades, genetics studies in *L. vannamei* have been developed, aiming the establishment of technologies to manage their stocks (Ma *et al.*, 2009). Nowadays, there is a relevant amount of transcriptome data available at public databases for this species (Li *et al.*, 2012; Ghaffari *et al.*, 2014). However, the majority of these data shows no information related to the presence of molecular markers, such as SNP and SSR.

In the present work, a transcriptome from the Pacific white shrimp *L. vannamei* were established and data mining analyses were performed for characterization of gene products and non-redundant SSR and SNP markers within coding regions. This information may be used for future genetic studies in *L. vannamei* and also in other taxonomically related shrimp species.

2. Material and Methods

2.1. cDNA libraries and sequencing

cDNA libraries of muscle (MC), ocular peduncle (OP) and hepatopancreas (HP) from captive adults, hatched larvae in nauplius (NP) and mysis (MY) stages, and eggs (EG) obtained immediately after spawning were separately constructed. Total RNA from each tissue sample was isolated using Trizol reagent (Life Technologies) and the mRNA was purified using PolyATtract mRNA Isolation System III kit (Promega). cDNA libraries were constructed using the Smart PCR cDNA Synthesis kit (Clontech). DYEnamic™ ET terminator Cycle Sequencing (GE Healthcare) reagent was used and the sequencing was carried out on MegaBACE 1000 DNA Analysis System (GE Healthcare).

2.2. Trimming and assembly

UniVec database (<http://www.ncbi.nlm.nih.gov/tools/vecscreen/univec/>) was used to trim rDNA and adaptor sequences. The assembly was performed using *De novo* assembly algorithm, available at Geneious software (Kearse *et al.*, 2012). Phred base calling algorithm (Ewing *et al.*, 1998) was implemented and Phred quality values (Q) lower than 20 (low quality) were cut-off from the analysis, and values higher than 40 were considered as high quality. Only reads with at least 80% of high quality base callings (%Q80) were considered.

2.3. Data mining for SSR and SNP

The contigs were submitted to search SSR in the Geneious software using the Phobos algorithm (Mayer, 2008). The analysis was performed considering total consensus and the established search parameters: imperfect search, repeat unit length ranging from 2 to 6 bp and perfection between 40% and 100%. SNP search was performed, considering the variant minimum

frequency of 20%, minimum e-value equal to 10^{-7} and considering the occurrence of polymorphisms inside and outside the codon domain regions.

2.4. Transcriptome annotation

The contigs were submitted to automatic annotation in nrNCBI (<http://www.ncbi.nlm.nih.gov/>), cross-referenced to GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>), Protein Data Bank (PDB, <http://www.rcsb.org/pdb/home/home.do>), SwissProt (<http://www.uniprot.org/>), Protein Information Resource (PIR, <http://pir.georgetown.edu/>) and Protein Research Foundation (PRF, <http://www.prf.or.jp/index-e.html>), in order to identify their respective gene product. The e-value cut-off considered was equal to $1.0E^{-5}$. Analyses also included the establishment of GO (Gene Ontology) terms, annotated up to third level was performed in BLAST2GO (Götz *et al.*, 2008), available at <http://www.blast2go.com/>. Sequence Analysis and Classification Database (InterPro, <http://www.ebi.ac.uk/interpro/>) was used to establish the protein family and domain. TSA (Transcriptome Shotgun Assembly Sequence Database) (<http://www.ncbi.nlm.nih.gov/genbank/tsa>) search was also performed to verify which sequences are new for *L. vannamei* species.

3. Results and Discussion

3.1. Transcriptome assembly and annotation

A total of 44.881 unigenes with an N50 length of 641 bp was produced. After trimming, 25.837 remained reads were assembled, resulting in 2.984 contigs. The reads evidencing the best quality and reliability were distributed along the libraries, as followed: 5.426 (HP), 5.026 (OP), 4.300 (NP), 3.815 (EG), 3.636 (MC), and 3.634 (MY) (Table 1). The number of EST per contigs ranged from 2 to 714 reads (Figure 1).

Table 1: Assembly summary of EST-transcriptome analyses from *L. vannamei*.

Number of Reads	Before Trimming	After Trimming
Total	44.881	25.837
Hepatopancreas (HP)	8.789	5.426
Ocular penduncle (OP)	8.624	5.026
Nauplius (NP)	7.400	4.300
Eggs (EG)	7.285	3.815
Muscle (MC)	6.458	3.636
Mysis (MY)	6.325	3.634
Contigs	-----	2.984
N50	-----	641 bp

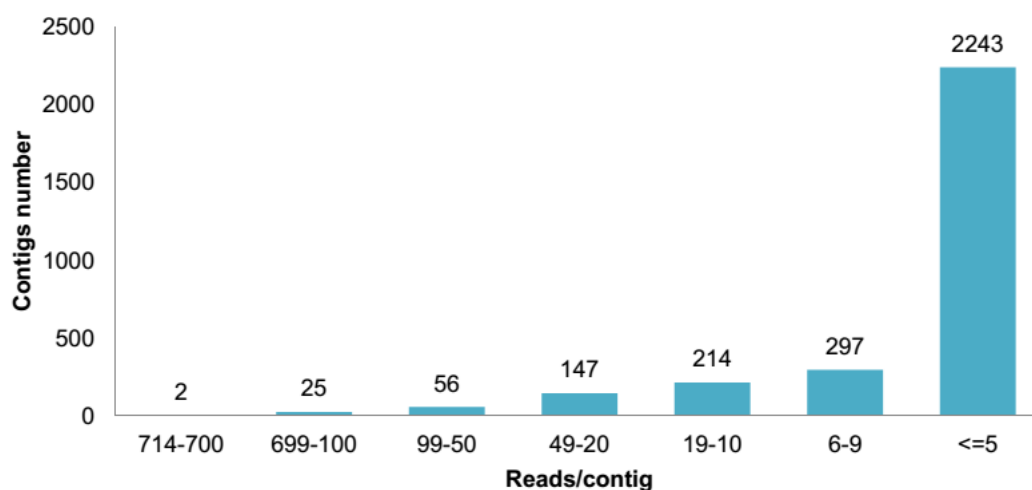


Figure 1: Reads number observed in contigs assembly.

The automatic annotation process implemented using nrNCBI and BLAST2GO tools identified gene products for 1.219 (41%) of 2.984 contigs. This percentage is higher when compared to some of the data available in the literature for shrimp EST annotation, such as 38% for *L. vannamei* (O'Leary *et al.*, 2006) and 16.7% for *Marsupenaeus japonicus* (Yamano and Unuma, 2006). Regarding similarity distribution, the majority of sequences showed values close to 90%. The

e-values ranged from highly reliable $1e^{-180}$ to the cut-off point equal to $1e^{-5}$ (Figure 2). The proteins family/domain could be identified for 77% of annotated proteins.

The transcriptome annotation data obtained herein showed a high similarity to proteins previously described for *L. vannamei*, *Penaeus monodon*, *Daphnia pulex*, and others (Figures 3 and 4). Recent analyses of transcriptomes, produced by Next Generation Sequencing (NGS), show also a high similarity between *L. vannamei* transcripts and proteins described for a wide variety of organisms, including *D. pulex* (Ghaffari *et al.*, 2014).

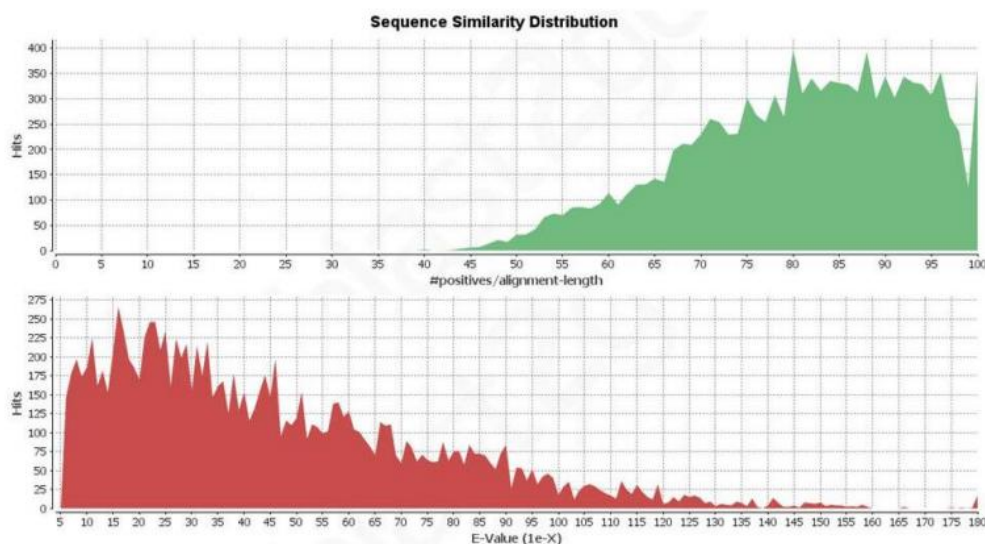


Figure 2: Sequence similarity distribution related to similarity (green) and e-value (red) values found in blastX hits.

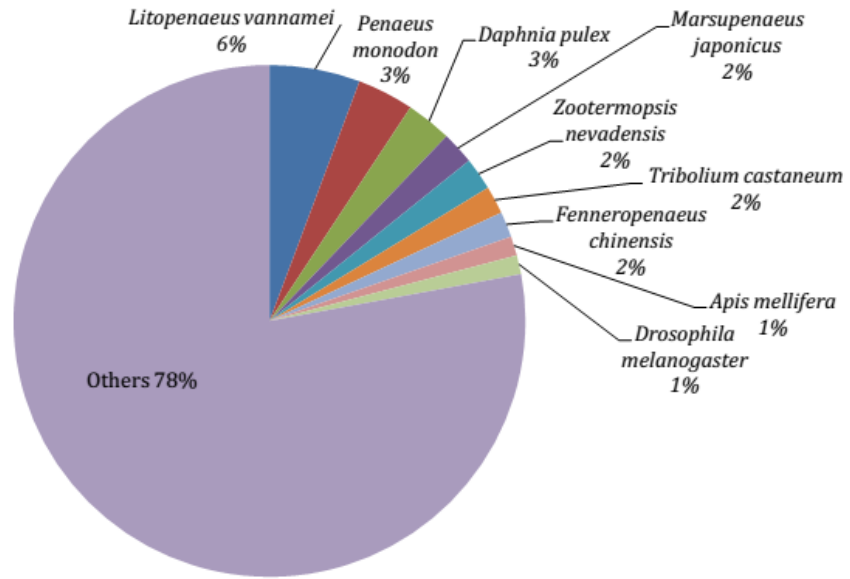


Figure 3: BlastX results from all-hits species distribution.

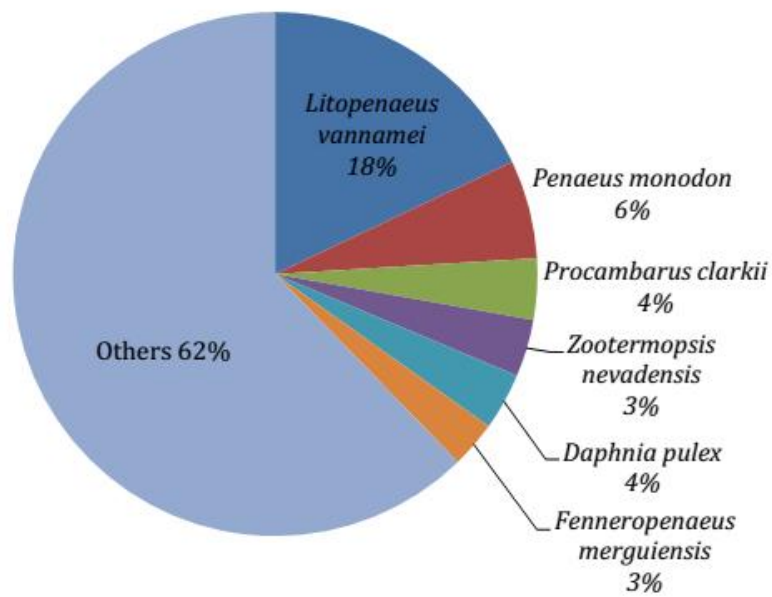


Figure 4: Blast results from top-hits species distribution.

The GO annotation, performed for 1.219 contigs with hits, established biological process, molecular function and cellular component for 836 contigs. For biological process, the proteins major participant steps are primary and organic metabolism, followed by cellular metabolism (Figure 5). Concerning the molecular function, a great variety of roles can be seen, highlighting ion, protein binding and oxidoreductase activity (Figure 6). For the cellular component, the proteins can be more frequently expressed at cell parts, non-membrane-bounded organelles, organelles in general and protein complex (Figure 7). The data found herein are in agreement with O’Leary *et al.* (2006), in which the main annotations for biological process, molecular function and cellular component are metabolism, binding and organelles, respectively.

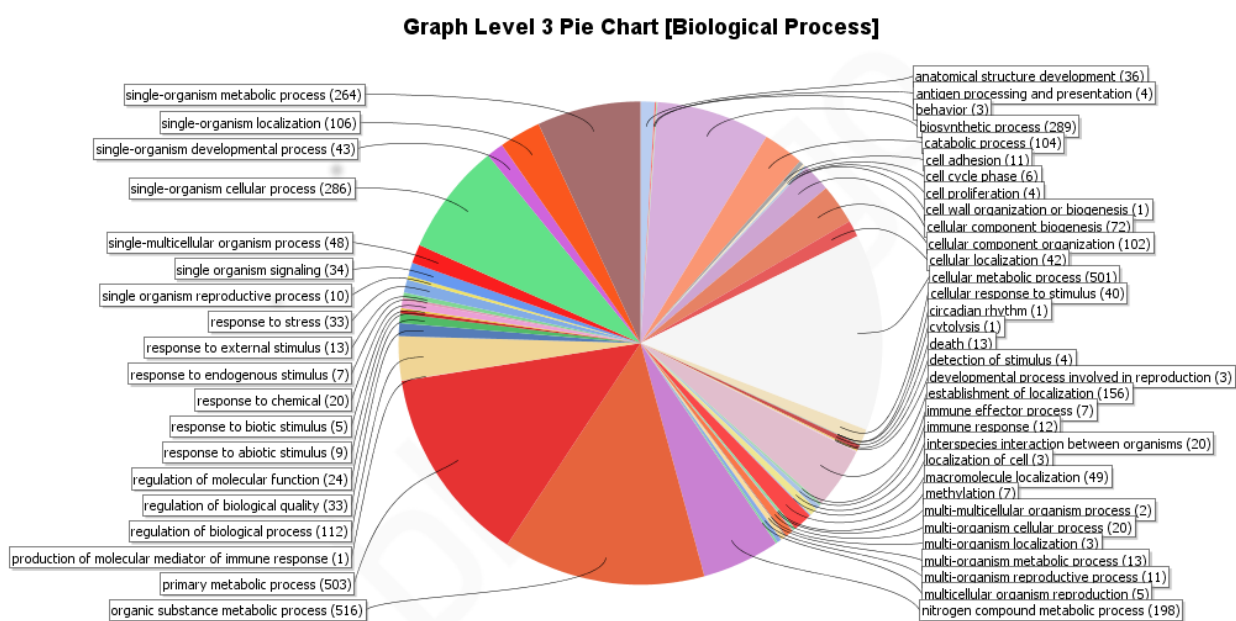


Figure 5: Biological processes identified in Gene Ontology (GO) database for the 2.984 contigs annotated.

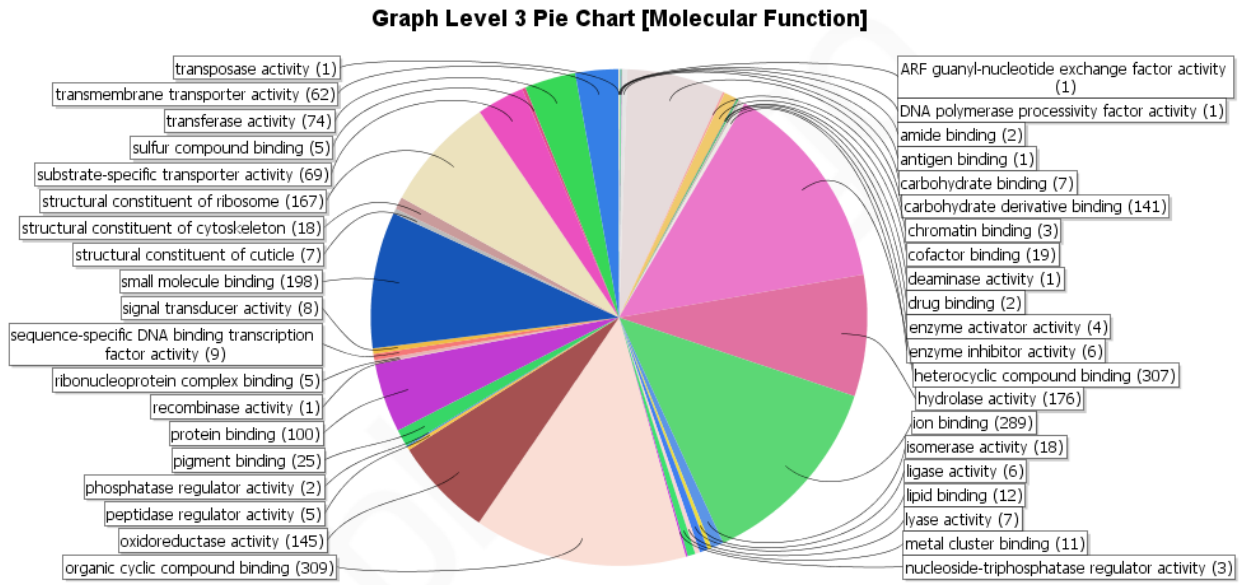


Figure 6: Molecular functions identified in Gene Ontology (GO) database for the 2.984 contigs annotated.

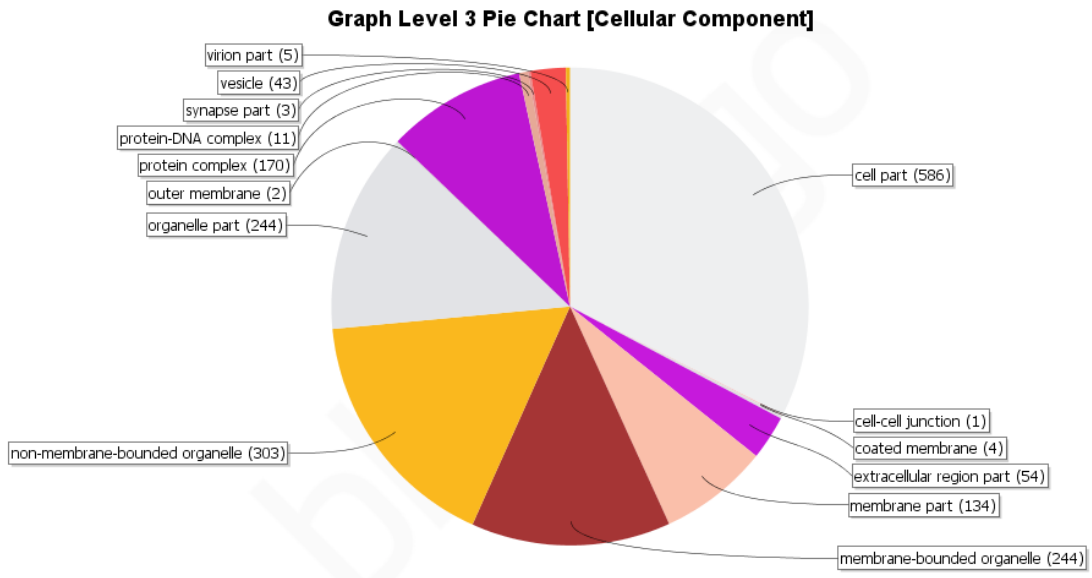


Figure 7: Cellular components identified in Gene Ontology (GO) database for the 2.984 contigs annotated.

A search in TSA NCBI database showed 2.507 contigs presenting hits. Out of the total contigs, 248 evidenced no similar sequences to *L. vannamei*, constituting new transcripts described for the first time for this species. This Transcriptome Shotgun Assembly project has been deposited at DDBJ/EMBL/GenBank under the accession GCVY00000000, BioProject: PRJNA77889, Sequence Read Archive: SRR1952625. The version described in this paper is the first version, GCVY01000000.1. Previously, close to 70 sequences were deposited at TSA with IDs ranging from JR494081 to JR494686 (Supplementary Material- Table S1).

3.2. Mining SSR within coding regions

The analysis for mining SSR evidenced 83 contigs presenting at least one SSR on their composition, amounting 1.983 di- to hexa-nucleotide SSR identified for *L. vannamei*. From this total, 1.656 were presented as perfect (83%) and 327 as compound (17%) microsatellites. Dinucleotide was the most abundant class of microsatellite, corresponding to approximately 72% of the total non-redundant SSR identified. Trinucleotide sequences were also frequent (18%), while tetra-, penta- and hexa-nucleotide microsatellites exhibited lower frequencies (7%, 2% and 1% respectively) (Table 2). The AC/AG dinucleotides were the most observed class. Among the trinucleotides, motifs containing AAG/ATC presented the highest frequency.

Table 2: Characterization of the identified non-redundant SSR and SNP loci through data mining analysis performed on transcriptome of *L. vannamei*.

	SSR		SNP/Indels
Total	1.983	Total	5.192
Compound	327	Indels	1.720
Perfect	1.656	SNP	3.472
2 bp	1.412	Transitions	1.144
3 bp	355	Transversions	2.328
4 bp	135	Di-allelic	3.436
5 bp	50	Tri-allelic	36
6bp	31	Tetra	00

In shrimp, previous data mining analysis searching for EST-SSR in *L. vannamei* (Perez *et al.*, 2005) *Fenneropenaeus chinensis* (Wang *et al.*, 2005), *P. monodon* (Maneeruttanarungroj *et al.*, 2006) and *M. japonicus* (Zhao *et al.*, 2007) showed different results. As reported by Perez *et al.* (2005) a predominance of trinucleotide repeats in the genome of *L. vannamei*, with motif (ATT)_n being the most observed. On the other hand, in *F. chinensis*, nearly 64% of the identified EST-SSR loci were dinucleotides, with (AT)_n motif as the most observed. In *P. monodon*, 40% SSR motif were trinucleotides and the (ATT)_n repeat was dominant, consistent with an earlier data suggesting that the genome of many crustacean species may be highly AT-rich (Wang *et al.*, 2005). However, mutation rate and DNA repair mechanism differences among different genomes could explain such variety of microsatellite motifs observed in the transcriptome of these species (Ellegren, 2004).

The fairly high frequency of trinucleotide microsatellites in expressed regions may be associated with a smaller perturbation degree that this type of repetition causes in the open reading frame regions. Indeed, some authors suggest that such repetitions may present a higher level of polymorphism than any other microsatellite type when in coding regions. In this manner, the coding amino acid remains the same and the variation found can only affect some protein properties, such as flexibility and binding affinity (Kashi *et al.*, 2006).

Furthermore, the predominance of few specific types of motifs in transcribed regions may be related to their gene expression roles. It has been suggested that specific SSR may promote a quicker adaptive advantage to specific environmental conditions, presenting a certain type of putative function (Li *et al.*, 2004). In the sugar cane genome, for instance, EST containing (ATT)₁₇ repeats seem to be related to proteins linked to the plant's resistance (Silva, 2001). In *P. monodon*, a (ATC)₄(TC)₃(TA)₄ microsatellite polymorphism was related to the WSSV (White Spot Syndrome

Virus) prevalence, being useful for disease-resistant population identification (Mukherjee and Mandal, 2009).

Despite of this important aspect, EST-SSR polymorphism studies in penaeids are scarce and most are limited to describe amplification success rates. In *P. monodon*, for instance, 82% of tested EST-SSR showed positive results in PCR (Maneeruttanarungroj *et al.*, 2006). For *F. chinensis*, approximately 60% of 30 loci were efficiently amplified [8]. In *L. vannamei*, 206 EST-SSR loci were established, with 51% of them producing satisfactory amplification results (Perez *et al.*, 2005). A similar result was obtained by Santos *et al.* (2012), for *L. vannamei*. From a total of 52 EST-SSR loci tested in PCR, 40% of them were polymorphic with an allelic diversity ranging from two to 20 and observed heterozygosity up to 86%. Moreover, when the 21 polymorphic loci were used in cross-species amplifications reactions, all of them showed successful results in at least one of the marine and freshwater shrimp species tested, demonstrating how this loci may be conserved between crustacean species, and may be used in the studies of other economical shrimp species (Marques *et al.*, 2013). Such a high level of transferability among taxonomically related species appears to be a common feature among shrimp species. Five out of nine *F. chinensis* EST-SSR loci tested were successfully amplified in *Fenneropenaeus merguensis*, *Fenneropenaeus penicillatus*, *L. vannamei* and *L. stylirostris*, demonstrating great transferability potential (Wang *et al.*, 2005).

Thereby, regarding EST-SSR loci and their high polymorphisms, cross-species amplification rates, and their association with putative molecular functions, including disease-resistance, these loci constitute an important source of genome information of *L. vannamei* and related species.

3.3. Mining SNP within coding regions

Data mining analysis revealed a great abundance of single nucleotide polymorphisms and indels (insertion/deletion) in the transcriptome of *L. vannamei*. One-base polymorphisms were

identified in 455 (15%) contigs. Close to 5.192 indels/SNPs were characterized, being 1.720 (33%) indels and 3.472 (67%) SNP. The most representative SNP class was di-allelic, and, although polymorphism confirmation requires verification after genotyping, all contigs containing SNP were composed by more than four EST, increasing the data reliability. Among the SNPs, 3.436 (99%) were di-allelics and 36 (1%) tri-allelics. Such result is expected, considering the frequency for single nucleotide substitutions at neutral positions around 1×10^{-9} /nucleotide and the low chance of two independent nucleotide mutations at the same position, (Brookes, 1999). From the total of di-allelic SNP, 64% were transitions and 36% transversions. Transitions tend to be more frequent in coding regions, where the selection pressure is stronger. When a nucleotide is replaced by another one of the same nature, the chance of this mutation passing through selection increases (Vignal *et al.*, 2002). Further studies, considering analyses of these polymorphisms may be carry out to investigate if these loci are under neutral, balanced or directional selection pressure.

The results observed here agree with some of the data found in literature, given that the transition and transversion rates reported by Yu *et al.* (2014) were 66.8% and 33.2%, respectively for *L. vannamei* transcriptome. The same is observed for the oyster *Crassostrea gigas*, which 76% of SNPs found were firstly classified as transitions and 24% as transversions (Wang *et al.*, 2009).

From the total of 445 contigs containing single polymorphisms, 299 showed blastX hits after manual annotation, consisting in SNP with known coding regions. The description of SNP located in genes is being highly efficient for polymorphisms characterization (Wang *et al.*, 2005; Marques *et al.*, 2013), metabolism studies (Guo *et al.*, 2013; Wang *et al.* 2014) and identification of QTL (Ma *et al.*, 2012), allowing the development of genetic improvement programs through MAS (Marker Assisted Selection). These markers may also be applied for differential expression analysis related to fitness traits (Vignal *et al.*, 2002; Liu and Cordes, 2004; Marti *et al.*, 2010) such as resistance to some pathogens (Zeng *et al.*, 2008).

When considering the association between complex traits, such as fitness, SNPs and their major di-allelic nature, it may be less informative when compared to multi-allelic loci, including SSR. However, when several SNPs are identified as being related to a trait, the haplotypes may explain more about the variation seen in phenotype. One-base variants tend to be more common in coding regions than tandem repeats and are likely responsible for differences observed in phenotypes (Vignal *et al.*, 2002).

Final considerations

In this study, the data mining evidenced 83 contigs containing SSR and 299 containing SNP. The data including their respective polymorphisms and gene products are now available at TSA database (<http://www.ncbi.nlm.nih.gov/genbank/tsa>). TSA IDs, e-value and similarity are available in Supplementary Material (Table A1 and A2). These markers are important tools for studies investigating the functional role of microsatellite sequences and single polymorphisms in the development, adaptation and survival of *L. vannamei* and related species, contributing to the better understanding of aspects related to structure and evolution of these sequences along the genome of eukaryotes. Besides, the *L. vannamei* transcriptome characterization provides potentially useful data for genetic studies of this important shrimp species and related species, aiding in the monitoring of genetic variability and studies involving aspects related to the construction of genetic maps, molecular function, metabolism and gene expression.

Supplementary Material

S1 Table. Contigs (83) containing SSR with gene product elucidated in automatic annotation. **(Disponível no arquivo anexado)**

S2 Table. Contigs (299) containing SNP with gene product elucidated in automatic annotation. **(Disponível no arquivo anexado)**

References

- Baranski M, Gopikrishna G, Robinson N, Katneni VK, Shekhar MS, Shanmugakarthik J *et al.*. The development of a high density linkage map for black tiger shrimp (*Penaeus monodon*) based on cSNPs. *PLoS One* 2014; 9(1): e85413.
- Brookes AJ. The essence of SNPs. *Gene* 1999; 234(2): 177–86.
- Duff R, Hoeh WR, Cook D, Weeks SC. Isolation and characterization of 13 polymorphic microsatellite loci from the clam shrimp *Eulimnadia texana* (Crustacea: Spinicaudata). *Molecular Ecology Notes* 2004; 4: 397-399.
- Ellegren H. Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics* 2004; 5: 435-445.
- Ewing B, Hillier L, Wendl MC, Green P. Base-Calling of Automated Sequencer Traces Using Phred. *Accuracy Assessment*. 1998; 175–185.
- Ghaffari N, Sanchez-Flores A, Doan R, Garcia-Orozco KD, Chen PL, Ochoa-Leyva A *et al.*. Novel transcriptome assembly and improved annotation of the whiteleg shrimp (*Litopenaeus vannamei*), a dominant crustacean in global seafood mariculture. *Scientific Reports* 2014; 4: 7081.
- Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ *et al.* High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research* 2008; 36(10): 3420–35.
- Guo H, Xian JA, Li B, Ye CX, Wang AL, Miao YT *et al.* Gene expression of apoptosis-related genes, stress protein and antioxidant enzymes in hemocytes of white shrimp *Litopenaeus vannamei* under nitrite stress. *Comparative Biochemistry and Physiology. Toxicology & Pharmacology* 2013; 157(4):366–71.
- Jiang H, Cai YM, Chen LQ, Zhang XW, Hu SN, Wang Q. Functional annotation and analysis of expressed sequence tags from the hepatopancreas of mitten crab (*Eriocheir sinensis*). *Marine Biotechnology* (New York, N.Y.) 2008; 11(3):317–26.
- Jones DB, Jerry DR, Khatkar MS, Raadsma HW, Zenger KR. A high-density SNP genetic linkage map for the silver-lipped pearl oyster, *Pinctada maxima*: a valuable resource for gene localisation and marker-assisted selection. *BMC Genomics* 2013; 14: 810.
- Kashi Y, King DG. Simple sequence repeats as advantageous mutators in evolution. *Trends in genetics* 2006; 22: 253-9.
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S *et al.* Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* (Oxford, England) 2012; 28(12):1647–9.
- Li C, Weng S, Chen Y, Yu X, Lu L, Zhang H *et al.* Analysis of *Litopenaeus vannamei* transcriptome using the next-generation DNA sequencing technique. *PLoS One* 2012; 7(10): e47442.
- Li H, Zhu D, Gao X. Mining single nucleotide polymorphisms from EST data of hard clam *Meretrix meretrix*. *Conservation Genetics Resources* 2010; 2:69-72.
- Li YC, Korol AB, Fahima TN. Microsatellite within genes: structure, function and evolution. *Molecular Biology and Evolution* 2004; 21: 991-1007.
- Liu Z.J & Cordes JF. DNA marker technologies and their applications in aquaculture genetics. *Aquaculture* 2004; 238(1-4), 1–37.
- Ma K, Qiu G, Feng J, Li J. Transcriptome analysis of the oriental river prawn, *Macrobrachium nipponense* using 454 pyrosequencing for discovery of genes and markers. *PLoS One* 2012; 7(6): e39727.

- Maneeruttanarungroj C, Pongsomboon S, Wuthisuthimethavee S, Klinbunga S, Wilson KJ, Swan J *et al.* 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* 2006; 37:363-368
- Marques CG, Santos CA, Galetti PM Jr., Freitas PD. Informative microsatellites for freshwater and marine shrimp species. *International Journal of Biodiversity and Conservation* 2013; 5:611–615.
- Marti SM, Onteru SK, Du ZQ, Rothschild MF. SNP analyses of the 5HT1R and STAT genes in Pacific white shrimp, *Litopenaeus vannamei*. *Spanish Journal of Agricultural Research* 2010; 8(1), 53–55.
- Mayer C. Phobos, a tandem repeat search tool for complete genomes. 2008.
- Mohanty P, Sahoo L, Parida K. Development of polymorphic EST-SSR markers in *Macrobrachium rosenbergii* by data mining. *Conservation Genetics Resources* 2013; 5(1): 133–136.
- Mukherjee K, Mandal N. A Microsatellite DNA Marker Developed for Identifying Disease-resistant Population of Giant Black Tiger Shrimp *Penaeus monodon*. *Journal of The World Aquaculture Society* 2009; 40: 274-280.
- O' Leary NA, Iii, Robalino J, Peck MET, Mckillen DJ, Gross PS. Analysis of multiple tissue-specific cDNA libraries from the Pacific whiteleg shrimp, *Litopenaeus vannamei*. *Comparative and General Pharmacology* 2006; 46(6): 931–939.
- Pérez F, Ortiz J, Zhinaula M, Gonzabay C, Calderón J, Volckaert FAMJ. Development of EST-SSR markers by Data Mining in three species of shrimp: *Litopenaeus vannamei*, *Litopenaeus stylirostris* and *Trachypenaeus biridy*. *Marine Biotechnology* 2005; 7:554-569.
- Santos CA, Rossini BC, Marques CG, Galetti PM, Freitas PD. Characterization and genomic annotation of polymorphic EST-SSR loci in *Litopenaeus vannamei* shrimp. *Aquaculture Research* 2012; 43:1567-1570.
- Siemon HNG, Chang A, Brown GD, Koop BF, Davidson WS. Type I microsatellite markers from Atlantic salmon (*Salmo salar*) expressed sequence tags. *Molecular Ecology Notes* 2005; 5:762-766.
- Silva JAG. Preliminary analysis of microsatellite markers derived from sugarcane expressed sequence tags (EST). *Genetics and Molecular Biology* 2001; 24:155-159.
- Vignal A, Milan D, San Cristobal MAE. A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics, Selection, Evolution* 2002; 34:275–305.
- Wang H, Li F, Xiang J. Polymorphic EST-SSR markers and their mode of inheritance in *Fenneropenaeus chinensis*. *Aquaculture* 2005; 249:107-114.
- Wang J, Qi H, Li L, Que H, Wang D, Zhang G. Discovery and validation of genic single nucleotide polymorphisms in the Pacific oyster *Crassostrea gigas*. *Molecular Ecology Resources* 2014; 123–135.
- Wang X, Guo X, Zhang Y, Meng X, Qiu X, Liu S *et al.* Development of polymorphic EST-derived SSR markers for the shrimp, *Fenneropenaeus chinensis*. *Conservation Genetics* 2008; 10(5):1455–1457.
- Wang X, Guo X, Zhang Y, Meng X, Qiu X., Liu S *et al.* Development of polymorphic EST-derived SSR markers for the shrimp, *Fenneropenaeus chinensis*. *Conservation Genetics* 2009; 1455–1457.
- Whankaew S, Tangphatsornruang S, Triwitayakorn K. Development of simple sequence repeat markers from expressed sequence tags of the black tiger shrimp (*Penaeus monodon*). *Molecular Ecology Resources* 2008; 8:1494-1496.
- Yamano K & Unuma T. Expressed sequence tags from eyestalk of kuruma prawn, *Marsupenaeus japonicus*. *Comparative Biochemistry and Physiology - A Molecular and Integrative Physiology* 2006; 143(2), 155–161.

- Yu H, Li Q. Exploiting EST Databases for the development and characterization of EST-SSR in the Pacific oyster (*Crassostrea gigas*). *Journal of Herdability* 2008; 1-7.
- Yu Y, Wei J, Zhang X, Liu J, Liu C, Li F, Xiang J. SNP discovery in the transcriptome of white Pacific shrimp *Litopenaeus vannamei* by next generation sequencing. *PLoS One* 2014; 9(1):e87218.
- Yu Z, Wang Y, Fu D. Development of Fifty-one novel EST-SSR loci in the Pacific oyster, *Crassostrea gigas* by data mining from the public EST database. *Conservation Genetic Resources* 2009 doi: 10.1007/s12686-009-9127-3
- Zeng D, Chen X, Li Y, Peng M, Ma N, Jiang W *et al.* Analysis of hsp70 in *Litopenaeus vannamei* and detection of SNPs. *Journal of Crustacean Biology* 2008; 28(4):727–730.
- Zhao Y, Li Q. Characterization of expressed sequence tag-derived microsatellites from the kuruma prawn, *Marsupenaeus japonicus*. *Molecular Ecology Notes* 2007; 7: 1248-1250.

Manuscrito 2 - IDENTIFICAÇÃO DE UM RICO CONJUNTO DE GENES RELACIONADOS A ESTRESSE E METABOLISMO DE ENERGIA ATRAVÉS DA ANÁLISE DE MARCAS DE SEQUENCIAS EXPRESSAS (EST) DO CAMARÃO *LITOPENAEUS VANNAMEI*

Resumo

A descoberta de novas funções gênicas tornou-se crucial para compreender a fisiologia dos crustáceos e as reações no metabolismo causadas por condições estressantes adversas. Dessa forma, os estudos com transcriptoma tornaram-se gradualmente mais eficientes, gerando uma grande quantidade de dados genéticos para serem processados com rapidez e precisão. Neste trabalho, 1.219 *contigs* de *Litopenaeus vannamei*, obtidos a partir do sequenciamento de Sanger, foram submetidos à anotação funcional manual para identificação de seus produtos gênicos e identificação de vias metabólicas do KEGG. Entre estes, 293 são enzimas, reunindo 541 *EC numbers*, com 127 únicos e inseridos em 94 diferentes vias mapeadas do KEGG, fornecendo dados informativos para futura análise funcional. As enzimas identificadas estão relacionadas principalmente a processos de transporte com função de ligantes de íons e com atuação nas mitocôndrias. Os principais produtos enzimáticos encontrados estão relacionados à proteção contra o estresse, seguidos daqueles participantes de processos abordando produção e gasto de energia. Estes achados contribuem para a caracterização do transcriptoma e podem ser úteis na compreensão do papel biológico dos tecidos utilizados na construção da biblioteca de cDNA, tais como músculo, ovos e larvas, envolvidos num metabolismo de energia rápida, além de hepatopâncreas e pedúnculo ocular.

Palavras-chave: *L. vannamei*, Enzimas, Anotação Funcional, Vias metabólicas, Xenobióticos.

Manuscript 2 - IDENTIFICATION OF A RICH GENE SET RELATED TO STRESS AND ENERGY METABOLISM BY ANALYSIS OF EXPRESSED SEQUENCE TAGS (EST) FROM *LITOPENAEUS VANNAMEI* SHRIMP

Camilla A. Santos¹, Pedro M. Galetti Jr.¹, Marcelo F. Carazzolle², Gonçalo G. Pereira² and Patrícia D. Freitas¹

¹Departamento de Genética e Evolução, Universidade Federal de São Carlos, São Carlos, SP, Brazil.

²Instituto de Biologia, Departamento de Genética e Evolução, Universidade Estadual de Campinas, Campinas, SP, Brazil.

Correspondence to: Camilla A Santos, Via Washington Luiz km 235, Phone +55-16-3351 8309, Fax: +55-16-3351 8377. Caixa Postal 676, CEP 13565-905, São Carlos-SP, Brazil.

E-mail: camilla.alves@yahoo.com.br

Abstract

Novel gene discovery and function have become crucial for understanding crustacean physiology and the reactions in metabolism caused by stressful and adverse conditions. For this purpose, transcriptome studies have gradually become more efficient, generating a large amount of genetic data to be processed fast and accurately. In this work, 1.219 contigs from *Litopenaeus vannamei* marine shrimp, with NCBI nr database hits, obtained from Sanger sequencing method, were submitted to manual functional annotation for accurate gene product and metabolic pathways identification. Among these, 293 are enzymes, gathering 541 EC numbers, with 127 unique and inserted in 94 different KEGG pathways, providing informative data for future functional analysis. The identified enzymes are mostly related to transport processes, binding roles and acting in the mitochondria. The main enzyme products found were those related to protection against stress, followed by those related to energy production and spent. These findings contribute to

transcriptome profiling and may be useful in understanding the biological role of the tissues used in the cDNA library construction, such as muscle, eggs and larvae, involved in a fast energy metabolism, in addition to hepatopancreas and ocular peduncle.

Keywords: *L. vannamei*, Enzymes, Functional Annotation, Pathways, Xenobiotics.

1. Introduction

Although there is an increasing amount of transcriptome data from next-generation sequencing technologies available for shrimp species (Ghaffari *et al.*, 2014; Yu *et al.*, 2014; Powell *et al.*, 2015), analysis of Expressed Sequence Tags (EST), based on transcriptome obtained from Sanger method, have still been the resource of a rich batch of information about many species (Du *et al.*, 2013; Liu *et al.*, 2013; Nirmala *et al.*, 2013), linking active coding genes to their expression products. This data also enables the identification of regulatory regions, such as transcriptional factors (Jung *et al.*, 2011; Davies *et al.*, 2012).

Functional annotation is performed in order to attribute a functional meaning for expressed sequences and identify their gene product and biological processes. Bioinformatics tools and databases (DBs) are queried to search for the most accurate information available. Enzyme Commission (EC) numbers can be obtained from different databases, such as Kyoto Encyclopedia of Genes and Genomes (KEGG) (www.genome.jp/kegg/) and ENZYME (<http://enzyme.expasy.org/>). KEGG gathers protein and functional information, enabling the connection between the genotypes and their respective phenotypes before environmental disturbances and providing a wider understanding of crustaceans metabolism (Kanehisa and Goto, 2000; Kanehisa *et al.*, 2008). Many pathways together can be seen as metabolic maps or networks that link genetic and biochemical data. For example, through an enzyme EC number and the potential pathways it is inserted it is possible to identify the type of chemical reaction performed by this enzyme and its interaction with

other molecules or cofactors enabling more efficient responses by the organisms (Reed *et al.*, 2006; Yamanishi *et al.*, 2009).

Transcriptome studies have been carried out in many organisms, including marine species, such as crab (Jiang *et al.*, 2008; Li *et al.*, 2014), oysters (Zhang *et al.*, 2014) and shrimp, including the *Litopenaeus vannamei* marine species (Pérez *et al.*, 2005; Ghaffari *et al.*, 2014). Identifying the protein and its respective coding gene have become necessary to provide more details about underlying mechanisms involving the crustacean survival and adaptation before disturbing factors. For example, stress caused by captive environments, high population densities, few feed availability and presence of toxics in the water tanks, which are frequent conditions in genetic improvement programs (Jiang *et al.*, 2008; Ma *et al.*, 2012). Therefore, the knowledge about the enzymes reactions and their roles in the cell metabolism may enhance the stocks productivity and decrease costs and biological losses, leading to a more sustainable farming (Muños and Escárate, 2014).

This approach may also be employed to improve selection and/or breeding strategies, such as Marker Assisted Selection (MAS) (Yue, 2013; Jones *et al.*, 2013). MAS consists in an approach particularly important for traits with low heritability levels and difficult to measure, such as those highly influenced by the environment, for example, egg production and stress resistance (Yue, 2014). In addition, the generated transcriptome data also enables to compare the physiological responses given by challenged and controlled animals during differential expression experiments (Zeng *et al.*, 2013). Therefore, the present work analyzed ESTs of *L. vannamei*, aiming to describe enzymatic products and their respective metabolic pathways, providing complementary transcriptome data about an important aquaculture species for which there is no reference genome available.

2. Material and Methods

2.1. EST Database

The Expressed Sequence Tags were originated from *L. vannamei* cDNA libraries constructed from the following tissues of captive adults: (i) muscle, (ii) ocular peduncle and (iii) hepatopancreas; (iv) nauplius and mysis (larvae stages) and (v) eggs were also sampled immediately after spawning also in captivity (Freitas *et al.*, unpublished data).

2.2. Functional annotation and pathways establishment

The sequences were first submitted to automatic blasts in nrNCBI database (<http://www.ncbi.nlm.nih.gov/>). Manually, SwissProt database (<http://www.uniprot.org/>) was queried to identify the codified protein and its function, EC numbers for enzymes and homologous organism information. Functional annotation of the Gene Ontology (GO) (www.geneontology.org) terms was done up to third level in order to obtain biological process, molecular function and cellular component. Sequence Analysis and Classification Database (InterPro, <http://www.ebi.ac.uk/interpro/>) was used to establish the protein family and domain. The considered e-value cut-off was equal to $1.0E^{-5}$. SwissProt and nrBLAST data were compared to verify if any contradiction remained in the results. In addition, other DBs such as ENZYME and KEGG were also queried. The gene product identification was only finished when the DBs annotation achieved a consensus.

After EC numbers identification, all enzymatic product sequences were submitted to analysis in KEGG. The participating pathways were visualized in Rapid Pathway Visualization Software 2.8 (PAICE), available at <http://sourceforge.net/projects/paice/>, under GPL v3 public license. Enzymes were classified according to their first EC digit number (ranging from one to six). Details about enzyme reaction, such as substrate and product were obtained from Enzyme Structures Database

(<http://www.ebi.ac.uk/thornton-srv/databases/enzymes/>). This Transcriptome Shotgun Assembly project has been deposited at DDBJ/EMBL/GenBank under the accession GCVY00000000, BioProject: PRJNA77889, Sequence Read Archive: SRR1952625.

3. Results and Discussion

3.1. Functional annotation and pathway identification

The 44.881 ESTs generated resulted in 2.984 contigs (Freitas *et al.*, unpublished data). From these, 1.219 contigs showed hits in NCBI nr database and were manually annotated for more accurate gene products and pathways identification. After functional annotation in KEGG database, 293 contigs (30%) were classified as enzymes. More than 98% of the identified enzymes products presented an e-value $< 10^{-10}$ and all of them showed identity (IDT) values higher than 60% (Figure 1). This Transcriptome Shotgun Assembly (TSA) project has been deposited at DDBJ/EMBL/GenBank under the accession GCVY00000000.1 (GCVY01000001-GCVY01000701).

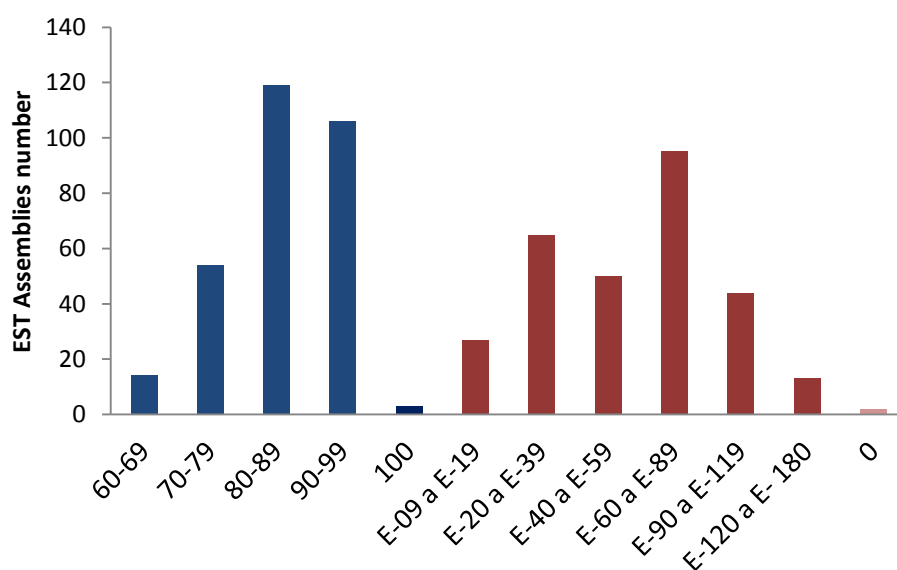


Figure 1: Identity (%) (blue) and E-value (red) distribution for the 293 enzyme products.

The GO annotation was carried out up to third level. The main categories found for the identified enzymes were metabolic processes and transport, for Biological Process; binding proteins, for Molecular Function; and mitochondria, for Cellular Component. From the 786 hits for Biological Process the most frequent ones were those regarding metabolic and catabolic steps (18%) ions, solute and ATP transport (17%), proteolysis (5%), followed by ATP synthesis and glycolysis, with 2% each. About the hits for GO Molecular Function (695), these were very diverse, highlighting binding proteins, ions and molecules (44%), oxidoreductases (12%), transferases (9%) and hydrolases (9%). For the Cellular Component (432), 17% of enzymes mostly act in mitochondria organelle or mitochondrial respiratory chain, followed by cytoplasm (14%), nucleus (7%) and extracellular (5%) (Figure 2). The complete GO annotation and EC numbers can be seen in Table S1 (Supplementary Data).

After comparing *L. vannamei* unigenes with the all matched terms in GO categories became clear a wide range of functions, biological processes and cellular components, since some genes were mapped to more than one term in each category. The unigenes majority is related to transport of several molecules, varied function, highlighting binding proteins, and takes place more often at the mitochondria. The enzymes were classified according to Enzymes Structure Database. They were distributed among the 293 products as the following catabolic reactions: 30% of oxidoreductases (1.-.-.), 15% of transferases (2.-.-.), 48% of hydrolases (3.-.-.), 2% of lyases (4.-.-.) and 5% of isomerases (5.-.-.).

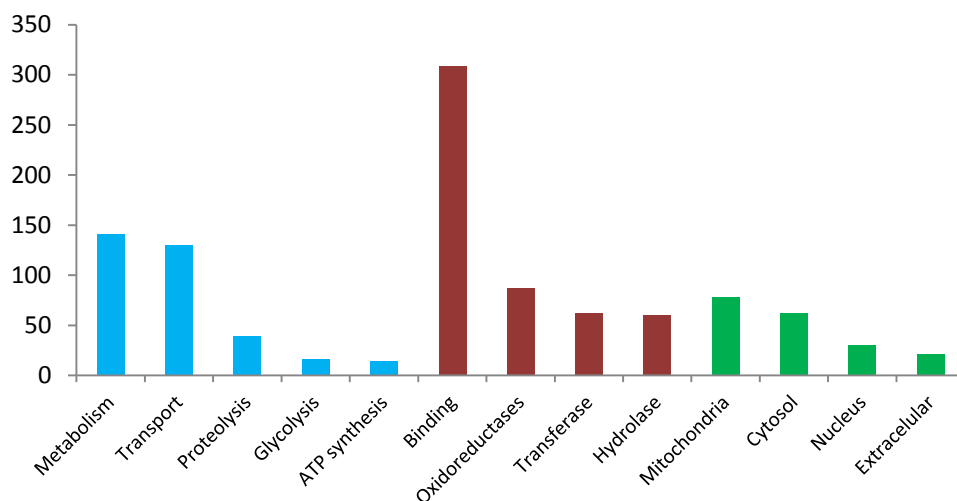


Figure 2: Some of the most representative hits from GO functional annotation for the 293 enzymatic products, highlighting Biological Process (blue bars), Molecular Function (red bars) and Cellular Component (green bars).

Ions and solutes transport proteins participate closely to oxidation-reduction activities inside the cell, including energy production and degradation activities in the mitochondria (Tubelvičiute *et al.*, 2014). The oxidoreductases, such as dehydrogenases and oxidases, were the most common functional roles, after binding (Figure 2). These enzymes are linked to survival reactions, such as immunity and protection against oxidative stress (Dinkova-Kostova and Talalay, 2010), reproduction, growth and energy production, besides some other metabolic actions linked to fitness (De Balsi *et al.*, 2014).

A total of 541 EC numbers were established from enzymatic products observed in 293 unigenes, resulting in 127 unique EC numbers and 94 pathways (Supplementary Material – Table S1). The most informative pathways are highlighted in Table 1. The most representative ones were related to (i) stress protection against reactive radicals and (ii) energy metabolism.

Table 1: Main enzymes present in most common pathways. Note molecular function and main participating pathways, based on KEGG database.

Enzyme/ EC number	Molecular Function	Pathway
Ferritin EC 1.16.3.1	Ferric iron binding/ Homeostasis/ Stress protection	Porphyrin Metabolism
GDP L-fucose synthase EC 1.1.1.271	Cellular metabolism	Fructose and mannose metabolism
S-formylglutathione hydrolase EC 3.1.2.12	Carboxylesterase	Methane metabolism
Fructose-1,6-biphosphatase EC 3.1.3.11	Fructose 1,6-biphosphato 1- phosphatase	Glycolysis/ Methane metabolism /Pentose pathway/ Krebs cycle
Phosphoenolpyruvate carboxykinase EC 4.1.1.32	Carboxyl-lyases	Krebs cycle
Fructose-bisphosphate aldolase EC 4.1.2.13	Fructose biphosphate aldolase	Glycolysis/ Fructose metabolism and mannose/ pentose pathway
Glucose-6-phosphate isomerase EC 5.3.1.9	Glucose-6-phosphatase	Glycolysis/ Pentoses pathway
Phosphoglycerate mutase EC 5.4.2.1	Transferase	Methane metabolism
Glutathione peroxidase EC 1.11.1.9	Glutathione peroxidase	Acid arachidonic metabolism / Glutathione metabolism
Pyruvate kinase EC 2.7.1.40	Pyruvate kinase and binding magnesium and potassium ions	Purine metabolism
Glutathione S-transferase EC 2.5.1.18	Glutathione transferase	Xenobiotics metabolism by P450 citochrome / Glutathione metabolism
Endochitinase EC 3.2.1.14	Hydrolase and chitinase	Amino and glycol metabolism
Cytochrome complex b-c1 EC 1.10.2.2	Cytochrome C reductase	Oxidative phosphorylation
NADH dehydrogenase complex EC 1.6.5.3/ 1.6.99.3	NADH desidrogenase	Oxidative phosphorylation
Cytochrome c oxidase EC 1.9.3.1	Iron/copper binding and cytochrome-c activity	Oxidative phosphorylation

3.3. Stress protection

Metabolism of xenobiotics by cytochrome P450 was the most complete pathway observed. Its relevance is linked to the wide use of organochlorines (OCs), organobromines (OBs) and organophosphates (OPs), in aquaculture, and their constant threat to shrimp homeostasis,

causing cellular metabolic stress damage. Although crustaceans do not have their own immune system, hepatopancreas plays this role and stress levels may be estimated in mitochondria through differential expression of enzymes involved in Reactive Oxygen Species (ROS) elimination (Robalino *et al.*, 2009; Masci *et al.*, 2014).

Oxidative stress happens when ROS levels are high, changing cellular metabolism and damaging the organelles (Lushchak, 2011). These radical production may lead to several DNA damages, such as deletions, mutations and other lethal genetic effects (Lesser, 2006; Sharma *et al.*, 2012). Disturbances, such as temperature and salinity levels variation may result in physiological and oxidative stress in organisms, increasing the production of radicals and triggering cellular damage (Tu *et al.*, 2008; Du *et al.*, 2013).

The main enzyme involved in metabolism of xenobiotics by cytochrome P450 cells identified here is glutathione transferase (2.5.1.18). It enables the use of glutathione to capture some ROS from oxidation, minimizing the damage caused to the cell. Crustaceans are one of the most efficient biological converters in nature (Rewitz *et al.*, 2006), however, many of these exotic compounds are transformed into still toxic chemicals (Davies *et al.*, 2012; Won *et al.*, 2013). Glutathione may work as a cofactor for antioxidant enzymes, such as glutathione peroxidase (1.11.1.9) and glutathione transferase.

In a study conducted by Zhou *et al.* (2010), glutathione transferase was the most efficient antioxidant enzyme for protecting against oxidation when considering environment temperature variation (heat shock). In this case, *L. vannamei* individuals were exposed to heat shock from 28° to 15°C during 6 hours. Additionally, other enzymes may also act on the protection against ROS, such as superoxide dismutase (SOD) (1.15.1.1), characterized as an efficient biomarker for the penaeid group before the stress caused by heat shock. This could be observed when *L. vannamei* individuals were submitted to salinity variation for 24 hours, causing SOD and glutathione

peroxidase activities to increase significantly because of oxidative stress (Lushchak *et al.*, 2011). Therefore, it is possible to establish these two enzymes together as the main responsible for protecting the cells against oxidative damages (Tu *et al.*, 2008; Zhou *et al.*, 2010), among the ones identified in this study. Another antioxidant enzyme found herein was ferritin (1.16.3.1) that may influence in animal survival after a long period of hypoxia, protecting the cells from stress radicals (Larade and Storey, 2004).

Several antioxidant enzymes were expected to be described in the present study. This fact can be observed probably due to animal tissues selected to construct the cDNA library, such as hepatopancreas (Figure 3). Muscle, ocular peduncle and hepatopancreas tissues were originated from captive animals, increasing the stress levels in the organism. When facing this environment, the organism is submitted to stress conditions, such as high population density levels and less food availability, which triggers a metabolic oxidative stress, leading to an increase of ROS rates and antioxidant enzymes expression levels, as a defense (Lesser, 2006; Liu *et al.*, 2007; Tian *et al.*, 2011).

As noticed, the oxidative stress process brings many damages to shrimp fitness in general. Therefore, it becomes harder for aquatic organisms to control the physiological equilibrium when any type of exotic compound is increased in the water. This may be related to the great amount of energy dispended by the organisms to keep their homeostasis, being the energy wasted inversely proportional to the energy remaining for growth (Lesser, 2006; Tu *et al.*, 2010). Therefore, all enzymatic products identified herein may help to achieve a clearer interpretation of the responses observed in aquatic organisms under stress conditions.

3.4. Energy production

Another pathway with many EC numbers identified here is glycolysis, The enzymes found, such as glucose-6-phosphato isomerase (5.3.1.9), converting glucose 6-phosphate in fructose 6-phosphate and fructose-1,6-biphosphatase (3.1.3.11) were related to production and/or degradation of carbohydrates and energy. They were responsible for energy release by breaking phosphate bonds. Glucose level control is central in crustaceans and its concentration is well defined along the day. In spite of this, this group can tolerate a large gradient. During the day, sugar level is low and stable and in the afternoon, it increases until reaches a peak after dusk (Sathyanandam *et al.*, 2008). Higher glucoses levels are also seen in premolt individuals when compared with postmolt or intermolt ones (Pratoomthai *et al.*, 2012).

Mathew *et al.* (2007) observed *Penaeus monodon* individuals infected by WSSV and some enzyme related to energy production had their activity monitored, highlighting the possible response caused by pathogen stress. Twenty-four hours after the infection, the amount of food ingested by the shrimp had decreased significantly. Consequently, glucose 6-phosphatase (5.3.1.9) and fructose 1,6-diphosphatase (1.1.1.37) activities decreased during the infection period. This resulted is an energetic crisis because gluconeogenesis is severely inhibited by the low levels of the last two enzymes, leading many animals to death. Carbohydrates availability in the cell and the manner some enzymes degrade them releasing energy, deeply influence organism homeostasis and survival (Pratoomthai *et al.*, 2012).

Another enzyme related to carbohydrate metabolism is endochitinase (3.2.1.14). To digest this saccharide, chitinase breaks chitin into smaller sugars, and the glycosidase (3.2.1.-) is responsible for hydrolyzing saccharide to monosaccharide. A better understanding about endochitinase role may be useful in a more efficient growth control in shrimp, being applied in genetic improvement programs. Additionally, endochitinase enzyme acts on chitin catabolism

and directly in crustacean molt process (Proespraiwong *et al.*, 2010). This enzyme is closely related to growth which is only possible in the intermolt stage. Molt, known as the process when chitin is degraded and the animal can grow, is a relevant step for arthropods homeostasis.

In this manner, knowing the enzymes and the pathway in which they are inserted, allow more efficient control on key processes in shrimp performance. Nevertheless, when analyzing all enzymatic products identified in this study, physiological and environmental variables must be considered as well, given that these may also cause stress. Considering this, the best is to take into account all of these factors in order to obtain a more reliable metabolic response (Tu *et al.*, 2010 Jubeaux *et al.*, 2012).

3.5. Most expressed genes

Thousands of ESTs (4.184) were used for the assembly of the 293 enzymatic products. Among the genes which presented hundreds of ESTs on their composition, the most expressed ones were those related to cytochrome oxidase subunits enzymes (Figure 3), with 593 reads in only one contig assembly. The most representative EST classes under analysis were obtained from larvae stages (mysis and nauplius) (1.194 ESTs) and eggs (950) cDNA libraries, showing a high abundance of respiratory proteins and other types that are essential for vital actions and demand a lot of metabolic energy, mostly represented by cytochrome oxidase subunit enzymes (1.9.3.1). This is expected given that both tissues require a significant amount of ATP due to the intense metabolism assigned to development stages.

A representative amount of EST was also originated from muscle library (766), characterized mainly for actin (3.6.1.3) and myosin (3.6.1/ 3.6.1.15) enzymes, related to muscle contraction and maintenance. Regarding the data obtained from hepatopancreas (649), were observed enzymes related to stress protection, such as glutathione transferase and glutathione

peroxidase, involved in the elimination of ROS. The ocular peduncle (625) showed the predominance of proteins related to respiratory chain, such as cytochrome oxidase and ribosomal (4.2.99.18) ones. As can be observed in the literature, this tissue is related to reproduction maturation in crustaceans (Asusena *et al.*, 2012), with an active metabolism, probably being these one of the many reasons responsible for the predominance of oxidative phosphorylation and translation proteins (Figure 3).

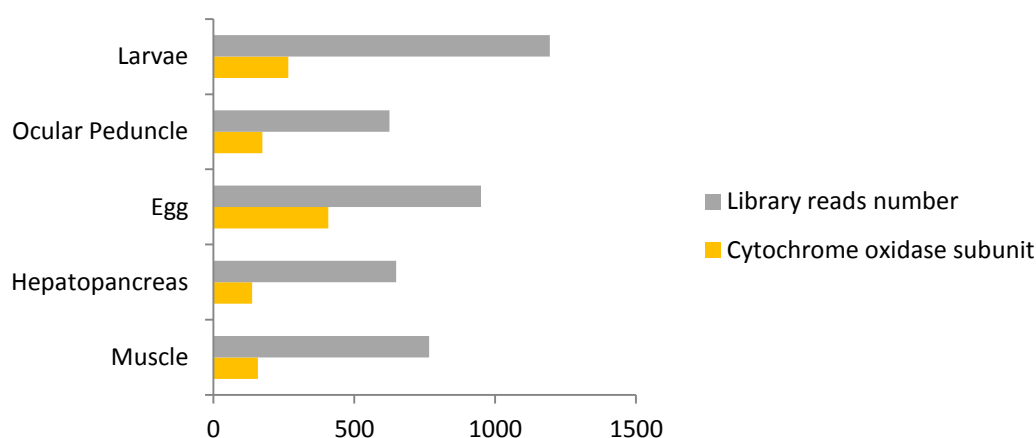


Figure 3: Total EST number (4.184) used in the assembly of the 293 contigs with enzymatic products (gray bars). EST number present in the more expressed gene (cytochrome oxidase) assembly in the data annotated (orange bars).

Final considerations

The amount of relevant data generated by transcriptome projects leads to functional analysis to provide a biological meaning to raw data, increasing the available information related to gene expression and metabolism. In this context, the present work contributes to analysis of transcriptome profiling and gene function in crustacean group. The data gathered here might be useful in understanding the biological role of some stress related tissues, such as hepatopancreas, as well as some involved in a fast energy production and spent, such as muscle, eggs and larvae.

Therefore, this data can be applied in approaches that consider aspects related to mechanisms involving stress adaptation, innate immunity and reproduction in shrimp.

Acknowledgements

This work was supported by Associação Brasileira de Criadores de Camarão (ABCC), Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). The authors are thankful to ShEST Project (*Litopenaeus vannamei* Genome EST Project) by the availability of the EST database.

Supplementary Material

Table S1: List of identified EC numbers for the 293 contigs. **(Disponível no arquivo anexado)**

References

- Asusena, A.-C. J. (2012). The Effects of Eyestalk Ablation on the Reproductive and Immune Function of Female *Macrobrachium americanum*. *Journal of Aquaculture Research & Development*, 03(8), 156.
- Davies, M. N., Volta, M., Pidsley, R., Lunnon, K., Dixit, A., Lovestone, S., Mill, J. (2012). Functional annotation of the human brain methylome identifies tissue-specific epigenetic variation across brain and blood. *Genome Biology*, 13(6), R43.
- De Balsi, K. L., Wong, K. E., Koves, T. R., Slentz, D. H., Seiler, S. E., Wittmann, A. H., Muoio, D. M. (2014). Targeted metabolomics connects thioredoxin-interacting protein (TXNIP) to mitochondrial fuel selection and regulation of specific oxidoreductase enzymes in skeletal muscle. *The Journal of Biological Chemistry*, 289(12), 8106–20.
- Dinkova-Kostova, A. T., & Talalay, P. (2010). NAD(P)H:quinone acceptor oxidoreductase 1 (NQO1), a multifunctional antioxidant enzyme and exceptionally versatile cytoprotector. *Archives of Biochemistry and Biophysics*, 501(1), 116–23.
- Du, Z., Zhu, L., Dong, M., Wang, J., Wang, J., Xie, H., & Zhu, S. (2012). Effects of the ionic liquid [Omicron]PF6 on antioxidant enzyme systems, ROS and DNA damage in zebrafish (*Danio rerio*). *Aquatic Toxicology (Amsterdam, Netherlands)*, 124-125, 91–3.
- Du, F. K., Xu, F., Qu, H., Feng, S., Tang, J., & Wu, R. (2013). Exploiting the transcriptome of Euphrates Poplar, *Populus euphratica* (Salicaceae) to develop and characterize new EST-SSR markers and construct an EST-SSR database. *PLoS One*, 8(4), 1-11.

- Ghaffari, N., Sanchez-Flores, A., Doan, R., Garcia-Orozco, K. D., Chen, P. L., Ochoa-Leyva, A., ... Criscitiello, M. F. (2014). Novel transcriptome assembly and improved annotation of the whiteleg shrimp (*Litopenaeus vannamei*), a dominant crustacean in global seafood mariculture. *Scientific Reports*, *4*, 7081.
- Howe, G. T., Yu, J., Knaus, B., Cronn, R., Kolpak, S., Dolan, P., Dean, J. F. D. (2013). A SNP resource for Douglas-fir: de novo transcriptome assembly and SNP detection and validation. *BMC Genomics*, *14*(1), 137. doi:10.1186/1471-2164-14-137
- Kanehisa M & Goto S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, *28*(1), 27-30.
- Kanehisa M., Araki M., Goto S., Hattori M., Hirakawa M., Itoh M., Katayama T. (2008). KEGG for linking genomes to life and the environment. *Nucleic acids research*, *36*.
- Jiang, H., Cai, Y.-M., Chen, L.-Q., Zhang, X.-W., Hu, S.-N., & Wang, Q. (2008). Functional annotation and analysis of expressed sequence tags from the hepatopancreas of mitten crab (*Eriocheir sinensis*). *Marine Biotechnology (New York, N.Y.)*, *11*(3), 317–26.
- Jones, D. B., Jerry, D. R., Khatkar, M. S., Raadsma, H. W., & Zenger, K. R. (2013). A high-density SNP genetic linkage map for the silver-lipped pearl oyster, *Pinctada maxima*: a valuable resource for gene localisation and marker-assisted selection. *BMC Genomics*, *14*, 810.
- Jubeaux, G., Simon, R., Salvador, A., Quéau, H., Chaumot, A., & Geffard, O. (2012). Vitellogenin-like proteins in the freshwater amphipod *Gammarus fossarum* (Koch, 1835): functional characterization throughout reproductive process, potential for use as an indicator of oocyte quality and endocrine disruption biomarker in males. *Aquatic Toxicology (Amsterdam, Netherlands)*, *112-113*, 72–82.
- Jung, H., Lyons, R. E., Dinh, H., Hurwood, D. a, McWilliam, S., & Mather, P. B. (2011). Transcriptomics of a giant freshwater prawn (*Macrobrachium rosenbergii*): de novo assembly, annotation and marker discovery. *PLoS One*, *6*(12), 1-14.
- Larade, K., & Storey, K. B. (2004). Accumulation and translation of ferritin heavy chain transcripts following anoxia exposure in a marine invertebrate. *The Journal of Experimental Biology*, *207*(Pt 8),
- Lesser M. P. (2006). Oxidative stress in marine environments: biochemistry and physiological ecology. *Annual review of physiology*, *68*(3), 253-78.
- Li, E., Wang, S., Li, C., Wang, X., Chen, K., & Chen, L. (2014). Transcriptome sequencing revealed the genes and pathways involved in salinity stress of Chinese mitten crab, *Eriocheir sinensis*. *Physiological Genomics*, *46*(5), 177–90.
- Liu Chun.-H., Tseng Mei-C., Cheng W. (2007). Identification and cloning of the antioxidant enzyme, glutathione peroxidase, of white shrimp, *Litopenaeus vannamei*, and its expression following *Vibrio alginolyticus* infection. *Fish & shellfish immunology*, *23*(1), 34-45.
- Liu, H., Li, S., Hu, P., Zhang, Y., & Zhang, J. (2013). Isolation and characterization of EST-based microsatellite markers for *Scatophagus argus* based on transcriptome analysis. *Conservation Genetics Resources*, *5*(2), 483–485.
- Lushchak V. I. (2011). Environmentally induced oxidative stress in aquatic animals. *Aquatic toxicology (Amsterdam, Netherlands)*, *101*(1), 13-30.
- 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*(6), e39727.

- Mathew S., Kumar A., Nair K., Anandan R., Gopalan P., Nair N. V. (2007). Biochemical studies on changes associated with enzymes of glucose metabolism in white spot syndrome virus (WSSV) infected with *Penaeus monodon* (Fabricius). *Journal of Biotechnology*, 6, 1944-1948.
- Masci, M., Orban, E., & Navigato, T. (2014). Organochlorine pesticide residues: an extensive monitoring of Italian fishery and aquaculture. *Chemosphere*, 94, 190–8.
- Muñoz, V., & Gallardo-Escárate, C. (2014). TLR and IMD signaling pathways from *Caligus rogercresseyi* (Crustacea: Copepoda): in silico gene expression and SNPs discovery. *Fish & Shellfish Immunology*, 36(2), 428–34.
- Nirmala, X., Schetelig, M. F., Yu, F., & Handler, A. M. (2013). An EST database of the Caribbean fruit fly, *Anastrepha suspensa* (Diptera: Tephritidae). *Gene*, 517(2), 212–7.
- Perez, F., Ortiz, J., Zhinaula, M., Gonzabay, C., Caldero, J., Pe, F., & Volckaert, F. A. M. J. (2005). Development of EST-SSR Markers by Data Mining in Three Species of Shrimp : *Litopenaeus vannamei* , *Litopenaeus stylirostris* , and *Trachypenaeus birdy*, 7, 554–569.
- Pratoomthai, B., Sakaew, W., Udomkit, A., Wongprasert, K., Chang, E. S., & Withyachumnarnkul, B. (2012). Decreased level of crustacean hyperglycemic hormone (CHH) in black tiger shrimp *Penaeus monodon* suffering from Monodon Slow-Growth Syndrome (MSGs). *Aquaculture*, 350-353, 19–25.
- Proespraiwong P., Tassanakajon A., Rimphanitchayakit V. (2010). Chitinases from the black tiger shrimp *Penaeus monodon*: phylogenetics, expression and activities. *Comparative biochemistry and physiology. Part B, Biochemistry & molecular biology*, 156(2), 86-96.
- Reed J. L., Famili I., Thiele I., Palsson B. O. (2006). Towards multidimensional genome annotation. *Nature reviews. Genetics*, 7(2), 130-41.
- Robalino, J., Carnegie, R. B., O’Leary, N., Ouvry-Patat, S. a, de la Vega, E., Prior, S., Gross, P. S. (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(1-3), 110-8.
- Rewitz K. F., Styrihave B., Løbner-olesen A., Andersen O. (2006). Marine invertebrate cytochrome P450: Emerging insights from vertebrate and insect analogies. *Comparative biochemistry and physiology. Part A, Molecular & integrative physiology*, 143, 363 - 381.
- Sathyanandam S., Vasudevan S., Natesan M. (2008). Serotonin modulation of hemolymph glucose and crustacean hyperglycemic hormone titers in *Fenneropenaeus indicus*. *Aquaculture*, 281(1-4), 106-112.
- Sharma, V., Anderson, D., & Dhawan, A. (2012). Zinc oxide nanoparticles induce oxidative DNA damage and ROS-triggered mitochondria mediated apoptosis in human liver cells (HepG2). *Apoptosis : An International Journal on Programmed Cell Death*, 17(8), 852–70.
- Tian, J., Chen, J., Jiang, D., Liao, S., & Wang, A. (2011). Transcriptional regulation of extracellular copper zinc superoxide dismutase from white shrimp *Litopenaeus vannamei* following *Vibrio alginolyticus* and WSSV infection. *Fish & Shellfish Immunology*, 30(1), 234–40.
- Tu H., Silvestre F., Bernard A., Douny C., Phuong N., Tao C. (2008). Oxidative stress response of black tiger shrimp (*Penaeus monodon*) to enrofloxacin and to culture system. *Aquaculture*, 285(1-4), 244-248.

- Tu H. T., Silvestre F., Phuong N. T., Kestemont P. (2010). Effects of pesticides and antibiotics on penaeid shrimp with special emphases on behavioral and biomarker responses. *Environmental toxicology and chemistry / SETAC*, 29(4), 929-38.
- Tubeleviciute, A., Teese, M. G., & Jose, J. (2014). Escherichia coli kduD encodes an oxidoreductase that converts both sugar and steroid substrates. *Applied Microbiology and Biotechnology*.
- Won, E.-J., Rhee, J.-S., Shin, K.-H., Jung, J.-H., Shim, W. J., Lee, Y.-M., & Lee, J.-S. (2013). Expression of three novel cytochrome P450 (CYP) and antioxidative genes from the polychaete, *Perinereis nuntia* exposed to water accommodated fraction (WAF) of Iranian crude oil and Benzo[α]pyrene. *Marine Environmental Research*, 90, 75–84.
- Yamanishi Y., Hattori M., Kotera M., Goto Susumu, Kanehisa Minoru. (2009). E-zyme: predicting potential EC numbers from the chemical transformation pattern of substrate-product pairs. *Bioinformatics*, 25(12), 179-86.
- Yue, G. H. (2013). Recent advances of genome mapping and marker-assisted selection in aquaculture. *Fish and Fisheries*, 1–21.
- Yue, G. H. (2014). Recent advances of genome mapping and marker-assisted selection in aquaculture. *Fish and Fisheries*, 15(3), 376–396.
- Zeng, D., Chen, X., Xie, D., Zhao, Y., Yang, C., Li, Y., Chen, X. (2013). Transcriptome analysis of Pacific white shrimp (*Litopenaeus vannamei*) hepatopancreas in response to Taura syndrome Virus (TSV) experimental infection. *PLoS One*, 8(2), 1-8.
- Zhang, L., Li, L., Zhu, Y., Zhang, G., & Guo, X. (2014). Transcriptome analysis reveals a rich gene set related to innate immunity in the Eastern oyster (*Crassostrea virginica*). *Marine Biotechnology (New York, N.Y.)*, 16(1), 17–33.
- Zhou J., Wang L., Xin Y., Wang W., He W., Wang, A. (2010). Effect of temperature on antioxidant enzyme gene expression and stress protein response in white shrimp, *Litopenaeus vannamei*. *Journal of Thermal Biology*, 35(6), 284-289.

Manuscrito 3 - ESTUDOS DE POLIMORFISMOS LOCALIZADOS NA CADEIA PRINCIPAL DE GENES DE PROTEÍNAS DO CAMARÃO *LITOPENAEUS VANNAMEI* E AVALIAÇÃO DE AMPLIFICAÇÃO HETERÓLOGA

Resumo

Os marcadores moleculares vêm mostrando sua ampla utilidade em estudos de diversidade genética e associação entre genótipos e fenótipos. Neste trabalho, foram selecionados quatro genes minerados em *Litopenaeus vannamei*, três contendo microssatélites (serina arginina, 60S ribossomal e troponina) e um contendo SNP (crustacianina), para avaliar uma possível relação entre genótipos e fenótipos de peso em duas famílias de camarão. Os resultados mostraram que os locos 60S ribossomal e crustacianina foram polimórficos, porém, não foi constatada associação entre os genótipos desses locos e o peso dos animais para a população amostrada. Os testes de amplificação heteróloga utilizando os marcadores de microssatélites mostraram sucesso para a maioria das espécies avaliadas. Embora os marcadores testados não tenham mostrado associação entre os genótipos e fenótipos de peso na população amostrada, podem ser usados como marcadores potenciais em outros estudos genéticos para *L. vannamei* e outras espécies relacionadas.

Palavras-chave: *L. vannamei*, 60S ribossomal, crustacianina, troponina, serina-arginina.

Manuscript 3 – MAIN CHAIN PROTEIN POLYMORPHISMS STUDIES IN GENES FROM *LITOPENAEUS VANNAMEI* AND CROSS-SPECIES AMPLIFICATION EVALUATION IN OTHER PRAWN SPECIES

Camilla Alves Santos¹, Ana Karina Texeira², João Luís Rocha³, Ana Carolina Guerrelhas³ and Patrícia Domingues de Freitas¹

¹Departamento de Genética e Evolução, Universidade Federal de São Carlos, São Carlos, SP, Brazil.

²Genearch Aquacultura, Pititinga, Rio do Fogo, RN, Brazil

³Aquatec Larvicultura de Camarão Marinho, Barra do Cunhaú, Canguaretama, RN, Brazil.

Correspondence to: Camilla Alves Santos¹, Via Washington Luiz km 235, Phone +55-16-3351 8309, Fax: +55-16-3351 8377. Caixa Postal 676, CEP 13565-905, São Carlos-SP, Brazil.

E-mail: camilla.alves@yahoo.com.br

Abstract

Molecular markers have been showing their wide utility in studies of genetic diversity and also association between genotypes and phenotypes. In this paper, four genes mined for *Litopenaeus vannamei*, three containing microsatellite (serine arginine, 60S ribosomal and troponin) and one containing SNP (crustacyanin), were selected for evaluating a possible relation between genotypes and weight phenotypes in two shrimp closed lines. The results showed that 60S ribosomal and crustacyanin loci were polymorphic, however, the genotype effect was not significant for the variance on total body weight. Cross-species amplification tests using the microsatellite markers showed efficient amplification for the majority of the evaluated species. Although the markers tested have showed no association between the genotypes and weight phenotypes in the population sampled, they may be used as potential markers in further genetic studies for this shrimp and other related species.

Keywords: *L. vannamei*, 60S ribosomal, crustacyanin, troponin, serine-arginine

Short Communication Format

Molecular markers, such as Simple Sequence Repeats (SSR) or microsatellites and Single Nucleotide Polymorphisms (SNP) within genes, have showed their wide application in aquaculture in the last decade (Jiang *et al.*, 2008; Yu and Li, 2008). The association between genotypes and fitness traits can be highlighted for several aquaculture species, including the description of a SSR region linked to resistance to White Syndrome Spot Virus (WSSV) disease in *Penaeus monodon* (Mandal and Mukherjee, 2009), and SNPs associated to survival and resistance identified in *Litopenaues vannamei* (Zeng *et al.*, 2008, Marti *et al.*, 2010). Another advantage related to polymorphism studies in coding regions is their high transferability rates among closely related species. Several studies have been describing the potential of use of SSR and SNP linked to Expressed Sequence Tags (EST) shrimp species which genome is still poorly known (Perez *et al.*, 2005, Freitas *et al.*, 2007, Santos *et al.*, 2012).

In this work we evaluated polymorphism in *L. vannamei* shrimp for three microsatellite loci and one SNP located within candidate genes related to immune response and growth. All molecular markers were located into the protein main chain. We also investigated a possible association between the genotypes and the total body weight phenotypes in two captive *L. vannamei* families challenged for high growth performance rates. In addition we performed cross-species amplification tests for the microsatellite loci in seven other marine shrimp: *Metapenaeus monocerus*, *Penaeus monodon*, *Fenneropenaeus indicus*, *Xiphopenaeus kroyeri*, *Farfantepenaeus brasiliensis*, *Litopenaeus schmitti* and *Farfantepenaeus paulensis*; and two freshwater ones: *Macrobrachium amazonicum* and *Macrobrachium jelskii*.

The genes serine arginine, 60S ribosomal, troponin and crustacyanin were obtained from the ShEST Project available at TSA database BioProject PRJNA77889, under access number

GCVY01000000.1. Sets of specific primer pairs were constructed for each loci using Primer 3 (Rozen and Skaletsky, 1999) software, considering the flanking regions for the SSR and SNP (Table 1).

Table 1: Information about the four loci tested in association tests. Note the marker type, TSAID, the gene product, function and primers sequence and melting temperature.

Locus/type	Protein	ID	Function	Primers	T (°C)
Trop/SSR	Troponin	JR494085	Muscle contraction	F:AGCCTGTGTACGAGGAGGAA R:AAAATAACGCGACGACATCC*	64
SerArg/SSR	Serine-arginine splicing factor	JR494104	Controls alternative splicing	F: GATATGGCAGGCTTCGTGAT R: TACCTGTCAAGCCAAGCT*	58
60Srib/SSR	60S ribosomal	JR494145	Exonuclease activity (E.C. = 4.2.99.18). Translation repair	F CAGCTACTGACATCAACTTC* R CTCCTTCTTGGCTTCCTCTCT	58
Crust/SNP	Crustacyanin	JR494393	Immune response	F ATCGAAATGAGCGGCTACAA R: TGTGTGTCAAGCTAG	55

*primer with M13 complementary

Polymerase Chain Reaction (PCR) were performed for DNA samples of 200 Specific Pathogen Free (SPF) shrimps from Genearch Genetic Improvement Company (Rio Grande do Norte State, Brazil). The animals belonged to families challenged for high survival and growth rates. Pleopods of 100 animals (50 males and 50 females) from two families (F1 and F3) were sampled and stored in alcohol 100% at -20° C. The animal whole body weight was measured by analytical balance and sex was identified by visualization of the external morphology. The samples of each family were separated in two groups: heavier lineage (50 animals - 25 females and 25 males - with the best weight rates) and lighter lineage (50 animals - 25 females and 25 males - with the worst rates) (Figure 1) (Table 2).

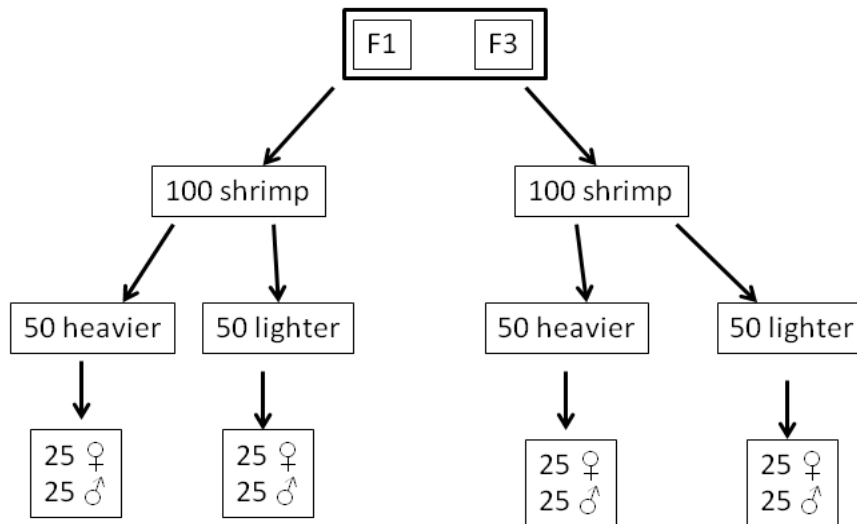


Figure 1: The schema used in shrimp population sampling.

Table 2: Weight mean and standart deviation (SD) values observed in families, groups and sex, followed by the whole animal body weight for the 200 animals collected.

	Families		Lineages		Sex		Total animal body weight
	F1	F3	Low	High	M	F	
Means (g)	14.5	15.1	11.8	17.8	14.7	14.9	14.8
SD (g)	3.64	3.04	1.91	0.75	3.13	3.58	3.36

The samples from the nine shrimp species used in cross-amplification tests were obtained from the Tissue Bank at the Laboratory of Molecular Biodiversity and Conservation (University of Sao Carlos, São Carlos, SP, Brazil). DNA extraction was performed using the phenol protocol proposed by Sambrook *et al.* (1989). PCR was carried out in 10 μ l of mix containing 50 ng/ μ l DNA, 0.2 mM dNTPs, 1 x PCR buffer (200 mM Tris-HCl, pH 8.4 and 500 mM KCl; Invitrogen Life Technologies), 1.5 mM MgCl₂, 1 U of *Taq* DNA Polymerase, 8 pmoles/ μ l of reverse primer, 2 pmoles/ μ l of the M13-tailed forward primer and 8 pmoles/ μ l of the fluorescently-labeled M13 primer for SSR loci (Schuelke, 2000). The M13 primer was not used for the locus with SNP marker. The thermal cycler was programmed for an initial denaturation of 5 min at 95 °C followed by 35

cycles of 30 s at 94 °C, 30 s at the annealing temperature, 30 s at 72 °C and a final extension of 20 min at 72 °C. The annealing temperatures were established using gradient tests that were performed in a Mastercycler Gradient Thermal Cycler (Life Technologies). Alleles were genotyped in Illumina Automatic Sequencer and scored using the Geneious software (Kearse *et al.*, 2012). Gene product was identified using blastN and blastX against nrNCB, SwissProt (UniProt Consortium, 2014) InterPro (Hunter *et al.*, 2009) and GO (Ashburner *et al.*, 2000) databases. The PCR products from the cross-species amplification were also sequenced and blastN (nrNCBI) searches were performed in order to confirm the protein and the species it belonged to. Alleles frequencies were estimated using Genepop version 3.4. The genotypic frequencies and Variance Analysis (ANOVA) were calculated using GLM (*General Linear Model*) Procedure employed in SAS software, version 9.3. The effects from family, sex and genotypes from each locus were considered. Sex was taken as a fixed effect and family and genotypes as random ones. The dependent variable was animal whole body weight, and p-value < 0.05 was assumed for data significance.

The alleles and genotypes frequencies for the four loci selected for genotype x phenotype association tests are in Table 3. For the 60S ribosomal protein (60Srib), the frequencies of 333 bp and 239 bp were 73% and 27%, respectively. As can be noted, the genotype 333/333 (66%) is more frequent in F3 whereas 239/333 (62%) and 239/239 (71%) in F1. Considering the SNP marker in crustacyanin gene (Crust) the C allele frequency is 94% and for allele T is 6%. The genotypes are equally distributed between families F1 and F3 (Table 3). The SSR and SNP markers in troponin and serine-arginine genes were removed from the association testing analysis, given that the first one is monomorphic and the second shows the genotype 594/594 in 98% of total animals and the other two remaining combinations in much lower frequencies, not enough for the statistical strength needed.

Table 3: The allelic and genotypic frequencies for the two polymorphic loci used in genotypes x phenotypes tests.

60Srib (Alleles/Frequency)		Crust (Alleles/Frequency)			
333	73%	C		94%	
239	27%	T		6%	
60Srib (Genotypes/Weight mean)	F1	F3	Crust (Genotypes/ Weight mean)	F1	F3
333/333 15.3 g	34%	66%	CC 15.0 g	51%	49%
239/333 14.6 g	62%	38%	CT 13.9 g	52%	48%
239/239 13.8 g	71%	29%	-	-	-

The effects significance of family and lineages were tested together nested, called here as Test A, along with sex and genotypes effects for each loci individually. Lineages are known as the heavier and lighter animals from each of the two families sampled. Only the family and lineage effect on the total variance observed for weight was significant (p-value <0.001). The determination coefficient (R^2) was close to 83% for all loci, and family and lineage shows high Mean Square (MS) rates, which highlight the great effect of families and lineages on weight variation for the model assumed (Table 4).

Table 4: Variance Analysis (ANOVA) results for the effects tested. Note the great influence of family and lineage (Family A mean squares) in the model used.

Loci	Family and Lineage	Sex	Locus	R²
Test A				
60Srib (Pr > F)	<0.001	0.42	0.81	0.82
Mean Square	557.10	1.26	0.41	
Crust (Pr > F)	<0.001	0.32	0.15	0.83
Mean Square	583.76	1.83	3.76	
Test B				
60Srib (Pr > F)	<0.001	0.42	0.81	0.82
Mean Square	557.10	1.26	0.41	
Crust (Pr > F)	<0.001	0.32	0.15	0.83
Mean Square	583.76	1.83	3.76	

Given that no significance in genotypes means were observed in the previous analysis, the two families and lineages were subdivided in four groups (Test B), to verify whether the weight variance could be partially assigned to the genotypes. The following animals represented the four groups called Test B: (i) F1 higher and (ii) lower weight and (iii) F3 higher and (iv) lower weight values animals. Although, the results found for Test B were exactly the same as those observed in Test A (Table 4), resulting in no association between the genotypes found and the weight of the collected animals, in the population analyzed. Some possible interactions between the two loci or among the loci and the other effects were also tested, but no association could be observed. However, in the case of a quantitative trait, such as weight, which is usually influenced by several genes of small effects each, it is expected that the few loci analyzed in this study explain too little or nothing of the weight variance observed (Zhang *et al.*, 2012; Jung *et al.*, 2013). Additionally, this variance may be assigned to several factors, such as environmental variance in phenotypes expression, bias in the sampling etc (Visscher *et al.*, 2008). Hence, further studies are recommended to perform association analyses for these genes in larger populations.

Regarding the cross-species amplification, the best rates were found for the SSRs into the SerArg and 60Srib genes (89% each). This can be justified due to these proteins participate of main metabolic steps, such as splicing control and repairs in translation, tending to be more conserved. For *P. monodon*, *F. indicus*, *F. brasiliensis* and *F. paulensis* positive amplification patterns for all SSR loci were observed (Table 5).

Table 5: Cross amplification results for three *L. vannamei* loci tested in seven marine and two fresh water shrimp species.

	SerArg	Trop	60Srib	Success
<i>Penaeus monodon</i>	X	X	X	100%
<i>Fenneropenaeus indicus</i>	X	X	X	100%
<i>Farfantepenaeus brasiliensis</i>	X	X	X	100%
<i>Farfantepenaeus paulensis</i>	X	X	X	100%
<i>Metapenaeus monocerus</i>	X	-	X	66%
<i>Litopenaeus schmitti</i>	-	X	X	66%
<i>Xiphopenaeus kroyeri</i>	X	-	-	33%
<i>Macrobrachium jelskii</i>	X	-	X	66%
<i>Macrobrachium amazonicum</i>	-	-	-	0%
Success	89%	55%	89%	

This finding may be assigned to *Farfantepenaeus* and *Litopenaeus* genders being more close to each other taxonomically (Baldwin *et al.* 1998; Maggioni *et al.* 2001). Therefore, even though SerArg and Trop markers had not been polymorphic for the population used in the association tests, they may be potential markers for species populations which present a small amount or even no molecular markers described yet, such as the ones used here.

Acknowledgements

This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP- 2012/17322-8 e 2012/13069-6), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

References

- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., & Harris, M. A. (2000). Gene Ontology: tool for the unification of biology. *Nature genetics*, 25(1), 25-29.
- Baldwin, J. D., Bass, a L., Bowen, B. W., & Clark, W. H. (1998). Molecular phylogeny and biogeography of the marine shrimp *Penaeus*. *Molecular Phylogenetics and Evolution*, 10(3), 399-407. <http://doi.org/10.1006/mpev.1998.0537>
- Freitas, P. D., Calgaro, M. R., & Galetti Jr, P. M. (2007). Genetic diversity within and between broodstocks of the white shrimp *Litopenaeus vannamei* (Boone, 1931) (Decapoda, Penaeidae) and its implication for the gene pool conservation. *Brazilian Journal of Biology = Revista Brasileira de Biologia*, 67(4 Suppl), 939-43. Retrieved from

<http://www.ncbi.nlm.nih.gov/pubmed/18278363>

- Hunter, S., Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Binns, D., ... & Finn, R. D. (2009). InterPro: the integrative protein signature database. *Nucleic acids research*, 37(suppl 1), D211-D215.
- Jiang H, Cai YM, Chen LQ, Zhang XW, Hu SN, Wang Q. Functional annotation and analysis of expressed sequence tags from the hepatopancreas of mitten crab (*Eriocheir sinensis*). *Marine Biotechnology* (New York, N.Y.) 2008; 11(3):317–26.
- Jung, H., Lyons, R. E., Hurwood, D. a., & Mather, P. B. (2013). Genes and growth performance in crustacean species: A review of relevant genomic studies in crustaceans and other taxa. *Reviews in Aquaculture*, 5(2), 77–110. <http://doi.org/10.1111/raq.12005>
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., ... Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics (Oxford, England)*, 28(12), 1647–9. <http://doi.org/10.1093/bioinformatics/bts199>
- Maggioni, R., Rogers, a D., Maclean, N., & D’Incao, F. (2001). Molecular phylogeny of western Atlantic Farfantepenaeus and Litopenaeus shrimp based on mitochondrial 16S partial sequences. *Molecular Phylogenetics and Evolution*, 18(1), 66–73. <http://doi.org/10.1006/mpev.2000.0866>
- Marti, S. M., Onteru, S. K., Du, Z. Q., & Rothschild, M. F. (2010). Short communication . SNP analyses of the 5HT1R and STAT genes in Pacific white shrimp , Litopenaeus vannamei. *Spanish Journal of Agricultural Research*, 8(1), 53–55.
- Mukherjee, K., & Mandal, N. (2009). A microsatellite DNA marker developed for identifying disease-resistant population of giant black tiger shrimp, Penaeus monodon. *Journal of the World Aquaculture Society*, 40(2), 274–280. <http://doi.org/10.1111/j.1749-7345.2009.00250.x>
- Perez, F., Ortiz, J., Zhinaula, M., Gonzabay, C., Caldero, J., Pe, F., & Volckaert, F. A. M. J. (2005). Development of EST-SSR Markers by Data Mining in Three Species of Shrimp : Litopenaeus vannamei , Litopenaeus stylirostris , and Trachypenaeus birdy, 7, 554–569. <http://doi.org/10.1007/s10126-004-5099-1>
- Rozen, S., & Skaletsky, H. (1999). Primer3 on the WWW for general users and for biologist programmers. *Bioinformatics methods and protocols*, 365-386.
- Santos, C. a, Rossini, B. C., Marques, C. G., Galetti, P. M., & Freitas, P. D. (2011). Characterization and genomic annotation of polymorphic EST-SSR loci in Litopenaeus vannamei shrimp. *Aquaculture Research*, 1–4. <http://doi.org/10.1111/j.1365-2109.2011.02956.x>
- Sambrook J., Fritshi E. F. *Molecular Cloning: A Laboratory Manual*. 2. ed. New York: Cold Spring Laboratory, 1989.
- Schuelke, M. (2000). An economic method for the fluorescent labeling of PCR fragments A poor man ’ s approach to genotyping for research and high-throughput diagnostics . *PRism*, 18(February), 1–2.
- UniProt Consortium. (2014). UniProt: a hub for protein information. *Nucleic acids research*, gku989.
- Visscher, P. M., Hill, W. G., & Wray, N. R. (2008). Heritability in the genomics era — concepts and misconceptions. *Nature Reviews Genetics*, 9(4), 255–266. <http://doi.org/10.1038/nrg2322>
- Yu H, Li Q. Exploiting EST Databases for the development and characterization of EST-SSR in the Pacific oyster (*Crassostrea gigas*). *Journal of Herdability* 2008; 1-7.
- Zeng, D., Zeng, D., Chen, X., Chen, X., Li, Y., Li, Y., Ma, N. (2008). ANALYSIS OF HSP70 IN LITOPENAEUS VANNAMEI AND DETECTION OF SNPS Digang Zeng, Xiaohan Chen, Yongmei Li, Min Peng, Ning Ma, Weiming Jiang, Chunling Yang,

and Ming Li, 28(4), 727–730.

Zhang, H., Wang, Z., Wang, S., & Li, H. (2012). Progress of genome wide association study in domestic animals. *Journal of Animal Science and Biotechnology*, 3(1), 26. <http://doi.org/10.1186/2049-1891-3-26>

Capítulo 3

Manuscrito 4 - ANÁLISE DO TRANSCRIPTOMA PARA IDENTIFICAÇÃO DE GENES DE PROTEÍNAS COM FUNÇÃO RELACIONADA A CRESCIMENTO E IMUNIDADE E DETECÇÃO DE POLIMORFISMOS DE BASE ÚNICA (SNPs) NO CAMARÃO *LITOPENAEUS VANNAMEI*.

Camilla A. Santos¹, Flávio Farias², Ana Karina Teixeira², Ana Carolina Guerrelhas³, Sônia C.S. Andrade⁴ e Patricia D. Freitas¹

¹Departamento de Genética e Evolução, Universidade Federal de São Carlos, São Carlos, SP, Brazil.

²Genearch Aquacultura, Pititinga, Rio do Fogo, Rio Grande do Norte, RN, Brazil.

³Aquatec Larvicultura de Camarão Marinho, Barra do Cunhaú, Canguaretama, Rio Grande do Norte, RN, Brazil.

⁴Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, SP, Brazil.

Correspondence to: Camilla Alves Santos. Via Washington Luiz km 235, Phone+55-16-3351-8309, Fax:+55-16-3351-8377. Caixa Postal 676, CEP13565-905, São Carlos-SP, Brazil.

E-mail: camilla.alves@yahoo.com.br

Resumo

Litopenaeus vannamei é uma das espécies de camarão mais importantes para a aquicultura mundial. Apesar disso, pouca informação é conhecida sobre os genes e proteínas desse peneídeo e outros crustáceos taxonomicamente próximos. Nesse trabalho, foi utilizada a tecnologia de sequenciamento de RNA (RNA-seq) do hepatopâncreas e músculo de indivíduos de *L. vannamei* separados em dois experimentos, (i) de rápido crescimento; (ii) exposição ao patógeno da *White Spot Syndrome Virus* (WSSV). Foram gerados 63.105 transcritos, com tamanho médio de 2.511 pb e N50 de 3.464 pb. Desse total, 22.989 (36%) apresentam *hits* contra a base de dados SwissProt.

Cerca de 15.500 *Single Nucleotide Polymorphisms* (SNPs) (frequência $\geq 50\%$) foram detectados em 3.572 unigenes, com o número de SNPs/unigenes variando de 1 a 70. Foram identificados 5.029 SNPs exclusivos dos animais avaliados para rápido crescimento e 6.529 nos camarões expostos ao WSSV. Três mil e noventa e sete SNPs comuns a ambas as condições também foram identificados. Dentre os locos com SNPs, 2.575 e 1.011 retornaram *hits* no Gene Ontology e KEGG, respectivamente, com 897 KEGG Orthology (KO) únicos, distribuídos em 333 vias mapeadas do KEGG. Os achados para os transcriptomas descritos aqui contribuem significativamente para os estudos de genômica e transcriptômica no grupo dos crustáceos, com ênfase especial para abordagens futuras visando o desenvolvimento de programas de seleção na aquicultura.

Palavras-chave: *L. vannamei*, Polimorfismos, Crescimento, WSSV, RNA-seq.

1. Introdução

Entre os camarões peneídeos estão incluídas espécies de enorme valor comercial, sendo estas responsáveis por um terço de toda pesca mundial anual de crustáceos (Wang *et al.*, 2013). O camarão branco do Pacífico *Litopenaeus vannamei* representa a principal espécie de camarão comercializada (Ma *et al.*, 2009). Embora a produção desse crustáceo se apresente como uma atividade de sucesso, os lucros esperados através de seu cultivo, no entanto ainda são freados por fatores como crescimento lento e baixa sobrevivência (Jung *et al.*, 2013; Lv *et al.*, 2014), além do contágio por patógenos como fungos, bactérias e vírus (Wang *et al.*, 2013; Liu *et al.*, 2014; Rao *et al.*, 2015). O vírus causador da Síndrome da Mancha Branca (WSSV; *White Spot Syndrome Virus*), por exemplo, além de retardar o crescimento corpóreo dos animais, pode ocasionar até 100% de mortalidade nos tanques de cultivo (FAO, 2016). Tais fatores ocasionam anualmente grandes perdas econômicas na carcinicultura mundial (Chen *et al.*, 2015).

Apesar da enorme importância comercial dos camarões peneídeos, os mecanismos moleculares envolvidos no crescimento e imunidade ainda não são conhecidos em detalhes. Poucos genes determinantes para o crescimento (Jung *et al.*, 2013; Jindra *et al.*, 2013; Lv *et al.*, 2014) e

resposta imune contra micro-organismos oportunistas (Rattanaporn e Utarabhand, 2011; Gajula *et al.*, 2013; Fan *et al.*, 2016) já foram descritos para crustáceos. Esta realidade torna evidente a necessidade de ampliar a quantidade de informações sobre genes e transcritos envolvidos nas respostas celulares e fisiologia dos crustáceos. Além disso, a falta de um genoma de referência para camarões ressalta a importância dos estudos abordando o sequenciamento de próxima geração (NGS) como fonte relevante de informações (Santos *et al.*, 2014; Powell *et al.*, 2015).

A abordagem de RNA-seq tem tornado possível o sequenciamento de genes com alta cobertura, possibilitando também a identificação em larga escala de polimorfismos relacionados a características de desempenho em espécies não modelo, como os crustáceos (Cui *et al.*, 2013; Yu *et al.*, 2014). Marcadores codominantes, como os Polimorfismos de Base única (*Single Nucleotide Polymorphisms*; SNPs) são altamente informativos e abundantes nos genomas (Brookes, 1999; Vignal *et al.*, 2002). Esses polimorfismos tem se mostrado eficientes em estudos de mapas de ligação e associação entre dados moleculares e fenotípicos em diversas espécies, como caranguejos, (Cui *et al.*, 2013) e camarões de água doce (Jin *et al.*, 2013). No entanto, apenas um estudo de SNPs identificados em *L. vannamei* por NGS foi reportado (Yu *et al.*, 2014) até o momento.

Nesse estudo, foi utilizada a abordagem de RNA-seq para a construção de um transcriptoma de hepatopâncreas e músculo de indivíduos de *L. vannamei*. Foram utilizadas amostras de dois grupos experimentais (i) músculo de animais avaliados para rápido crescimento e alta sobrevivência e (ii) hepatopâncreas e músculo de animais expostos ao vírus da Mancha Branca. Grande parte das respostas imunes nos crustáceos acontece no hepatopâncreas, devido a muitas proteínas envolvidas em reconhecimento e eliminação de patógenos estarem armazenadas nesse órgão (Guo *et al.*, 2013; Chen *et al.*, 2015). Os tratamentos e tecidos foram unidos no esforço de maximizar a representatividade dos transcritos e a cobertura do transcriptoma. Esse estudo tem como objetivo

principal identificar polimorfismos (SNPs) em genes candidatos relacionados principalmente a crescimento, ganho de peso e respostas imunes.

2. Material e Métodos

2.1. Obtenção das amostras

Os exemplares de *Litopenaeus vannamei* avaliados para crescimento utilizados nesse experimento foram originados da Empresa de Melhoramento Genético Genearch, localizada na cidade de Rio do Fogo, no estado do Rio Grande do Norte (RN), Brasil. A escolha das amostras foi realizada entre as 20 famílias SPF (*Specific Pathogen Free*) constituintes do mesmo *batch*, submetidas ao programa de melhoramento genético para altas taxas de sobrevivência e rápido crescimento. Nessa etapa do processo de seleção, as famílias encontravam-se em triplicatas e com cerca de dois meses de cultivo (peso em torno de 1.5 g). Com base nos dados de desempenho para crescimento pertencentes à empresa, seis camarões (dois de cada um dos tanques em triplicata) de cada uma das quatro famílias com maior média de peso, em gramas, foram coletados, totalizando 24 animais nessa condição. O mesmo procedimento foi realizado com os 24 animais das quatro famílias de menor média de peso do mesmo *batch* e amostras de músculo (pleópodos) foram coletadas.

Pós-larvas (PLs) BH12, híbridas de animais SPF, foram obtidas na Empresa de Larvicultura Aquatec, localizada na cidade de Canguaretama, no estado do Rio Grande do Norte, Brasil, e colocadas em contato com o vírus da Mancha Branca (WSSV) em uma fazenda com alta incidência da doença. As PLs utilizadas nesse trabalho pertencem a uma linhagem SPF comercial de alta performance de crescimento, sendo mantidas em ambiente sanitário monitorado para a presença de diversas doenças, incluindo a WSSV. Foram selecionados oito animais, já pesando cerca de 6 g, quatro deles com sintomas e quatro sem sintomas da doença. Amostras de pleópodos e

hepatopâncreas dos oito animais foram coletadas, totalizando 16 amostras. Amostras de hemolinfa foram submetidas a análises prévias de PCR Quantitativo em Tempo Real (qPCR) para a confirmação da presença/ausência do vírus, seguindo procedimento proposto por Silva *et al.* (2011). A eficiência de amplificação RT-PCR foi determinada pelo método descrito por Pfaffl *et al.* (2002) e a especificidade da amplificação foi confirmada pela análise da curva de dissociação (Ririe *et al.* 1997). Todos os tecidos coletados das 64 amostras utilizadas foram acondicionadas em *RNA later* (Thermo Fisher Scientific, Waltham, MA, USA) e mantidas em biofreezer (-80°C) para extração do RNA total.

2.2. Extração de RNA, controle de qualidade e sequenciamento

Foram selecionados oito animais avaliados para rápido crescimento e quatro expostos ao WSSV, totalizando 12 camarões a serem utilizados na montagem *de novo* e construção da referência. Esse número de amostras foi escolhido por conter um animal de cada uma das oito famílias avaliadas para rápido crescimento e amostras dos tecidos hepatopâncreas e músculo de indivíduos doentes e saudáveis (expostos ao WSSV). Dessa forma, a representatividade dos transcritos foi maximizada permitindo também o aumento da cobertura do transcriptoma. A extração de RNA total das 64 amostras de músculo e hepatopâncreas foi realizada utilizando uma versão modificada do protocolo de Trizol®/clorofórmio (Chomczynski *et al.*, 1995) seguida da purificação com LiCl 7.5 M para eliminação de possíveis traços de fenol. A quantidade foi verificada em fluorômetro Qubit (Thermo Fisher Scientific) e em espectrofotômetro NanoDrop (Thermo Fisher Scientific) e as amostras com razões entre 1.8 – 2.2 foram consideradas suficientemente puras. A integridade das amostras foi confirmada no equipamento BioAnalyser (Agilent Technologies Inc., Santa Clara, CA, USA), aquelas com valores de *RNA Integrity Number* (RIN) ≥ 6 foram armazenadas a -80° C. Cada uma das amostras deu origem a uma biblioteca construída

através do kit TruSeq RNA Library Preparation V2 (Illumina Inc). A quantificação das mesmas, antes de serem carregadas no sequenciador, foi feita por qPCR com o Library Quantification Kit Universal (KAPA Biosystems, Wilmington, MA, USA). As bibliotecas foram agrupadas e corridas em plataforma Illumina HiSeq 2500 com 2 x 100 pb *paired-end*, utilizando-se o kit TruSeq SBS V3 (Illumina Inc.), no Laboratório de Biotecnologia Animal, Escola Superior de Agricultura "Luiz de Queiroz"-Universidade de São Paulo (ESALQ-USP), Piracicaba-SP, Brasil.

2.3. Limpeza das sequências e montagem *de novo*

A qualidade dos dados brutos gerados após sequenciamento foi visualizada no software FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) (versão 0.10.1). Todas as *reads* foram filtradas por valor de qualidade 23 e 30 do Phred (QS) e comprimento mínimo de 65 pb utilizando o programa SeqyClean (v.1.9.9) (<https://github.com/ibest/seqyclean>). Esse programa também remove sequências de contaminantes (*primers* e vetores) utilizando como base de dados o Univec (<https://www.ncbi.nlm.nih.gov/tools/vecscreen/univec/>).

Após a filtragem das *reads*, foi realizada a montagem da referência no software Trinity (Grabherr *et al.* 2011). Para tal, 12 bibliotecas foram selecionadas para serem utilizadas na montagem. As sequências foram normalizadas utilizando-se o script *insilico_read_normalization.pl* do Trinity, com 50X adotado como valor mínimo de cobertura. Para a montagem foi definido 300 pb como tamanho mínimo dos *contigs* formados. Os pacotes CD-Hit (Li e Godzik 2006) e DustMasker (Morgulis *et al.*, 2006) foram utilizados para remover *contigs* redundantes com mais de 95% de similaridade e sequências de baixa complexidade, respectivamente. Após a filtragem, os *contigs* foram anotados contra a base de dados do SwissProt com valor de *e-value* máximo de $1e10^{-3}$. O conjunto completo de *reads* das amostras utilizadas ficará disponível no NCBI Sequence Read Archive (SRA) em breve.

O pacote TransDecoder (<http://transdecoder.sourceforge.net/>) foi utilizado na identificação das regiões codificantes candidatas dos *contigs*. O *pipeline* do Trinotate (<https://trinotate.github.io/>) foi utilizado para anotação das sequências através dos seguintes bancos de dados: Uniprot (uniref90 + SwissProt) com valor de corte de $1e10^{-5}$, Gene Ontology (GO) (Ashburner *et al.*, 2000) com os termos Processo Biológico, Função Molecular e Componente Celular do GO e KEGG (*Kyoto Encyclopedia of Genes and Genomes*) (Kanehisa *et al.*, 2012), com a análise das principais vias metabólicas encontradas.

2.4. Detecção dos Polimorfismos de Base Única (SNPs)

As 48 amostras de animais avaliados para rápido crescimento e as 16 de camarões expostos ao WSSV foram mapeadas contra a referência construída pelo pacote Trinity. No mapeamento, foi utilizado o software Bowtie2 versão 2.2.6 (Langmead e Salzberg, 2012). Os arquivos de alinhamento foram utilizados para a identificação dos SNPs. O pacote Samtools (versão 1.3) (Li *et al.*, 2009) foi utilizado para detectar os SNPs através do comando `mpileup` usando os seguintes parâmetros: `-g -u` (calcular as probabilidades dos genótipos e gerar um arquivo BCF não comprimido), `-q20` (valor mínimo de qualidade de mapeamento para um alinhamento), `-Q20` (valor mínimo de qualidade de uma base), `-C50` (valor de qualidade de mapeamento mínimo usada para desconsiderar *reads* com muitos erros de alinhamento ou *mismatches*), `-A` (não ignorar pares de *reads* com problemas) e `-B` (não habilitar um realinhamento probabilístico, evitando a detecção de falsos SNPs por erros no alinhamento). Os SNPs foram chamados através do Bcftools (versão 1.3) (Li *et al.*, 2009). SNPs com qualidade da base variante < 30 e soma da cobertura dos *reads* com alelos alternativos nas fitas *forward* e *reverse* ($DP4$) < 10 foram excluídos das análises para evitar artefatos. Foram consideradas nesse trabalho apenas as variantes com um mínimo de 50% de frequência no total de 64 amostras. O comando `--max-missing 0.5` foi utilizado no vcftools (versão 0.1.14) (Danecek *et al.*, 2011) para

filtrar os polimorfismos com frequência $\geq 50\%$. A taxa de transições e transversões foi calculada no bcftools através do comando `--stats`. A Frequência do Menor Alelo (MAF) foi obtida através do comando `--maf 0.05` no vcftools.

3. Resultados e Discussão

3.1. Sequenciamento do transcriptoma e montagem *de novo*

Cerca de 30 milhões de *reads* pareadas (29.376.967) foram trimadas para contaminantes e sequências de rRNA ou de baixa qualidade, restando 28.654.475 (97.5%) fragmentos que participaram da construção da referência. Foram obtidos 63.105 transcritos com tamanho médio de 2.511 pb e N50 = 3.464 pb. Desse total, 22.989 (36%) apresentaram *blastX hits* no SwissProt (Tabela 1). Foram considerados como unigenes os *contigs* filtrados para $< 95\%$ de similaridade e isoformas após a construção da referência. A porcentagem de unigenes com *hits* no SwissProt encontrado nesse trabalho é superior ao observado para os transcriptomas de outras espécies de camarões, como *Macrobrachium rosenbergui* (26.47%) (Rao *et al.*, 2016) e *Penaeus monodon* (11%) (Nguyen *et al.*, 2016). Os parâmetros de tamanho médio e N50 dos *contigs* refletem a qualidade da montagem aqui realizada, a qual apresenta valores superiores aos já reportados em outros transcriptomas de *L. vannamei*, como por exemplo, em Guo *et al.* (2013) com valores de tamanho médio de 561 pb (N50 = 736 pb). Após a eliminação dos *contigs* redundantes, sequências de baixa complexidade e isoformas foram identificadas as regiões codificantes candidatas restando 14.124 unigenes. Esses foram anotados contra o banco nr-NCBI (não redundante), no qual 11.256 (18%) retornaram *hits*, evidenciando um valor superior ao encontrado para *P. monodon* (13%) em Nguyen *et al.* (2016). Foi realizada a anotação funcional contra a base do GO e identificação das vias do KEGG no pacote Trinotate, cujos resultados estão na Tabela 1.

Tabela 1: Estatísticas das 12 amostras utilizadas para a construção da referência após a normalização e filtragem no SeqyClean. São mostradas também a avaliação do transcriptoma de referência e a visão geral da anotação funcional dos transcritos contendo SNPs.

Transcritos das 12 amostras utilizadas na construção da referência após filtragem no SeqyClean						
Tecido	Crescimento		Saudáveis/Doentes (WSSV)			
	Músculo (8 bibliotecas)	Músculo (2 bibliotecas)	Hepatopâncreas (2 bibliotecas)			
Tratamento	Maior crescimento	Menor crescimento	Saudável	Doente	Saudável	Doente
Fragmentos <i>paired-end</i>	142.094.916	150.593.354	44.029.454	44.972.016	59.850.408	47.141.162
Total	292.688.270		89.001.470		106.991.570	
488.681.310						
Transcritos das 12 amostras após normalização no Trinity						
Número de <i>reads</i> após a normalização					29.376.967	
Número de <i>reads</i> remanescentes após filtragem de <i>reads</i> de baixa qualidade e com adaptadores (<i>Reads</i> participantes da montagem da referência)					28.654.475 (97.5%)	
Análise do Transcriptoma de referência						
Total de transcritos gerados					63.105	
Unigenes com <i>blastX hits</i> no SwissProt					22.989 (36%)	
Tamanho médio					2.511 pb	
N50					3.464 pb	
Transcrito mais longo (pb)					20.601	
Número de transcritos > 1000 pb					27.887	
Número de unigenes não redundantes (trimagem de > 95% similaridade e isoformas). Participaram da detecção dos SNPs					20.943	
Número unigenes não redundantes com ORFs					14.124	
Anotação funcional das proteínas no Trinotate						
Unigenes com <i>blastX hits</i> (nr)					11.256 (18%)	
SNPs (frequência >= 50%)						
Unigenes com SNPs					3.572	
Número de SNPs					15.465	
Unigenes com anotação no GO e SNPs					2.575 (72%)	
Unigenes com KEGG <i>hits</i> e SNPs					1.011 (30%)	
No. KOs					897	
Número de vias metabólicas do KEGG participantes					333	

3.2. Detecção dos SNPs

Um total de 15.465 potenciais SNPs (frequência >= 50%) foi identificado em 3.572 unigenes, sendo 14.929 bi-alélicos e 536 multi-alélicos. O número médio de SNPs/unigene obtido foi de 4.3.

Outros transcriptomas Illumina de crustáceos reportaram valores médios de 3.7 e 1.1 SNPs/unigene para o caranguejo *Eriocheir sinensis* (Cui *et al.*, 2013) e o camarão *Macrobrachium nipponense* (Jin *et al.*, 2013), respectivamente. Foram encontrados 15.009 (97%) SNPs com a Frequência do Menor Alelo (MAF) \geq 5%. Cerca de 28% dos SNPs apresenta o segundo alelo mais comum com frequência de 0.45 a 0.5 nas amostras. A frequência encontrada indica que o conjunto amostral utilizado apresenta um número de locos heterozigotos com o segundo alelo mais frequente distribuído nas amostras avaliadas. Esses dados são corroborados ao se analisar a heterozigosidade dos locos, dos quais 3.356 (21%) são heterozigotos.

O número de SNPs por unigene variou de 1 a 70, com 874 (24%) unigenes contendo um mínimo de 5 SNPs/loco (Tabela S1 - Material Suplementar). Em relação à substituição de bases, foram encontradas 10.387 transições (67%) e 5.078 transversões (33%). Entre todas as substituições observadas, as transições C/T foram as mais frequentes. As transversões por sua vez, apresentaram as taxas de G/C, G/T e A/C próximas, enquanto de A/T foi ligeiramente mais elevada (Tabela 2). Uma possível explicação para a maior quantidade de transições observada é um possível viés na alta taxa de desaminação do grupo metil da citosina, transformando-a em uma timina (C/T), que é visto como G/A na fita reversa (Vignal *et al.*, 2002). As transversões aparecem em menor número provavelmente devido às diferenças na estrutura molecular de purinas e pirimidinas e número de pontes de hidrogênio ligantes (Jin *et al.*, 2013; Lv *et al.*, 2014). A taxa de transições por transversões (ts/tv) foi de 2.02:1. É comum o encontro de uma maior taxa de transições em relação as transversões em transcriptomas de organismos aquáticos, como os camarões *M. nipponense* (1.02:1) (Jin *et al.*, 2013) e *L. vannamei* (2:1) (Yu *et al.*, 2014) e o peixe *Takifugu rubripes* (1:0.59) (Cui *et al.*, 2014). No entanto, os dados gerados aqui contrastam com os do trabalho de Cui *et al.* (2013), no qual o número de transversões superou o de transições no caranguejo *E. sinensis*, com taxa tv/ts de 1.84:1. As estatísticas de cobertura de *reads* para cada SNPs foi calculada (Figura S1 -

Material Suplementar), com média de 383 X, valores superiores aos encontrados em outros estudos utilizando dados de RNA-seq de Illumina com *L. vannamei* presentes na literatura, como em Yu *et al.*, 2014, com média de cobertura de 44 X. Mais que 50% do total de SNPs identificado apresentou cobertura maior que 500 X.

Tabela 2: Taxas de transições e transversões observadas nos SNPs identificados no transcriptoma de *L. vannamei*.

Tipo	Transições			Transversões		
Substituição	CT	GA	AC	AT	GC	GT
Total	5.353 (46.3%)	5.206 (45.0%)	1.317 (11.3%)	1.844 (15.9%)	1.027 (8.8%)	1.266 (10.9%)

Foram identificados 5.029 polimorfismos exclusivos para as amostras dos animais avaliados para rápido crescimento, 6.529 para os expostos ao WSSV e 3.907 comuns a ambas as condições (Tabela S1 - Material Suplementar).

3.3. BlastX hits em espécies de artrópodes encontrados nas bases do SwissProt

Após o *blast* das sequências contra o banco de dados do SwissProt, foi dada maior ênfase aos genes que apresentaram *hits* para espécies de artrópodes e com funções relacionadas a desempenho, como crescimento e resistência a doenças. Dentre os principais marcadores encontrados exclusivamente nas amostras dos animais avaliados para rápido crescimento estão os contidos nos genes: receptor de rianodina (RY44), miosina (MYSA), hormônio juvenil esteroide (JHE), quitina, integrantes de cutícula, pró-resilina (RESIL) e paramiosina (MYSP1). O número de SNPs por loco variou de 2 a 27 (Tabela 3). A proteína rianodina atua principalmente nos músculos e

acredita-se que a sua função dentro do grupo dos invertebrados seja conservada, sendo que no nematelminto *Caenorhabditis elegans* mutações nesse gene resultaram em diminuição do movimento muscular e crescimento lento (Maryon *et al.*, 1998). Em *Drosophila melanogaster*, estudos anteriores demonstraram que quando os receptores da rianodina são bloqueados, ocorre interrupção da contração muscular (Littleton e Ganetzky, 2000; Sullivan *et al.*, 2000) e consequente comprometimento da função do músculo.

Na miosina, principal proteína constituinte dos músculos (Harrington e Rogers, 1984), diversas variantes alélicas foram encontradas nos genes da cadeia pesada. Este alto polimorfismo observado pode estar associado às distintas capacidades de contração das inúmeras variantes das proteínas que compõem a miofibrila, incluindo principalmente a miosina (Jung *et al.*, 2013). A miosina, juntamente com a paramiosina, atua na formação das miofibrilas musculares, sendo juntas responsáveis pela estabilização do sarcômero (Liu *et al.*, 2005). Existem evidências de que o aumento na produção de miosina e reorganização das miofibrilas esteja associado ao crescimento durante e após a muda em camarões (Cesar e Yang, 2007; Li *et al.*, 2014).

O gene JHE codifica o hormônio juvenil esteroide que atua no desenvolvimento sexual dos artrópodes e o seu precursor participa do controle de processos como muda, metamorfose e reprodução (Jindra *et al.*, 2013). É possível que alguns dos polimorfismos encontrados aqui confirmem vantagens adaptativas, por exemplo, ao impedir a maturação sexual precoce e favorecer a muda, resultando em animais que apresentam maiores taxas de crescimento (Shechter *et al.*, 2008). As proteínas constituintes de quitina estão entre as responsáveis pela digestão do exoesqueleto de quitina dos artrópodes, possibilitando a ecdise e o crescimento do animal (Li *et al.*, 2015). Por sua vez, as proteínas integrantes da cutícula, como as observadas aqui, atuam na produção e enrijecimento de um novo exoesqueleto de quitina após a muda (Zhang *et al.*, 2014). O enrijecimento precoce da cutícula ou o desenvolvimento de um exoesqueleto de tamanho menor

que o esperado prejudicam o crescimento do animal (Li *et al.*, 2015). Sendo assim, a existência de variantes alélicas nesses genes pode influenciar no ganho de peso nos animais e consequentemente no crescimento corpóreo. Em relação à resilina, sabe-se que esta proteína é constituinte da cutícula dos artrópodes e que quando o gene que a codifica é silenciado há um atraso significativo na muda, prejudicando o crescimento, como observado no lagostim *Cherax quadricarinatus* (Tynyakov *et al.*, 2015).

Os SNPs exclusivos das amostras de camarões expostos ao WSSV, se encontram principalmente nos genes da cadeia pesada de dineína (DYHC), hemocianina B (HCYB) e C (HCYC), estresse térmico 22 (HSP22) e 67B2 (HS6B), fator anti-lipopolissacarídeo (ALFs), lectinas tipo C, Sparzle 3 (Spz3), crustina, crustacianina 1 (CRA1) e 2 (CRA2) e resposta viral. O número de polimorfismos variou de 1 a 33 por loco (Tabela 3). A dineína é uma das proteínas motoras do citoesqueleto, sendo o movimento produzido pela cadeia pesada da proteína (Asai e Koonce, 2001). É possível que alguns dos polimorfismos encontrados nesse trabalho estejam relacionados à fagocitose no citoesqueleto, com algumas variantes podendo acarretar em alguma vantagem imunológica. Em artrópodes, uma das formas mais comuns de defesa contra micro organismos oportunistas é a endocitose com a consequente encapsulação e destruição do patógeno (Wongpanya *et al.*, 2007).

As proteínas hemocianina, crustina e crustacianinas possuem ampla ação contra vírus, especialmente o WSSV, atuando em respostas ao estresse e sobrevivência do animal (Fan *et al.*, 2016), na aglutinação do patógeno na hemolinfa e lise celular (Shockey *et al.*, 2009; Zheng *et al.*, 2016). Por sua vez, as lectinas atuam no reconhecimento do patógeno através de estruturas moleculares específicas da molécula (Song *et al.*, 2010). Assim, os polimorfismos identificados nesses genes podem se constituir numa importante fonte de informação para estudos futuros que avaliem a associação dessas variantes com a resposta imune.

Variantes alélicas em genes responsáveis pela resposta imune são esperadas, uma vez que alguns polimorfismos podem favorecer o reconhecimento mais eficiente do patógeno, conferindo vantagens adaptativas. Por não possuírem uma memória imunológica, dependem majoritariamente de suas proteínas atuantes na defesa contra patógenos e dos polimorfismos presentes em seus respectivos genes. Estas variantes podem fornecer resultados mais eficientes na defesa contra patógenos em comparação a outros indivíduos da mesma espécie que não as possuem (Tassanakajon *et al.*, 2014). A estrutura molecular das proteínas dessa natureza evolui de acordo com a evolução dos micro organismos oportunistas em que entram em contato, sendo uma possível justificativa para os polimorfismos encontrados aqui nos genes de proteínas imunes (Antony *et al.*, 2011; Imjongjirak *et al.*, 2011). Um exemplo dessa evolução constante pode ser observado nos fatores anti-lipopolissacarídeos (ALF), nos quais é necessário uma grande especificidade na ligação entre a molécula de ALF e a estrutura molecular do patógeno, resultando em uma alta diversidade de sequências na proteína, com inúmeras isoformas (Tassanakajon *et al.*, 2014) (Tabela 3). Foram identificados variantes também no gene da proteína Spatzle 3, cuja função esta relacionada a ativação da via metabólica *Toll*, importante no recrutamento de peptídeos antimicrobianos, como os ALF e as crustinas, em artrópodes (Shi *et al.*, 2013; Wang *et al.*, 2012). As variantes observadas podem influenciar no reconhecimento de patógenos específicos, como o WSSV, que é alvo da proteína Spz3, conferindo vantagens adaptativas aos animais. Outros genes polimórficos encontrados aqui são os das proteínas de estresse térmico HSP22 e HS6B. Proteínas dessa natureza estão ligadas a situações de estresse severo desencadeado perante perturbações ambientais as quais os animais estão expostos (Feder, 1999).

Também foram encontrados SNPs comuns a ambas as condições de crescimento e exposição ao WSSV, em genes como os das proteínas crustacianina, estresse térmico 60, ALF, coagulante de hemolinfa e peneidina (Tabela 3). A coagulação da hemolinfa é um dos principais

mecanismos de defesa contra micro organismos oportunistas adotado pelos crustáceos, com a consequente interrupção na proliferação do patógeno (Cheng *et al.*, 2008). As peneidinas são exclusivas da família dos peneídeos e fornecem uma proteção sistêmica ao organismo contra bactérias gram-positivas e alguns fungos, além de serem encontradas em diversos tecidos, como hemócitos, músculo e cutícula (Cuthbertson *et al.* 2004; Du e Jin, 2015). Já é conhecida a relação antagônica entre crescimento e imunidade em camarões (Argue *et al.*, 2002; Cock *et al.*, 2009). Dessa forma, é interessante que os locos envolvidos nessas características sejam selecionados juntos a fim de que os animais submetidos a programas de melhoramento apresentem boas taxas de crescimento e em paralelo, quando expostos a patógenos, apresentem respostas imunes eficientes. Os SNPs comuns às condições de crescimento e exposição à Síndrome da Mancha Branca encontrados aqui são potenciais marcadores a serem utilizados na aquicultura.

Tabela 3: Locos com maior número de SNPs exclusivos para as amostras de rápido crescimento, exposição ao WSSV e comuns a ambas as condições com *hits* para espécies de artrópodes. São detalhados o nome da proteína e o número de SNPs/loco.

Proteína	No. SNP/loco
Crescimento	
Receptor de rianodina (RY44) [<i>Drosophila melanogaster</i>] (Q24498)	27
Miosina (MYSA) [<i>Drosophila melanogaster</i>] (P05661)	25
Hormônio esteroide juvenil (JHE) [<i>Pandalopsis japonica</i>]	9
Quitina desacetilase [<i>Tribolium castaneum</i>]	9
Integrante da cutícula (CU03) [<i>Homarus americanus</i>] (P81387)	5
Integrante da cutícula (CU04) [<i>Homarus americanus</i>] (P81388)	2
Pró-resilina (RESIL) [<i>Drosophila melanogaster</i>] (Q9V7U0)	4
Paramiosina (MYSP1) [<i>Drosophila melanogaster</i>] (P35415)	2
Imunidade	
Cadeia pesada de dineína (DYHC) [<i>Drosophila melanogaster</i>] (P37276)	33
Hemocianina cadeia C (HCYC) [<i>Panulirus interruptus</i>] (P80096)	27
Hemocianina cadeia B (HCYB) [<i>Panulirus interruptus</i>] (P10787)	1
Proteína de estresse térmico (HSP22 [<i>Drosophila O melanogaster</i>] (P02515)	21
Proteína de estresse térmico 67B2 (HS6B) [<i>Drosophila melanogaster</i>] (P22978)	9
Fator anti-lipopolissacarídeo (ALPS) [<i>Portunus trituberculatus</i>] (C0KJQ4)	10
Fator anti-lipopolissacarídeo isoforma 5 [<i>Fenneropenaeus chinensis</i>]	2
Lectina tipo C [<i>Portunus trituberculatus</i>]	4
Lectina tipo C [<i>Fenneropenaeus chinensis</i>]	1
Lectina tipo C [<i>Penaeus monodon</i>]	3
Lectina 5 tipo C [<i>Fenneropenaeus merguensis</i>]	3
Lectina 3 tipo C [<i>Fenneropenaeus merguensis</i>]	1
Lectina 1 tipo C [<i>Litopenaeus vannamei</i>]	1
Sparzle 3 (Spz3) [<i>Litopenaeus vannamei</i>]	5
Crustina tipo 1 [<i>Panulirus japonicus</i>]	2
Crustina tipo 2 [<i>Panulirus japonicus</i>]	3
Provável crustina [<i>Litopenaeus vannamei</i>]	2
Crustina-PI 2 [<i>Pacifastacus leniusculus</i>]	1
Crustacianina 1 (CRA1) [<i>Homarus americanus</i>] (P58989)	1
Crustacianina 2 (CRA2) [<i>Homarus americanus</i>] (P80007)	1
Resposta viral [<i>Penaeus monodon</i>]	1
Crescimento e Imunidade	
Miosina (MYSA) [<i>Drosophila melanogaster</i>] (P05661)	92
Crustacianina subunidade A2 [<i>Homarus gammarus</i>] (P80007)	15
Anti-lipopolysaccharide factor [<i>Portunus trituberculatus</i> (C0KJQ4)	12
60 kDa heat shock protein, mitochondrial [<i>Drosophila melanogaster</i>] (O02649)	10
Hemolymph clottable protein [<i>Penaeus monodon</i>] (Q9U572)	10
Penaeidin-3c [<i>Litopenaeus vannamei</i>] (P81060)	7

Os animais utilizados neste trabalho foram todos obtidos de um ambiente cativo e alguns expostos a um patógeno letal, o que certamente desencadeia respostas ao estresse em seus organismos. Características relacionadas a desempenho geralmente apresentam níveis mais elevados de diversidade genética quando comparadas as não relacionadas. Essa hipótese é suportada principalmente pelos traços de *fitness* serem controlados por um número grande de locos e ao se considerar as frequências mutacionais desses inúmeros locos (Houle, 1992). Muitas vezes as mutações acumuladas são um reflexo de um ambiente específico em que os organismos vivem, os quais podem ser adaptações particulares para uma determinada população (Merila e Sheldon, 1999).

3.4. Anotação funcional no GO e análise de vias metabólicas

Cerca de 11.250 unigenes apresentaram blastX *hits* no nr-NCBI dentre os 14.124 alinhados. Após a detecção dos SNPs, foram observados 3.572 unigenes contendo SNPs (frequência $\geq 50\%$), com 2.575 retornando *hits* (72%) contra o banco de dados do GO (Tabela S2 - Material Suplementar). A anotação funcional no GO no nível 2 retornou 3.607 termos únicos da categoria Processos Biológico (PB), sendo os mais frequentes: transcrição (GO:0006351), tradução (GO:0006412), catabolismo de ATP (GO:0006200), apoptose (GO:0006915) e transporte de proteínas (GO:0015031). Para Função Molecular (FM) com 1.433 *hits* únicos, aparecem os papéis envolvidos com a ligação de: íons (GO:0046872); ATP (GO:0005524); DNA (GO:0003677) e RNA (GO:0003723) e constituinte do ribossomo (GO:0003735) e músculo (GO:0008307). Em relação ao Componente Celular (CC) foi possível a identificação de 696 termos únicos constando entre os mais frequentes o citoplasma (GO:0005737), núcleo (GO:0005634), membrana celular (GO:0016021) e mitocôndria (GO:0005739) (Figura 1).

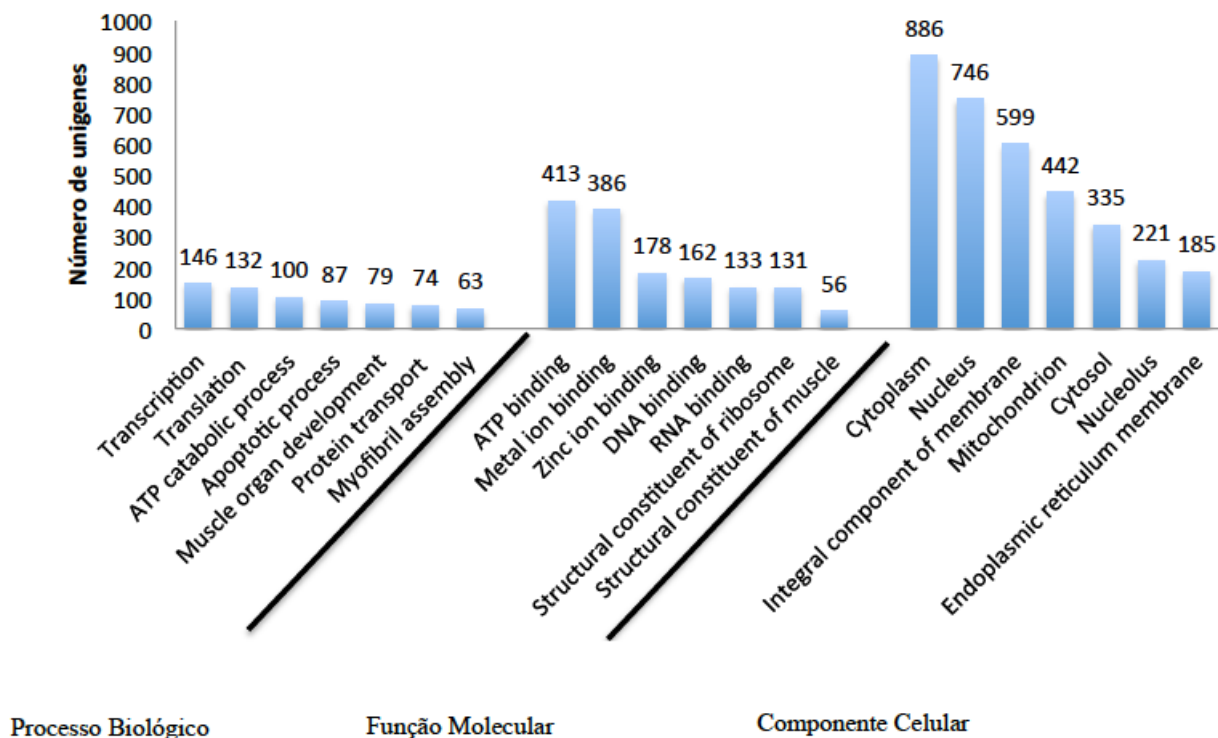


Figura 1: Distribuição dos termos do GO encontrados do nível 2 para as categorias Processo Biológico, Função Molecular e Componente Celular. O eixo y representa o número de locos que compõe cada termo do GO e o eixo x as três categorias desse banco de dados.

Em relação aos genes com um mínimo de 10 SNPs/loco e *hits* no GO, Processos Biológicos como tradução (GO:0006412), contração (GO:0006936) e desenvolvimento (GO:0007517) musculares, montagem da miofibrila (GO:00030239), montagem (GO:0030241) e organização (GO:0031033) do filamento de miosina, locomoção (GO:0040011), diferenciação da célula muscular (GO:0042692) e organização do sarcômero (GO:0045214) aparecem descritos nos genes mais polimórficos. Como função Molecular destaca-se o papel de constituinte do músculo (GO:0008307). Com base no foco dado aos genes com mais variantes alélicas que compõem os termos do GO (Yu *et al.*, 2014), eventos celulares relacionados a funções musculares e tradução aparecem

prioritariamente, sugerindo que estes genes podem estar envolvidos em eventos de desempenho, como o crescimento. Dentre os principais locais de ação dos genes mais polimórficos estão citoplasma (GO:0005737), mitocôndria (GO:0005739), membrana (GO:001602) e sarcômero (GO:0030017) (Tabela S3 - Material Suplementar).

Ao se realizar a procura das sequências de proteínas preditas para *L. vannamei* contra a base de dados do KEGG, foram encontrados *hits* para 1.011 (30%) unigenes com SNPs, consistindo em 897 KEGG Orthology (KO) únicos distribuídos em 333 vias mapeadas na base do KEGG (Tabela S4 - Material Suplementar). A porcentagem de *hits* contra o KEGG encontrada aqui foi condizente com valores reportados em outros trabalhos, como em Cui *et al.* (2014) em que, por exemplo, 31% dos unigenes com SNPs apresentaram KEGG *hits* para o peixe *T. rubripes*, sendo a mesma taxa observada para o camarão *M. nipponense* (Jin *et al.*, 2013). Em relação ao número de vias mapeadas no KEGG, o valor encontrado aqui foi superior a outros transcriptomas *L. vannamei*, como em Yu *et al.*, 2014 em que foram totalizadas 240 vias mapeadas no KEGG.

Merecem destaque as seguintes vias com seus respectivos números de KOs identificados: ko01100 vias de metabolismo (346), ko01110 biossíntese de compostos secundários (98), ko04141 processamento de proteína pelo retículo endoplasmático (54), ko01200 metabolismo de carbono (49), ko03010 ribossomo (44) e ko04144 endocitose (37) (Figura 2). Esses resultados concordam com a anotação do GO já descrita anteriormente aqui, na qual Processos Biológicos como tradução, transporte de proteínas e apoptose, além de Componentes Celulares como citoplasma e mitocôndria estão entre os termos do GO mais comuns encontrados nesse trabalho. O número de KOs nas vias de metabolismo, biossíntese de compostos secundários, processamento de proteínas pelo retículo endoplasmático, ribossomo, e transporte de RNA sugerem um metabolismo elevado de proteínas e carboidratos para a produção de energia. Esse cenário ilustra produtos que podem estar sendo empregados em processos como crescimento e defesa imunológica. Os animais

avaliados aqui foram submetidos à seleção para rápido crescimento, o que caracteriza a intensa demanda por proteínas e energia para o ganho de peso. Além disso, o fato de parte das amostras ter sido exposta ao WSSV favorece a expressão de genes relacionados a respostas imunes. Um exemplo disso é presença de proteínas/enzimas encontradas nesse trabalho pertencentes a vias relacionadas a metabolismo e defesa, como a endocitose. A fagocitose, um dos tipos de endocitose, é um dos mecanismos de defesa contra patógenos usados pelos crustáceos contra vírus e bactérias, no qual estes são englobados pela célula, inativados e destruídos (Cerenius *et al.*, 2008; Liu *et al.*, 2014).

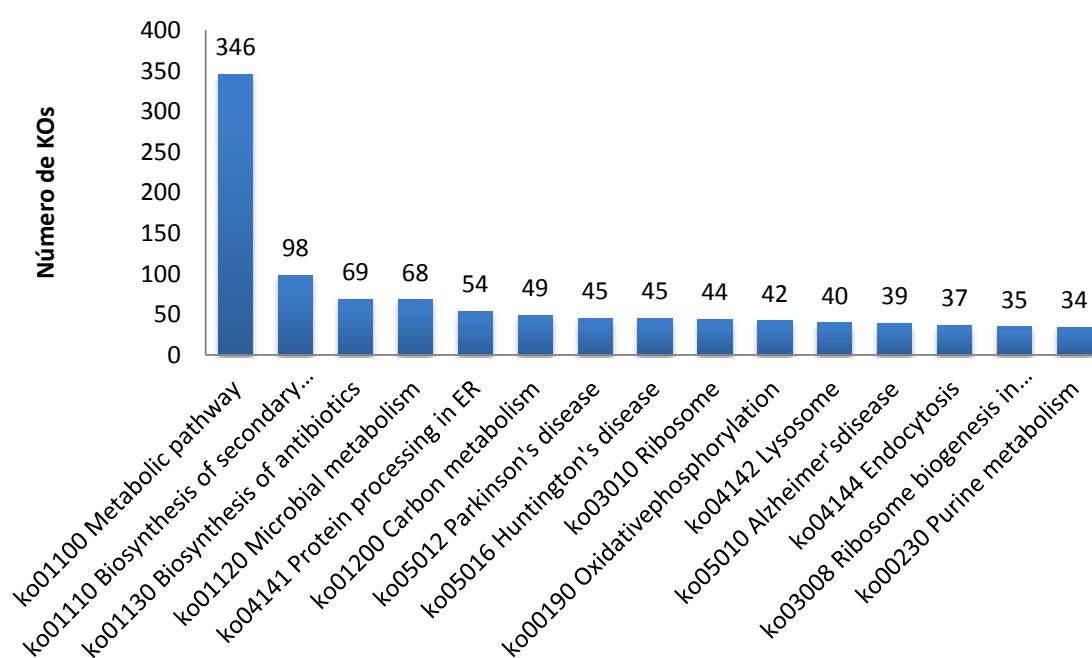


Figura 2: Distribuição do top 15 das vias mapeadas no KEGG. O eixo x representa as vias encontradas e o eixo y mostra o número de KO para cada via correspondente.

No entanto, a distribuição dos KOs ao longo das vias mapeadas variou muito, como pode ser visto pelo número de vias com correspondência encontrada (Tabela S4 - Material Suplementar),

sendo que a maioria apresentou *hits* com processos e/ou doenças não aplicáveis aos invertebrados, não sendo possível uma extrapolação de função para os crustáceos. A ausência de um genoma de referência e o número ainda reduzido de estudos de transcriptomas em crustáceos restringe, portanto, o entendimento mais detalhado do papel de algumas proteínas dentro de vias metabólicas específicas.

Considerações finais

A montagem *de novo* do transcriptoma de tecidos de indivíduos de *L. vannamei* possibilitou a identificação de SNPs em genes importantes relacionados a crescimento e imunidade nos crustáceos. Uma grande variedade de genes com potenciais funções relacionadas ao *fitness* do animal apresentaram diversas variantes alélicas. Além disso, foram identificados SNPs exclusivos para as amostras dos camarões avaliados para rápido crescimento e para a exposição ao WSSV, além de polimorfismos comuns a ambas as condições. Dessa forma, os dados gerados aqui acrescentam informações relevantes ao transcriptomas de peneídeos, uma vez que estudos reportando SNPs detectados através de NGS em *L. vannamei* ainda são raros na literatura. Os achados do presente trabalho podem ser aplicados na aquicultura em estudos de associação genótipo X fenótipo que visem a Seleção Assistida por Marcadores (MAS).

Agradecimentos

Os autores agradecem o apoio financeiro das agências de fomento: Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) (Processo número 2012/17322-8), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), além do Centro de Genômica Funcional do Laboratório de

Biociotecnologia Animal e ao Prof. Dr. Luiz Lehmann Coutinho (ESALQ-USP), Piracicaba-SP, Brasil, pelo serviço de sequenciamento e uso do servidor para processamento dos dados.

Material Suplementar

Figura S1: Gráfico ilustrando os intervalos de cobertura das *reads* utilizadas na detecção dos SNPs (**Pág. 149**).

Tabela S1: SNPs exclusivos das amostras dos animais avaliados para rápido crescimento, expostos ao WSSV e comuns a ambas as condições. Também é mostrado o número de locos com SNPs para cada condição e o número de SNPs/loco. **(Disponível no arquivo anexado)**

Tabela S2: Anotação dos unigenes contendo SNPs no Gene Ontology. **(Disponível no arquivo anexado)**

Tabela S3: Genes com mais de 10 SNPs/loco que compõem os termos do GO Processo Biológico, Função Molecular e Componente Celular. **(Disponível no arquivo anexado)**

Tabela S4: Anotação dos unigenes contendo SNPs no KEGG. **(Disponível no arquivo anexado)**

Referências

- Antony, S. P., Singh, I. S. B., Sudheer, N. S., Vrinda, S., Priyaja, P., & Philip, R. (2011). Molecular characterization of a crustin-like antimicrobial peptide in the giant tiger shrimp, *Penaeus monodon*, and its expression profile in response to various immunostimulants and challenge with WSSV. *Immunobiology*, 216(1–2), 184–194. <http://doi.org/10.1016/j.imbio.2010.05.030>
- Argue, B. J., Arce, S. M., Lotz, J. M., & Moss, S. M. (2002). Selective breeding of Pacific white shrimp *Litopenaeus vannamei* for growth and resistance to Taura Syndrome Virus.
- Asai, D. J., & Koonce, M. P. (2001). The dynein heavy chain: structure, mechanics and evolution. *Trends in cell biology*, 11(5), 196-202.
- Ashburner M., Ball C. A., Blake J.A., Botstein D., Butler H., Cherry J.M., Davis A.P., Dolinski K., Dwight S.S., Eppig J.T., Harris M.A, Hill D.P., Issel-Tarver L., Kasarskis A., Lewis S., M. J. C. (2000). Gene ontology: Tool for the identification of biology. *Natural Genetics*, 25(may), 25–29.
- Brookes, A. J. (1999). The essence of SNPs. *Gene*, 234(2), 177–86. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10395891>
- Cerenius, L., Lee, B. L., & Söderhäll, K. (2008). The proPO-system: pros and cons for its role in invertebrate immunity. *Trends in Immunology*, 29(6), 263–271. <http://doi.org/10.1016/j.it.2008.02.009>
- Cesar, J. R. O., & Yang, J. (2007). Expression Patterns of Ubiquitin , Heat Shock Protein 70 , α -Actin and β -Actin Over the Molt Cycle in the Abdominal Muscle of Marine Shrimp *Litopenaeus Vannamei*. *Comparative Biochemistry & Physiology*, 559(October 2006), 554–559. <http://doi.org/10.1002/mrd>
- Chen, K., Li, E., Li, T., Xu, C., Wang, X., Lin, H., Chen, L. (2015). Transcriptome and molecular pathway analysis of the hepatopancreas in the Pacific white shrimp *Litopenaeus vannamei* under chronic low-salinity stress. *PLoS ONE*,

10(7), 1–22. <http://doi.org/10.1371/journal.pone.0131503>

- Cheng, W., Tsai, I. H., Huang, C. J., Chiang, P. C., Cheng, C. H., & Yeh, M. S. (2008). Cloning and characterization of hemolymph clottable proteins of kuruma prawn (*Marsupenaeus japonicus*) and white shrimp (*Litopenaeus vannamei*). *Developmental and Comparative Immunology*, 32(3), 265–274. <http://doi.org/10.1016/j.dci.2007.05.009>
- Chomczynski P, Mackey K (1995). Short technical reports. Modification of the TRI reagent procedure for isolation of RNA from polysaccharide- and proteoglycan-rich sources. *BioTechniques* 19: 942–945.
- Cock, J., Gitterle, T., Salazar, M., & Rye, M. (2009). Breeding for disease resistance of Penaeid shrimps. *Aquaculture*, 286(1–2), 1–11. <http://doi.org/10.1016/j.aquaculture.2008.09.011>
- Cuthbertson, B. J., Büllsbach, E. E., Fievet, J., Bachère, E., & Gross, P. S. (2004). A new class (penaeidin class 4) of antimicrobial peptides from the Atlantic white shrimp (*Litopenaeus setiferus*) exhibits target specificity and an independent proline-rich-domain function. *The Biochemical Journal*, 381, 79–86. <http://doi.org/10.1042/BJ20040330>
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... & McVean, G. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156–2158.
- Du, Z.-Q., & Jin, Y.-H. (2015). Molecular Characterization and Antibacterial Activity Analysis of Two Novel Penaeidin Isoforms from Pacific White Shrimp, *Litopenaeus vannamei*. *Applied Biochemistry and Biotechnology*, 177(8), 1607–1620. <http://doi.org/10.1007/s12010-015-1840-7>
- Fan, L., Wang, A., Miao, Y., Liao, S., Ye, C., & Lin, Q. (2016). Comparative proteomic identification of the hepatopancreas response to cold stress in white shrimp, *Litopenaeus vannamei*. *Aquaculture*, 454, 27–34. <http://doi.org/10.1016/j.aquaculture.2015.10.016>
- Food and Agriculture Organization of the United Nations (FAO). Threats and risks of introducing alien shrimp species. Disponível em: <http://www.fao.org/docrep/009/a0086e/a0086e10.htm>. Acessado em 10 de setembro de 2016.
- Gajula M. P., Soni G., Babu G., Rai A., B. N. (2013). Journal of Applied Molecular Interaction Studies of Shrimp Antiviral Protein , PmAV with WSSV RING Finger. *Journal of Applied Bioinformatics & Computational Biology*, (2:1), 1–5.
- Grabherr, Manfred G., *et al.* "Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data." *Nature biotechnology* 29.7 (2011): 644.
- Guo, H., Ye, C.-X., Wang, A.-L., Xian, J.-A., Liao, S.-A., Miao, Y.-T., & Zhang, S.-P. (2013). Transcriptome analysis of the Pacific white shrimp *Litopenaeus vannamei* exposed to nitrite by RNA-seq. *Fish & Shellfish Immunology*, 35(6), 2008–16. <http://doi.org/10.1016/j.fsi.2013.09.019>
- Harrington WF, Rogers ME (1984) Myosin. *Annual Review of Biochemistry* 53: 35–73.
- Houle, D. (1992). Comparing evolvability and variability of quantitative traits. *Genetics*, 130(1), 195–204.
- Imjongjirak, C., Amparyup, P., & Tassanakajon, A. (2011). Molecular cloning, genomic organization and antibacterial activity of a second isoform of antilipoplysaccharide factor (ALF) from the mud crab, *Scylla paramamosain*. *Fish and Shellfish Immunology*, 30(1), 58–66. <http://doi.org/10.1016/j.fsi.2010.09.011>
- Jindra, M., Palli, S. R., & Riddiford, L. M. (2013). The Juvenile Hormone Signaling Pathway in Insect Development. *Annual*

Review of Entomology, 58(1), 181–204. <http://doi.org/10.1146/annurev-ento-120811-153700>

- Jung, H., Lyons, R. E., Hurwood, D. A., & Mather, P. B. (2013). Genes and growth performance in crustacean species: A review of relevant genomic studies in crustaceans and other taxa. *Reviews in Aquaculture*, 5(2), 77–110. <http://doi.org/10.1111/raq.12005>
- Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., & Tanabe, M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Research*, 40(Database issue), D109-14. <http://doi.org/10.1093/nar/gkr988>
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357–9. <http://doi.org/10.1038/nmeth.1923>
- Li, E., Wang, S., Li, C., Wang, X., Chen, K., & Chen, L. (2014). Transcriptome sequencing revealed the genes and pathways involved in salinity stress of Chinese mitten crab, *Eriocheir sinensis*. *Physiological Genomics*, 46(5), 177–90. <http://doi.org/10.1152/physiolgenomics.00191.2013>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16), 2078-2079.
- Li, W., & Godzik, A. (2006). Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, 22(13), 1658–1659. <http://doi.org/10.1093/bioinformatics/btl158>
- Li, X., Cui, Z., Liu, Y., Song, C., & Shi, G. (2013). Transcriptome analysis and discovery of genes involved in immune pathways from hepatopancreas of microbial challenged mitten crab *Eriocheir sinensis*. *PloS One*, 8(7), e68233. <http://doi.org/10.1371/journal.pone.0068233>
- Li, X., Xu, Z., Zhou, G., Lin, H., Zhou, J., Zeng, Q., ... Gu, X. (2015). Molecular characterization and expression analysis of five chitinases associated with molting in the Chinese mitten crab, *Eriocheir sinensis*. *Comparative Biochemistry and Physiology Part B Biochemistry and Molecular Biology*, 187, 110–120. <http://doi.org/10.1016/j.cbpb.2015.05.007>
- Littleton, J. T., & Ganetzky, B. (2000). Ion channels and synaptic organization: analysis of the *Drosophila* genome. *Neuron*, 26(1), 35-43.
- Liu, H., Miller, M. S., Swank, D. M., Kronert, W. A., Maughan, D. W., & Bernstein, S. I. (2005). Paramyosin phosphorylation site disruption affects indirect flight muscle stiffness and power generation in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the United States of America*, 102(30), 10522-10527.
- Liu, Y., Hou, F., He, S., Qian, Z., Wang, X., Mao, A., Liu, X. (2014). Identification, characterization and functional analysis of a serine protease inhibitor (Lvserpin) from the Pacific white shrimp, *Litopenaeus vannamei*. *Developmental and Comparative Immunology*, 43(1), 35–46. <http://doi.org/10.1016/j.dci.2013.10.012>
- Lv, J., Liu, P., Gao, B., Wang, Y., Wang, Z., Chen, P., & Li, J. (2014). Transcriptome analysis of the *Portunus trituberculatus*: De novo assembly, growth-related gene identification and marker discovery. *PLoS ONE*, 9(4). <http://doi.org/10.1371/journal.pone.0094055>
- M.E., F. (1999). Heat-shock proteins , molecular chaperones , and the stress response : ... Reproduced with permission of the copyright owner . Further reproduction prohibited without permission .

- Ma, K. Y., Chan, T.-Y., & Chu, K. H. (2009). Phylogeny of penaeoid shrimps (Decapoda: Penaeoidea) inferred from nuclear protein-coding genes. *Molecular Phylogenetics and Evolution*, 53(1), 45–55. <http://doi.org/10.1016/j.ympev.2009.05.019>
- Maryon, E. B., Saari, B., & Anderson, P. (1998). Muscle-specific functions of ryanodine receptor channels in *Caenorhabditis elegans*. *Journal of Cell Science*, 111(19), 2885-2895.
- Morgulis, A., Gertz, E. M., Schäffer, A. A., & Agarwala, R. (2006). A fast and symmetric DUST implementation to mask low-complexity DNA sequences. *Journal of Computational Biology*, 13(5), 1028-1040.
- Nguyen, C., Nguyen, T. G., Nguyen, L. Van, Pham, H. Q., Nguyen, T. H., Pham, H. T., ... Dinh, K. D. (2016). De novo assembly and transcriptome characterization of major growth-related genes in various tissues of *Penaeus monodon*. *Aquaculture*, 464, 545–553. <http://doi.org/10.1016/j.aquaculture.2016.08.003>
- Pfaffl M.W., Horgan G.W. & Dempfle L. (2002). Relative expression softwaretool (REST©) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Res* 30:1–10.
- Powell, D., Knibb, W., Remilton, C., & Elizur, A. (2015). De-novo transcriptome analysis of the banana shrimp (*Fenneropenaeus merguensis*) and identification of genes associated with reproduction and development. *Marine Genomics*, 22, 71–78. <http://doi.org/10.1016/j.margen.2015.04.006>
- Rao, R., Bhasu, S., Bing, R. Z. Y., Alinejad, T., Hassan, S. S., & Wang, J. (2016). A transcriptome study on *Macrobrachium rosenbergii* hepatopancreas experimentally challenged with white spot syndrome virus (WSSV). *Journal of Invertebrate Pathology*, 136, 10–22. <http://doi.org/10.1016/j.jip.2016.01.002>
- Rao, R., Bing Zhu, Y., Alinejad, T., Tiruvayipati, S., Lin Thong, K., Wang, J., & Bhasu, S. (2015). RNA-seq analysis of *Macrobrachium rosenbergii* hepatopancreas in response to *Vibrio parahaemolyticus* infection. *Gut Pathogens*, 7(April 2016), 6. <http://doi.org/10.1186/s13099-015-0052-6>
- Rattanaporn, O., & Utarabhand, P. (2011). Molecular cloning of a C-type lectin with two CRD domains from the banana shrimp *Fenneropenaeus merguensis*: Early gene up-regulation after *Vibrio harveyi* infection. *Journal of Invertebrate Pathology*, 106(2), 196–204. <http://doi.org/10.1016/j.jip.2010.09.002>
- Ririe, K.M., Rasmussen R.P. & Wittwer CT (1997). Product differentiation by analysis of DNA melting curves during the polymerase chain reaction. *Anal Biochem* 245(2):154–160.
- Santos, C. A., Blanck, D. V., & de Freitas, P. D. (2014). RNA-seq as a powerful tool for Penaeid shrimp genetic progress. *Frontiers in Genetics*, 5(AUG), 1–6. <http://doi.org/10.3389/fgene.2014.00298>
- Shechter, A., Berman, A., Singer, A., Freiman, A., Grinstein, M., Erez, J. Sagi, A. (2008). Reciprocal changes in calcification of the gastrolith and cuticle during the molt cycle of the red claw crayfish *Cherax quadricarinatus*. *Biological Bulletin*, 214(2), 122–134. <http://doi.org/10.1007/s11692-008-9012-2> [pii]
- Silva, S.M.B.C., Pinheiro A.C.A.S. & Coimbra M.R.M. (2011). Quantitation of infectious myonecrosis virus indifferent tissues of naturally infected Pacific white shrimp, *Litopenaeus vannamei*, using real-time PCR with SYBR Green chemistry. *J Virol Methods*, 177(2), 197–201.
- Song, K. K., Li, D. F., Zhang, M. C., Yang, H. J., Ruan, L. W., & Xu, X. (2010). Cloning and characterization of three novel

- WSSV recognizing lectins from shrimp *Marsupenaeus japonicus*. *Fish and Shellfish Immunology*, 28(4), 596–603.
<http://doi.org/10.1016/j.fsi.2009.12.015>
- Sullivan, K. M., Scott, K., Zuker, C. S., & Rubin, G. M. (2000). The ryanodine receptor is essential for larval development in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences*, 97(11), 5942-5947.
- Tassanakajon, A., Somboonwiwat, K., & Amparyup, P. (2014). Sequence diversity and evolution of antimicrobial peptides in invertebrates. *Developmental and Comparative Immunology*, 48(2), 324–341.
<http://doi.org/10.1016/j.dci.2014.05.020>
- Tynyakov, J., Bentov, S., Abehsera, S., Khalaila, I., Manor, R., Katzir Abilevich, L., ... Sagi, A. (2015). A Novel Chitin Binding Crayfish Molar Tooth Protein with Elasticity Properties. *Plos One*, 10(5), e0127871.
<http://doi.org/10.1371/journal.pone.0127871>
- Vignal A., Milan D., Sancristobal M., Eggen A. (2002). A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics, Selection, Evolution : GSE*, 34, 275–305. <http://doi.org/10.1051/gse>
- Wang, X. W., & Wang, J. X. (2013). Pattern recognition receptors acting in innate immune system of shrimp against pathogen infections. *Fish and Shellfish Immunology*, 34(4), 981–989. <http://doi.org/10.1016/j.fsi.2012.08.008>
- Wongpanya, R., Aoki, T., Hirono, I., Yasuie, M., & Tassanakajon, A. (2007). Analysis of Gene Expression in Haemocytes of Shrimp *Penaeus monodon* Challenged with White Spot Syndrome Virus by cDNA Microarray. *ScienceAsia*, 33(2), 165. <http://doi.org/10.2306/scienceasia1513-1874.2007.33.165>
- Yu, Y., Wei, J., Zhang, X., Liu, J., Liu, C., Li, F., & Xiang, J. (2014). SNP discovery in the transcriptome of white Pacific shrimp *Litopenaeus vannamei* by next generation sequencing. *PloS One*, 9(1), e87218.
<http://doi.org/10.1371/journal.pone.0087218>
- Zhang, L., Li, L., Zhu, Y., Zhang, G., & Guo, X. (2014). Transcriptome analysis reveals a rich gene set related to innate immunity in the Eastern oyster (*Crassostrea virginica*). *Marine Biotechnology (New York, N.Y.)*, 16(1), 17–33.
<http://doi.org/10.1007/s10126-013-9526-z>
- Zheng, L., Zhao, X., Zhang, P., Chen, C., Liu, S., Huang, R. Zhang, Y. (2016). Hemocyanin from Shrimp *Litopenaeus vannamei* Has Antiproliferative Effect against HeLa Cell In Vitro. *Plos One*, 11, e0151801.
<http://doi.org/10.1371/journal.pone.0151801>

Capítulo 4

Manuscrito 5 - ANÁLISES DE RNA-SEQ REVELAM DIFERENÇAS NO PERFIL DE EXPRESSÃO DE GENES INDUZIDOS EM FAMÍLIAS DE ALTO E BAIXO CRESCIMENTO DO CAMARÃO *LITOPENAEUS VANNAMEI*.

Camilla A. Santos¹, Flávio Farias², Ana Karina Teixeira², Ana Carolina Guerrelhas³, Sonia C.S. Andrade⁴ e Patricia D. Freitas¹

¹Departamento de Genética e Evolução, Universidade Federal de São Carlos, São Carlos, SP, Brazil.

²Genearch Aquacultura, Pititinga, Rio do Fogo, Rio Grande do Norte, RN, Brazil.

³Aquatec Larvicultura de Camarão Marinho, Barra do Cunhaú, Canguaretama, Rio Grande do Norte, RN, Brazil.

⁴Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, SP, Brazil.

Correspondence to: Camilla Alves Santos. Via Washington Luiz km 235, Phone+55-16-3351-8309, Fax:+55-16-3351-8377. Caixa Postal 676, CEP13565-905, São Carlos-SP, Brazil.

E-mail: camilla.alves@yahoo.com.br

Resumo

Um dos principais problemas enfrentados pela carcinicultura é o prejuízo no crescimento dos animais, devido a diversos motivos, como por exemplo, situações de estresse. Na busca de uma solução, torna-se necessário o crescente conhecimento de informação molecular sobre tais propriedades bionômicas, como o processo de ganho de peso. Com esse propósito, este estudo avaliou o transcriptoma obtido via sequenciamento do RNA (RNA-seq) de indivíduos de *Litopenaeus vannamei* submetidos a um programa de seleção para rápido crescimento, visando identificar proteínas envolvidas no processo de crescimento dos crustáceos e verificar possíveis diferenças entre os perfis de expressão dos genes induzidos em animais de alto e baixo desempenho de

crescimento. Para tal, uma análise de expressão diferencial foi realizada para a identificação dos genes diferencialmente expressos (DE) em camarões com maior e menor média de peso corporal. Foram encontrados 207 genes DE, sendo 162 (78%) *up-regulated* para alto e 45 (22%) para baixo crescimento (FDR < 0,05), com 95 genes DE exclusivamente expressos nos animais maiores. Destacaram-se majoritariamente genes relacionados à ecdise, como quitinases e proteínas constituintes da cutícula dos artrópodes. Após o enriquecimento funcional no Gene Ontology (GO), dez termos foram considerados enriquecidos. Os resultados descritos aqui vão contribuir significativamente para o avanço de uma área com pouca informação genômica conhecida e disponível, servindo de fonte para futuras investigações sobre bases moleculares relacionadas ao crescimento em crustáceos.

Palavras-chave: *L. vannamei*, Transcriptoma, Quitinase, Cutícula, Ecdise, Expressão Gênica.

1. Introdução

Um dos principais problemas da carcinicultura mundial é o prejuízo observado no desempenho dos animais devido a diversos fatores externos que influenciam o crescimento, como temperatura, pH e condições gerais estressantes do ambiente (Jung *et al.*, 2013; Chen *et al.*, 2015). No cultivo de camarões, espera-se obter animais com maior tamanho corpóreo e saudáveis, embora já seja conhecido que uma grande parcela da energia que deveria estar sendo utilizada no crescimento e reprodução, por exemplo, acaba sendo desperdiçada devido a condições desfavoráveis nos tanques de cultivo (Rao *et al.*, 2016). Algumas etapas envolvidas no crescimento dos artrópodes, como o processo de muda ou ecdise, de modo geral, são bem conhecidas. Durante essa fase os animais perdem o exoesqueleto de quitina antigo e o substituem por um novo e maior, permitindo o crescimento no período pós-ecdise e entre as mudas (intermudas) (Yudkovski *et al.*, 2007; Zhang *et al.*, 2014). No entanto, muito pouco ainda é conhecido sobre o mecanismo de ação de genes e proteínas determinantes para o crescimento dos crustáceos (Lv *et al.*, 2014; Zhang *et al.*, 2014).

Nos últimos anos, as técnicas de sequenciamento em larga escala ou *High Throughput Sequencing* (HTS), como o sequenciamento completo do transcriptoma (RNA-seq) tem possibilitado a descoberta de novos transcritos e as funções moleculares envolvidas em muitas características de interesse econômico, como crescimento e sobrevivência (Ghaffari *et al.*, 2014; Valenzuela-Miranda *et al.*, 2015), principalmente para grupos sem genoma de referência, como os camarões. Indivíduos de *Litopenaeus vannamei* pertencentes a famílias de maior e menor crescimento, criados sob as mesmas condições, foram coletados e uma análise de expressão diferencial foi realizada, com o objetivo de identificar os genes induzidos na comparação entre as condições de maior e menor crescimento, além de apontar algumas das proteínas relacionadas a desempenho nos camarões. Os resultados encontrados aqui trazem um conjunto de genes DE relacionados a crescimento em ambas as condições, acrescentando dados transcriptômicos importantes em uma área da fisiologia dos peneídeos em que pouca informação é conhecida.

2. Material e Métodos

2.1. Coleta de amostras e preparação das bibliotecas

A extração e sequenciamento de RNA, construção das bibliotecas e anotação dos transcritos estão descritas no Capítulo 3.

2.2. Genes Diferencialmente Expressos (DE)

Para a análise de genes diferencialmente expressos foi utilizado um total de 48 amostras de camarões *Specific Pathogen Free* (SPF) submetidos a um programa de melhoramento genético para rápido crescimento na empresa Genearch, localizada no estado do Rio Grande do Norte, Brasil. Destes, 24 camarões pertenciam a quatro famílias com as maiores médias de peso em gramas (alto crescimento), enquanto os outros 24 provieram de quatro famílias com as menores médias de peso

(baixo crescimento). No período da coleta, as famílias foram amostradas com peso variando de 0.99 – 1,91 g (Tabela 1) e cada uma estava representada em triplicata. As médias de peso das famílias de alto e baixo crescimento foram consideradas estatisticamente diferentes após a realização de um teste-t: alto crescimento (média = 1.74, desvio padrão = 0.13), baixo crescimento (média = 1.22, desvio padrão = 0.17), $t = -4.69$, $p = 0.003$ (bicaudal). Foram coletados seis camarões de cada família, sendo dois de cada uma das triplicatas, totalizando seis animais por família e 24 animais para cada um dos grupos analisados (Figura 1). Os *reads* das amostras utilizadas nas duas comparações foram mapeados contra a referência construída pelo Trinity, versão 2.1.1 (Grabherr *et al.*, 2011), utilizando-se o software Bowtie2, versão 2.2.6 (Langmead e Salzberg, 2012). Um maior detalhamento pode ser encontrado no Capítulo 2.

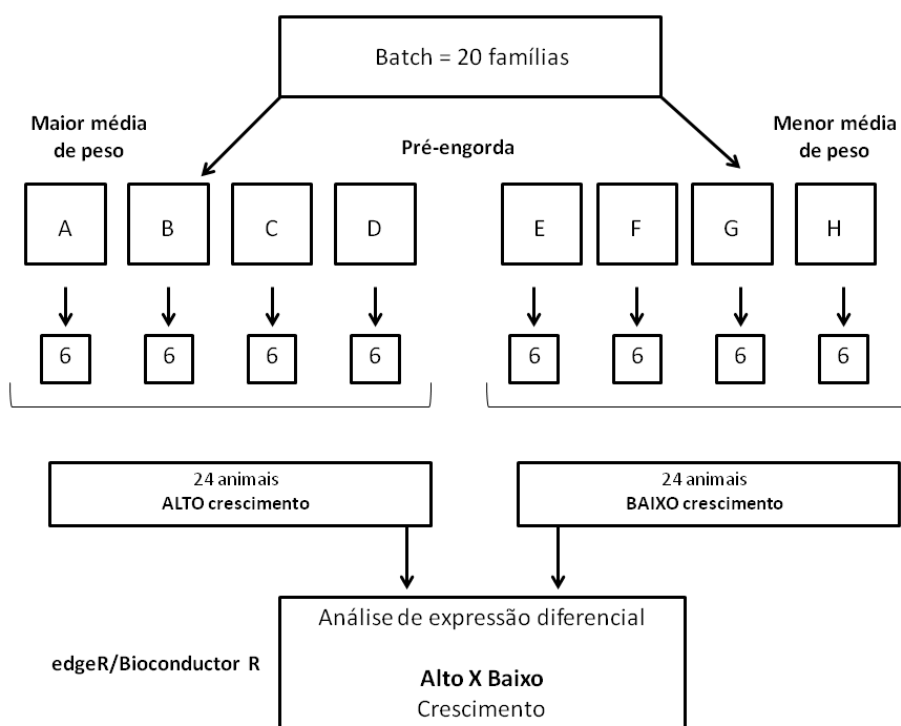


Figura 1: Esquema do desenho amostral dos 48 animais utilizados na análise de expressão diferencial no edgeR, para identificação dos genes DE entre Alto e Baixo crescimento.

Tabela 1: Peso médio, em gramas, das famílias avaliadas para alto e baixo crescimento no momento da coleta na fase de pré-engorda.

Condição de crescimento	Famílias	Peso (g)
Baixo	A	1.39
Baixo	B	1.31
Baixo	C	1.20
Baixo	D	0.99
Média		1.22
Alto	E	1.91
Alto	F	1.76
Alto	G	1.71
Alto	H	1.58
Média		1.74

2.2.1. Comparação entre as condições de alto e baixo crescimento

Para identificação dos genes diferencialmente expressos (DE), as condições de alto e baixo crescimento foram comparadas no pacote edgeR (Robinson *et al.*, 2010) implementado no software R/Bioconductor (versão 3.2, Gentleman *et al.*, 2004; Love *et al.*, 2015). O modelo adotado considerou como único fator o crescimento dos animais. Sendo assim, as dispersões foram estimadas utilizando-se a função estimateDisp(). Os dados foram ajustados de acordo com a distribuição binomial negativa do Teste Exato com a função exactTest(). Foi avaliada a diferença nos níveis de expressão dos genes dos animais de menor crescimento sobre os de maior crescimento, utilizando-se os animais pertencentes às mesmas famílias como réplicas biológicas. Os valores de log₂-FoldChange (log₂FC) e seus respectivos *p-values* foram calculados. Os valores de *p-value* ajustado ou *False Discovery Rate* (FDR), corrigido pelo método de Benjamini-Hochberg (BH) (Benjamini e Hochberg, 1995) foram obtidos, sendo que genes com FDR < 0.05 foram considerados diferencialmente expressos. Um *heatmap* dos 40 genes com menor valor de p corrigido foi plotado para ilustrar a diferença no perfil de expressão desses genes nas condições de alto e baixo

crescimento, utilizando-se os dados das contagens normalizados, sendo representado, no caso do edgeR, o log2 das contagens por milhão (logCPM).

2.3. Enriquecimento funcional no Gene Ontology (GO)

A anotação funcional das regiões codificantes dos transcritos da referência preditas pelo Transdecoder foi realizada pelo Trinotate (<https://trinotate.github.io/>). No *pipeline* do Trinotate, homologias nas sequências foram procuradas no banco de dados do Uniprot e posteriormente as categorias Processo Biológico, Função Molecular e Componente Celular do Gene Ontology (GO) (Ashburner *et al.*, 2000) foram atribuídas. Os termos do GO enriquecidos no conjunto de genes DEs foram detectados através do pacote GOseq, v. 1.24.0 (Young *et al.*, 2010), implementado no pacote R. As categorias do GO enriquecidas foram calculadas utilizando-se a aproximação de Wallenius (Wallenius, 1963) e os *p-values* foram ajustados através do método de BH. As categorias com *p-value* ajustado (FDR) < 0.1 foram consideradas como enriquecidas.

3. Resultados

3.1. Genes induzidos na comparação entre animais de Baixo X Alto crescimento

As 24 amostras de alto crescimento tiveram 831.930.442 *reads* mapeados, enquanto o número para as 24 amostras de baixo crescimento foi de 796.126.113. Isso totaliza 1.628.056.555 (96%) *reads* mapeados das amostras contra os 1.697.401.992 *reads* pós-filtragem das 48 amostras (Tabela S1 – Material Suplementar). Foram encontrados 207 genes DE (Tabela S2 - Material Suplementar), sendo 162 (78%) *up-regulated* para alto e 45 (22%) para baixo crescimento (FDR < 0,05) (Figura 2). Dentre os principais genes induzidos estão os (i) relacionados a crescimento e processo de muda (endoquitinase, constituintes da cutícula, ENSANGP00000021035-like e hormônio juvenil - JHE) e metabolismo de carboidratos (Pró-resilina, e transportador de trealose) e

(ii) imunidade (lectina, crustacianina, peptídeo antimicrobiano e fator anti-lipopolissacarídeo) e estresse (Proteínas de estresse térmico) (Tabela 2). Além disso, ao se verificar a contagem dos *reads* mapeados contra a referência, observou-se que dentre os 207 genes DE encontrados, 95 (46%) são exclusivos para alto crescimento. Dentre os genes DE exclusivos, nove possuem proteínas conhecidas em artrópodes, destacando-se as proteínas integrantes de cutícula, endoquitinase, pró-resilina e transportadora de trealose entre as relacionadas a desempenho, e as lectinas 2 e 3 envolvidas na resposta imune (Tabela 2). Todos os genes dessas proteínas mostraram-se fortemente induzidos para a condição de alto crescimento (valores de \log_2FC variando de -12.26 a -7.81).

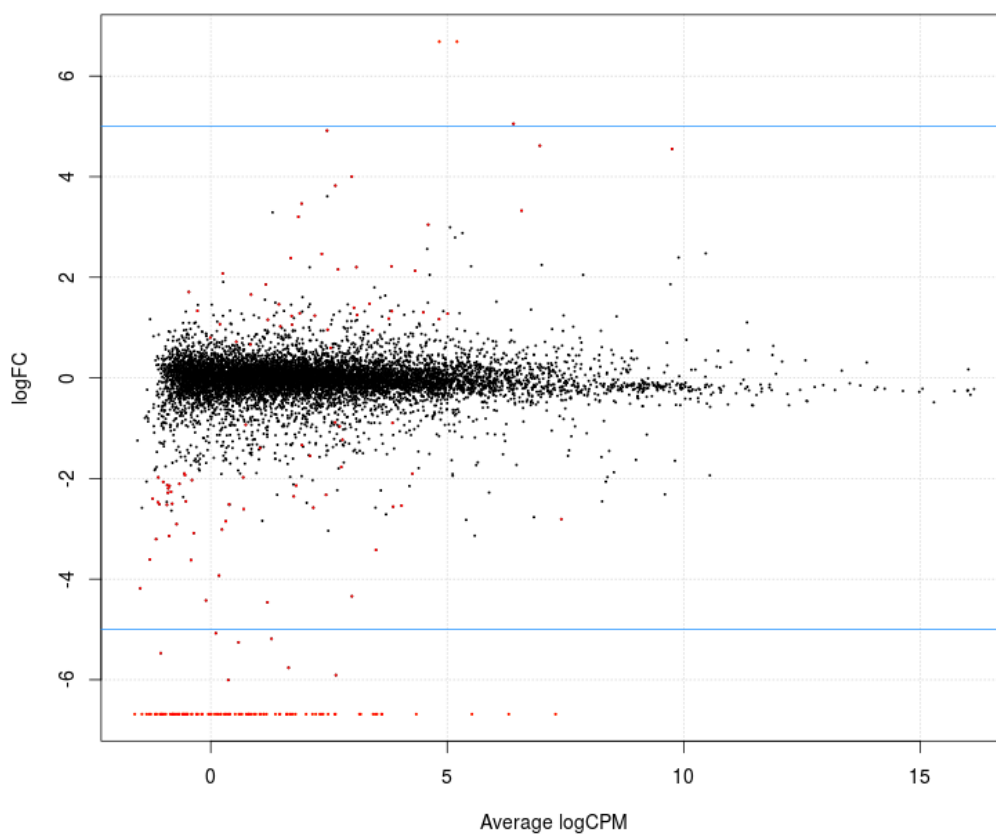


Figura 2: MA plot resultante da comparação entre alto (\log_2FC negativos) e baixo (\log_2FC positivos) crescimento. Os genes DE significativos são representados pelos pontos vermelhos ($FDR < 0.05$). Média dos valores de $\log_2FC = 5.07$ (linhas azuis).

Tabela 2: Genes candidatos envolvidos no ganho de peso e imunidade inata nos artrópodes na comparação alto (log2FC negativos) e baixo (log2FC positivos) crescimento. São mostrados também os genes DE exclusivos para alto crescimento (*).

Contig ID	No. acesso UniprotKB	Proteína	Símbolo do gene	log2FC FDR < 0.05	Condição de crescimento <i>up-regulated</i>
Crescimento e ecdise e metabolismo de açúcares					
*Locus_23910	Q7M4F4	Protéina SgAbd-1 estrutural da endocutícula	N/A	-12.2685	Alto
*Locus_22711	P36362	Endoquitinase	N/A	-11.3377	Alto
*Locus_3996				-9.9491	Alto
Locus_27571.	P81579	Proteína da cutícula AM1274	N/A	5.0517	Baixo
Locus_15931	P81578	Proteína da cutícula AM1239	N/A	2.1295	Baixo
Locus_11821	Q20AS9	ENSANGP0000021035-like	N/A	3.4640	Baixo
Locus_31558	F5A5Q6	Carboxilesterase, provável JHE	CXE1	-11.7558	Alto
Locus_23346	A0A023EFA6	Indutor do hormônio juvenil (JHE)	N/A	2.6354	Baixo
*Locus_16512	Q9V7U0	Pró-resilina	RESIL	-10.8468	Alto
Locus_9969				-9.8301	Alto
*Locus_755				-8.1978	Alto
*Locus_31894				-7.8168	Alto
*Locus_34246				-10.5516	Alto
Locus_25317	B0WC46	Facilitador do transporte de trealose	TRET1	-2.1402	Alto
Resposta imune e estresse					
*Locus_23169	B8R3L0	Lectina 2 tipo C	N/A	-10.1299	Alto
*Locus_22973	I6W5B6	Lectina 3	Lec3	-8.1098	Alto
Locus_17093	I6ZAW5	Precursor tipo 1 de peptídeo antimicrobiano	crusIb	-10.8047	Alto
Locus_21425	C0KJQ4	Fator lipopolissacarídeo (ALPS)	ALF	-1.9766	Alto
Locus_26180	P22978	Proteína de estresse térmico 67B2	Hsp67Bb	-5.4704	Alto
Locus_30488	Q5R7D3	Proteína 70 kDa de estresse térmico 1	HSPA1	-1.9323	Alto
Locus_31128	P02515	Proteína de estresse térmico 22	HSP22	1.2858	Baixo
Locus_30970	P80007	Subunidade A2 da crustacianina	N/A	-2.4706	Alto

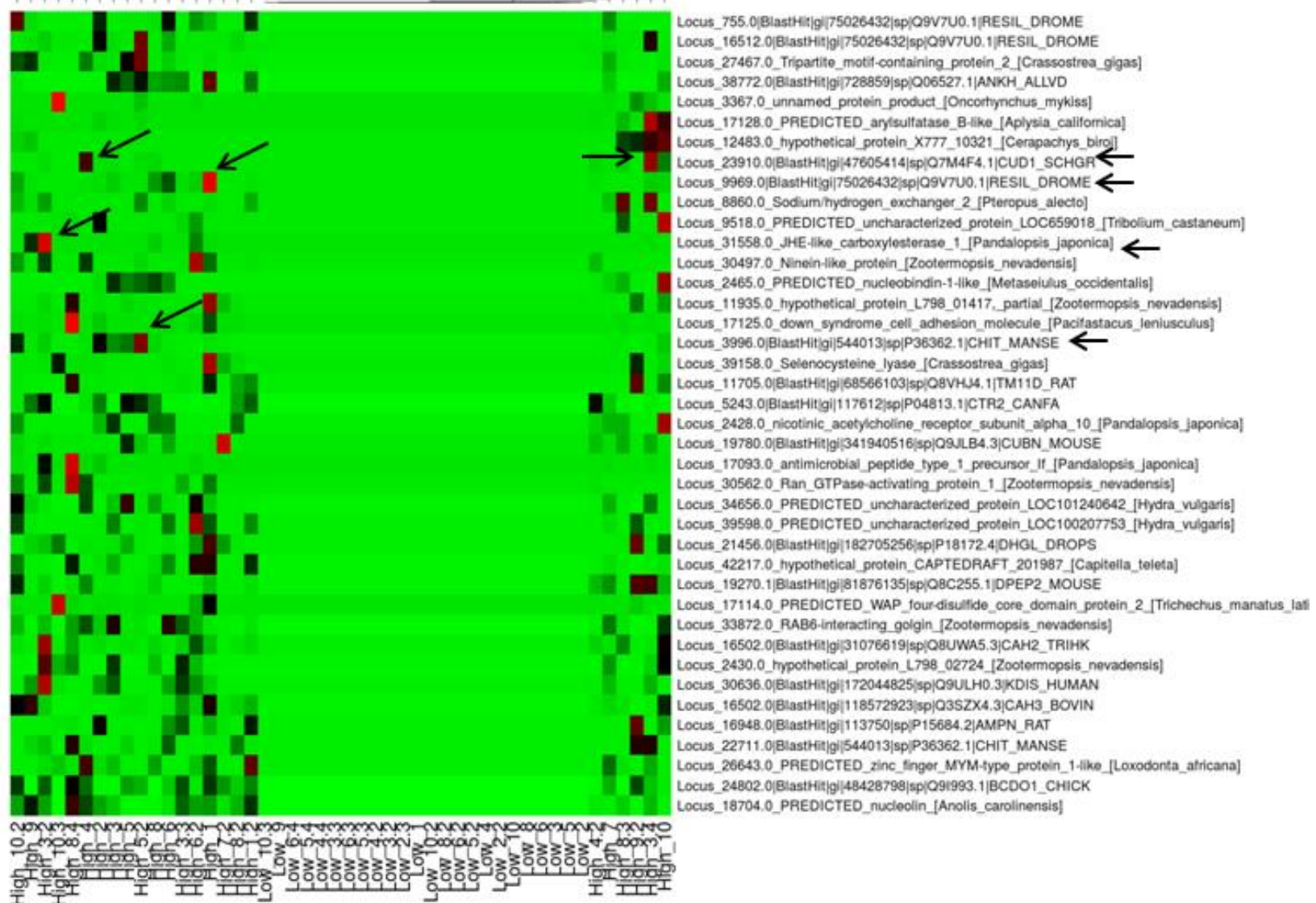


Figura 3 (Página 98): Representação dos 40 genes com menor valor de p-ajustado para a comparação entre os animais de alto (*High*) e baixo (*Low*) crescimento, utilizando-se os dados das contagens normalizados (logCPM). Notam-se genes mais expressos (quadrados vermelhos) apenas para a condição de maior crescimento. As setas pretas representam os genes mais induzidos nos animais de alto crescimento. A cor verde representa as amostras em que esses 40 genes apresentam baixa expressão. Identificação dos animais por condição e família: Baixo crescimento: Low_2, 3, 5, 6, 8 e 10 (Família A); Low_2.2, 4, 5.2, 6.2, 8.2 e 10.2 (Família B); Low_1, 2.3, 3.2, 4.2, 5.3 e 6.3 (Família C) e Low_3.3, 4.3, 5.4, 6.4, 9 e 10.3 (Família D). Alto crescimento: High_3, 5, 7, 8, 9 e 10 (Família E); High_3. 2, 4, 5.2, 7.2, 8.2, 1 (Família F); High_1.2, 3.3, 4.2, 5.3, 8.3 e 10.2 (Família G) e High_2, 3.4, 6.2, 8.4, 9.2 e 10.3 (Família H).

Baseando-se nos valores de log₂FC, os 40 genes com menor valor de *p-value* ajustado para a comparação de alto e baixo crescimento foram plotados em um *heatmap*, notando-se um perfil de expressão distinto entre as diferentes condições (Figura 3). Nos animais maiores, nota-se que algumas proteínas, tais como resilina, hormônio juvenil, estrutural da endocutícula e endoquitinase são fortemente expressas (quadrados vermelhos apontados na Figura 3). Por outro lado, não houve diferença evidente na expressão entre nenhum dos 40 genes observados nos animais de menor crescimento.

3.2. Enriquecimento no GO

Os genes diferencialmente expressos foram utilizados para realizar uma análise de enriquecimento funcional no pacote GOseq. Dez termos do GO significativamente enriquecidos foram encontrados (Tabela 3) para FDR < 0.1, sendo nove termos para os animais de alto e um para os de baixo crescimento. Dentre estes, termos do GO enriquecidos para os animais maiores foram

ligante de quitina (GO:0008061), metabolismo de quitina (GO:0006030), metabolismo de carboidratos (GO:1901135) e metabolismo de compostos organonitrogenados (GO:0097367). Para os animais de menor crescimento o único termo considerado enriquecido foi o de resposta de defesa a bactéria (GO:0042742).

Tabela 3: Resultado dos termos do GO dos genes diferencialmente expressos entre as condições de alto e baixo crescimento e as vias enriquecidas.

GO ID	Categoria*	Enriquecido no Tratamento	FDR	No. genes DE	Nome
GO:0008061	MF	Alto	1.60E-41	53	Chitin binding
GO:0006030	BP	Alto	1.96E-40	49	Chitin metabolic process
GO:1901071	BP	Alto	1.49E-39	49	Glucosamine-containing compound metabolic process
GO:0006040	BP	Alto	4.71E-39	49	Amino sugar metabolic process
GO:0006022	BP	Alto	7.77E-37	50	Aminoglycan metabolic process
GO:0005576	CC	Alto	1.13E-31	68	Extracellular region
GO:1901135	BP	Alto	5.81E-19	51	Carbohydrate derivative metabolic process
GO:1901564	BP	Alto	5.72E-15	53	Organonitrogen compound metabolic process
GO:0097367	MF	Alto	4.19E-13	59	Carbohydrate derivative binding
GO:0042742	BP	Baixo	0.085492	5	Defense response to bacterium

* Função Molecular (MF), Processo Biológico (BP) e Componente Celular (CC).

4. Discussão

Um maior número de genes *up-regulated* foi evidenciado para os animais de alto crescimento, sugerindo que estes demandam intensa síntese protéica e maior gasto energético, resultando em um possível ganho de peso mais rápido. Considerando a condição de alto crescimento, os genes das proteínas relacionadas à quitina (endoquitinase), estrutural da endocutícula, pró-resilina (RESIL), transportadora de trealose (TRET1) e hormônio juvenil (CXE1) (Tabela 2) apareceram fortemente induzidos nos animais de alto crescimento.

Uma das principais funções das quitinases é a digestão do exoesqueleto ou cutícula de quitina durante a muda, possibilitando o crescimento do animal. As quitinases são essenciais no processo de muda e crescimento, sendo que níveis irregulares dessa proteína podem causar deficiências no ganho de peso e até a morte (Li *et al.*, 2015). Esta enzima também atua no processo de maturação sexual dos crustáceos, uma vez que a precocidade na reprodução pode comprometer expressivamente o crescimento (Shechter *et al.*, 2008; Zhang *et al.*, 2014). Estudos realizados no caranguejo *Portunus trituberculatus* demonstraram que um metabolismo ativo e eficiente de quitina possibilita o maior crescimento corpóreo também nessa espécie (Lv *et al.*, 2014).

Em relação à pró-resilina, embora seu mecanismo de ação nos crustáceos ainda não esteja bem elucidado, já se sabe que esta proteína elástica é constituinte da cutícula dos artrópodes e parece ter relação com o processo de ecdise (Wong *et al.*, 2012). Além disso, a pró-resilina representa uma importante fonte de reserva energética, que se converte em açúcar de forma rápida e eficiente auxiliando, por exemplo, no deslocamento dos animais (Tynyakov *et al.*, 2015). Por sua vez, a trealose é o principal açúcar na hemolinfa dos crustáceos, atuando como principal fonte de energia em situações ambientais extremas, momento no qual atinge sua maior concentração na hemolinfa (Clegg *et al.*, 2000). A ecdise pode ser considerada uma etapa estressante para os artrópodes, os quais ficam muito vulneráveis e frágeis pela ausência de um exoesqueleto que atue na proteção. Podem ser constatados picos mais elevados de trealose nas fases anterior e posterior a muda, implicando também em um possível papel desse açúcar no metabolismo da quitina (Chung, 2008; Shi e Chung 2014).

O hormônio juvenil apareceu diferencialmente expresso nas condições de alto (CXE1) ($\log_2FC = -11.75$, $p = 0$) e baixo crescimento ($\log_2FC = 3.8216$, $p = 0.0004$), apresentando uma intensidade muito maior nos animais de alto crescimento. O hormônio participa do desenvolvimento e maturação sexual e o seu precursor, o metil farnesoato, é naturalmente

encontrado nos crustáceos, agindo no controle de alguns fatores de transcrição relacionados a muda, metamorfose e reprodução (Jindra *et al.*, 2013; Riddiford 2012). Essas proteínas juntas sugerem um cenário de intensa atividade muscular (crescimento) e metabolismo de carboidratos, provavelmente, relacionadas ao ganho de peso do camarão.

Sete genes DE exclusivos da condição de alto crescimento relacionados a desempenho nos artrópodes foram identificados: estrutural da endocutícula, TRET1, endoquitinase (Locus_22711 e Locus_3996) e RESIL (Locus_16512, Locus_755 e Locus_31894) (Tabela 2). Todos esses genes mostram-se bastante induzidos (log2FC variando de -12.26 a -7.81, $p < 0.05$) nos animais de alto crescimento, reforçando um cenário formado por camarões com intensa produção de proteínas relacionadas com o crescimento muscular. Uma evidência sobre os animais maiores apresentarem mais genes induzidos relacionados a crescimento pode ser observada na Figura 3, na qual nota-se a expressão de genes codificantes das proteínas resilina, estrutural de endocutícula e endoquitinase fortemente *up-regulated* apenas em animais de alto crescimento.

Nas famílias de menor crescimento, entre os genes mais induzidos ligados a desempenho estão os relacionados à estrutura da cutícula e o da proteína ENSANGP00000021035-like. Proteínas constituintes da cutícula dos artrópodes atuam na reconstrução do exoesqueleto e consequente na proteção do animal após a muda. O aumento da cutícula torna o exoesqueleto mais duro e resistente, dificultando ou até impedindo o crescimento do camarão (Li *et al.*, 2015). A função da proteína ENSANGP00000021035-like ainda não é amplamente conhecida, mas se sabe que esta possui função relacionada ao metabolismo de quitina durante a muda nos crustáceos. Como pode ser observado em Yudkovski *et al.*, (2007), essa proteína parece se ligar a quitina do exoesqueleto, podendo participar da digestão ou catabolismo de quitina no lagostim *Cherax quadricarinatus*. Nos animais de menor média de peso, nota-se que as proteínas relacionadas à reconstrução da cutícula são induzidas de maneira menos intensa quando comparadas as proteínas dos animais de maior

peso (Tabela 2), ilustrando uma velocidade de crescimento possivelmente menor e justificando, em parte, o menor peso observado nesses animais. Proteínas atuantes na digestão da quitina não são observadas como induzidas nos animais menores, atrasando a ruptura do exoesqueleto antigo e podendo constituir outro fator limitante do crescimento (Zhang *et al.*, 2014; Li *et al.*, 2015).

Dentre os genes relacionados à imunidade inata dos artrópodes, destacaram-se alguns genes exclusivamente induzidos nos animais de alto crescimento, incluindo as lectinas 2 ($\log_2FC = -10.12$, $p = 0$) e 3 (Lec3) ($\log_2FC = -8.10$, $p = 0$) (Lec3). As lectinas aparecerem entre os nove genes DE exclusivos para os animais de alto crescimento. Estas proteínas possuem forte atividade antiviral e antibacteriana e são armazenadas e induzidas com maior frequência em tecidos responsáveis pela imunidade nos crustáceos como hepatopâncreas, intestino e hematócitos. A lectina-C, como a do tipo 2 encontrada aqui, parece ter ação mais direta no reconhecimento, fagocitose, encapsulamento e aglutinação do patógeno, impedindo a sua disseminação nos tecidos (Tassanakajon *et al.*, 2013; Wang e Wang, 2013). Já o gene do peptídeo antimicrobiano *cruslb* ($\log_2FC = -10.80$, $p = 0$) apresenta atividade contra fungos e principalmente bactérias gram-negativas (Zhang *et al.*, 2010). Considerando que estas proteínas apareceram expressas mais ativamente nos animais de alto crescimento, é possível que estes animais possuam uma resposta imune mais eficiente e conseqüentemente um controle mais rápido dos possíveis efeitos danosos causados pelo patógeno (Lushchak, 2011; Tassanakajon *et al.*, 2013; Tassanakajon *et al.*, 2014), mesmo vivendo em um ambiente SPF.

Proteínas participantes de respostas ao estresse (genes *Hsp67Bb*, *HSPA1* e *HSP22*) também foram identificadas nos camarões maiores, como as Proteínas de Choque Térmico ou *Heat Shock Proteins* (HSPs), porém não expressas tão intensamente quanto as de imunidade. Cada tipo de HSP é induzido perante variantes de estresse específicos, tais como temperatura (Baringou *et al.*, 2016) e infecção por patógeno (Zhang *et al.*, 2013), por exemplo. Os animais utilizados nesse trabalho

foram coletados de um ambiente cativo com altas densidades populacionais, o que pode ser uma justificativa plausível para o encontro de proteínas relacionadas a estresse. Apesar dos animais amostrados serem originados de um ambiente SPF com controle de patógenos, algumas proteínas de imunidade apresentaram-se fortemente induzidas nos animais de alto crescimento, sugerindo uma resposta imune perante o contato com patógenos por parte desses animais. Já se sabe que existe uma relação antagônica entre as características envolvidas com crescimento e imunidade em crustáceos (Argue *et al.*, 2002; Cock *et al.*, 2009). Os animais precisam ser selecionados para crescimento e imunidade de maneira conjunta para se alcançar um equilíbrio e conseqüentemente boas taxas de sobrevivência. Dessa forma, os animais aqui amostrados foram submetidos a um programa de crescimento e sobrevivência, demonstrando bons resultados em relação ao ganho de peso. Paralelamente, proteínas de imunidade estão sendo continuamente produzidas, o que acarreta em respostas imunes mais eficiente mediante contato com um possível patógeno.

Em relação aos termos enriquecidos do GO, nota-se a predominância de proteínas ligantes de quitina e metabolismo de carboidratos nos animais de maior crescimento. O metabolismo de açúcares pode evidenciar uma grande demanda de energia pelo organismo para a manutenção de funções vitais de um modo geral, mas também sugere um alto metabolismo do açúcar quitina, constituinte do exoesqueleto dos artrópodes. Termos relacionados à produção e degradação de quitina caracterizam a muda, evento que possibilita o crescimento e ganho de peso do camarão (Lv *et al.*, 2014; Li *et al.*, 2015). A identificação desses termos como enriquecidos apenas nos animais de alto crescimento, reforça a relação de que as proteínas encontradas aqui contribuam para o crescimento do animal e aumento do peso corpóreo.

Considerações finais

A grande maioria dos genes diferencialmente expressos encontrado neste estudo é de proteínas relacionadas à ecdise, construção e digestão de cutícula e metabolismo de açúcares. Esses genes foram induzidos majoritariamente na condição de alto crescimento, indicando um metabolismo mais atuante nesses animais e a forte relação que esses produtos gênicos possuem com o crescimento e ganho de peso nos crustáceos. Em menor quantidade, porém induzidas significativamente nos camarões maiores, foram observadas também proteínas responsáveis por respostas imunes, como lectinas e peptídeos antimicrobianos, caracterizando uma pronta resposta imune nesses animais. Este trabalho apresenta uma variedade de proteínas descritas, contribuindo efetivamente para enriquecer uma área com limitada informação genômica disponível, como é o caso do crescimento de crustáceos, além de proporcionar uma maior compreensão do perfil de expressão do transcriptoma de crescimento desse grupo. No entanto, um grande número de proteínas hipotéticas e com produto desconhecido foi encontrado, enfatizando a necessidade de mais esforços na tentativa da identificação da função de proteínas chave para o ganho de peso nos crustáceos.

Agradecimentos

Os autores agradecem o apoio financeiro das agências de fomento: Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) (Processo número 2012/17322-8), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), além do Centro de Genômica Funcional do Laboratório de Biotecnologia Animal e ao Prof. Dr. Luiz Lehmann Coutinho (ESALQ-USP), Piracicaba-SP, Brasil, pelo serviço de sequenciamento e uso do servidor para processamento dos dados.

Material Suplementar

Tabela S1: Número de *reads* resultantes do sequenciamento Illumina, após a filtragem no SeqyClean e após o mapeamento para as bibliotecas de alto e baixo crescimento. **(Pág. 149)**

Tabela S2: Os 207 genes DE para a comparação alto X baixo crescimento com seus respectivos valores de log₂FC_b, *p* value ajustado (FDR) e log dos valores dos dados das contagens transformados por milhão (logCPM). **(Pág. 150)**

Referências

- Argue, B. J., Arce, S. M., Lotz, J. M., & Moss, S. M. (2002). Selective breeding of Pacific white shrimp *Litopenaeus vannamei* for growth and resistance to Taura Syndrome Virus.
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., & Harris, M. A. (2000). Gene Ontology: tool for the unification of biology. *Nature genetics*, 25(1), 25-29.
- Baringou, S., Rouault, J.-D., Koken, M., Hardivillier, Y., Hurtado, L., & Leignel, V. (2016). Diversity of cytosolic HSP70 Heat shock protein from decapods and their phylogenetic placement within Arthropoda. *Gene*. <http://doi.org/10.1016/j.gene.2016.06.061>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, 289-300.
- Chen, K., Li, E., Li, T., Xu, C., Wang, X., Lin, H., ... Chen, L. (2015). Transcriptome and molecular pathway analysis of the hepatopancreas in the Pacific white shrimp *Litopenaeus vannamei* under chronic low-salinity stress. *PLoS ONE*, 10(7), 1–22. <http://doi.org/10.1371/journal.pone.0131503>
- Cock, J., Gitterle, T., Salazar, M., & Rye, M. (2009). Breeding for disease resistance of Penaeid shrimps. *Aquaculture*, 286(1–2), 1–11. <http://doi.org/10.1016/j.aquaculture.2008.09.011>
- Ghaffari, N., Sanchez-Flores, A., Doan, R., Garcia-Orozco, K. D., Chen, P. L., Ochoa-Leyva, A., ... Criscitiello, M. F. (2014). Novel transcriptome assembly and improved annotation of the whiteleg shrimp (*Litopenaeus vannamei*), a dominant crustacean in global seafood mariculture. *Scientific Reports*, 4, 7081. <http://doi.org/10.1038/srep07081>
- Jindra, M., Palli, S. R., & Riddiford, L. M. (2013). The Juvenile Hormone Signaling Pathway in Insect Development. *Annual Review of Entomology*, 58(1), 181–204. <http://doi.org/10.1146/annurev-ento-120811-153700>
- Jung, H., Lyons, R. E., Hurwood, D. a., & Mather, P. B. (2013). Genes and growth performance in crustacean species: A review of relevant genomic studies in crustaceans and other taxa. *Reviews in Aquaculture*, 5(2), 77–110. <http://doi.org/10.1111/raq.12005>
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods*, 9(4), 357–359. <http://doi.org/10.1038/nmeth.1923>
- Love, M. I., Anders, S., Kim, V., & Huber, W. (2015). RNA-Seq workflow: gene-level exploratory analysis and differential expression. *F1000Research*, 4, 1070. <http://doi.org/10.12688/f1000research.7035.1>
- Lv, J., Liu, P., Gao, B., Wang, Y., Wang, Z., Chen, P., & Li, J. (2014). Transcriptome analysis of the *Portunus trituberculatus*: De novo assembly, growth-related gene identification and marker discovery. *PLoS ONE*, 9(4). <http://doi.org/10.1371/journal.pone.0094055>
- Manfred G. Grabherr, Brian J. Haas, Moran Yassour, Joshua Z. Levin, Dawn A. Thompson, Ido Amit, Xian Adiconis, Lin

- Fan, Raktima Raychowdhury, Qiandong Zeng, Zehua Chen, Evan Mauceli, Nir Hacohen, Andreas Gnirke, Nicholas Rhind, Federica di Palma, Bruce W., N., & Friedman, and A. R. (2013). Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature Biotechnology*, 29(7), 644–652. <http://doi.org/10.1038/nbt.1883>.Trinity
- Rao, R., Bhassu, S., Bing, R. Z. Y., Alinejad, T., Hassan, S. S., & Wang, J. (2016). A transcriptome study on *Macrobrachium rosenbergii* hepatopancreas experimentally challenged with white spot syndrome virus (WSSV). *Journal of Invertebrate Pathology*, 136, 10–22. <http://doi.org/10.1016/j.jip.2016.01.002>
- Robinson, M. D., & Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology*, 11(3), R25. <http://doi.org/10.1186/gb-2010-11-3-r25>
- Tassanakajon, A., Somboonwiwat, K., & Amparyup, P. (2014). Sequence diversity and evolution of antimicrobial peptides in invertebrates. *Developmental and Comparative Immunology*, 48(2), 324–341. <http://doi.org/10.1016/j.dci.2014.05.020>
- Tassanakajon, A., Somboonwiwat, K., Supungul, P., & Tang, S. (2013). Discovery of immune molecules and their crucial functions in shrimp immunity. *Fish & Shellfish Immunology*, 34(4), 954–67. <http://doi.org/10.1016/j.fsi.2012.09.021>
- Tynyakov, J., Bentov, S., Abehsera, S., Khalaila, I., Manor, R., Katzir Abilevich, L., Sagi, A. (2015). A Novel Chitin Binding Crayfish Molar Tooth Protein with Elasticity Properties. *Plos One*, 10(5), e0127871. <http://doi.org/10.1371/journal.pone.0127871>
- Valenzuela-Miranda, D., Del Río-Portilla, M. A., & Gallardo-Escárate, C. (2015). Characterization of the growth-related transcriptome in California red abalone (*Haliotis rufescens*) through RNA-Seq analysis. *Marine Genomics*, 24, 199–202. <http://doi.org/10.1016/j.margen.2015.05.009>
- Wang, X. W., & Wang, J. X. (2013). Pattern recognition receptors acting in innate immune system of shrimp against pathogen infections. *Fish and Shellfish Immunology*, 34(4), 981–989. <http://doi.org/10.1016/j.fsi.2012.08.008>
- Wong, D. C. C., Pearson, R. D., Elvin, C. M., & Merritt, D. J. (2012). Expression of the rubber-like protein, resilin, in developing and functional insect cuticle determined using a *Drosophila* anti-rec 1 resilin antibody. *Developmental Dynamics*, 241(2), 333–339. <http://doi.org/10.1002/dvdy.23724>
- Young, M. D., Wakefield, M. J., Smyth, G. K., & Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biology*, 11(2), R14. <http://doi.org/10.1186/gb-2010-11-2-r14>
- Zhang, F., Jiang, K., Sun, M., Zhang, D., & Ma, L. (2013). Multiplex immune-related genes expression analysis response to bacterial challenge in mud crab, *Scylla paramamosain*. *Fish and Shellfish Immunology*, 34(2), 712–716.
- Zhang, S., Jiang, S., Xiong, Y., Fu, H., Sun, S., Qiao, H., ... Gong, Y. (2014). Six chitinases from oriental river prawn *Macrobrachium nipponense*: cDNA characterization, classification and mRNA expression during post-embryonic development and moulting cycle. *Comparative Biochemistry and Physiology - B Biochemistry and Molecular Biology*, 167(1), 30–40. <http://doi.org/10.1016/j.cbpb.2013.09.009>
- Zhang, Y., Wang, L., Wang, L., Yang, J., Gai, Y., Qiu, L., & Song, L. (2010). The second anti-lipopolysaccharide factor (EsALF-2) with antimicrobial activity from *Eriocheir sinensis*. *Developmental and Comparative Immunology*, 34(9), 945–952. <http://doi.org/10.1016/j.dci.2010.04.002>

Capítulo 5

**Manuscrito 6 - ANÁLISE DO TRANSCRIPTOMA DE HEPATOPÂNCREAS E MÚSCULO DE CAMARÕES
LITOPENAEUS VANNAMEI EXPOSTOS AO VÍRUS DA SÍNDROME DA MANCHA BRANCA (WSSV)
ATRAVÉS DE RNA-SEQ**

Camilla A. Santos¹, Karin Kurkjian², Ana Carolina Guerrelhas³, Sônia C.S. Andrade⁴ e Patrícia D. Freitas¹

¹ Departamento de Genética e Evolução, Universidade Federal de São Carlos, São Carlos, SP, Brazil.

² Genearch Aquacultura, Pititinga, Rio do Fogo, Rio Grande do Norte, RN, Brazil.

³ Aquatec Larvicultura de Camarão Marinho, Barra do Cunhaú, Canguaretama, Rio Grande do Norte, RN, Brazil.

⁴ Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, SP, Brazil.

Correspondence to: Camilla Alves Santos. Via Washington Luiz km 235, Phone+55-16-3351-8309, Fax:+55-16-3351-8377. Caixa Postal 676, CEP13565-905, São Carlos-SP, Brazil.

E-mail: camilla.alves@yahoo.com.br

Resumo

O vírus causador da Síndrome da Mancha Branca ou *White Spot Syndrome Virus* (WSSV) consiste numa das principais ameaças ao cultivo do camarão branco do Pacífico *Litopenaeus vannamei*, uma das espécies de crustáceos comercialmente mais importantes para aquicultura mundial. Nesse estudo, a abordagem de sequenciamento de próxima geração, RNA-seq foi utilizada para identificar proteínas envolvidas nos processos de resposta imune de crustáceos e investigar as diferenças entre o perfil de expressão dos genes induzidos no hepatopâncreas e no músculo de indivíduos de *L. vannamei* infectados e não infectados pelo WSSV. Foram encontrados 5.816 genes diferencialmente expressos (DE) na comparação entre os tecidos, sendo 3.096 (53%) *up-regulated*

para hepatopâncreas e 2.720 (47%) para músculo. Para as condições saudáveis X doentes, considerando o efeito dos tecidos, foram identificados 1.017 genes DE induzidos nos animais saudáveis e 620 nos doentes, com uma maior diferença na resposta do hepatopâncreas à infecção. Entre as principais proteínas caracterizadas destacaram-se as relacionadas a estresse e ação antimicrobiana, com expressão principalmente no hepatopâncreas. Após o enriquecimento funcional no Gene Ontology, 663 termos significativamente enriquecidos foram encontrados no hepatopâncreas e músculo e 446 nos animais saudáveis e doentes. Esse estudo mostra a diferença marcante entre os perfis de expressão dos tecidos amostrados e destaca o papel fundamental do hepatopâncreas na imunidade dos crustáceos, evidenciando o diferente comportamento desse órgão em animais infectados e não infectados pelo WSSV.

Palavras-chave: *L. vannamei*, Genes Diferencialmente Expressos, Imunidade Inata, AMPs, PRPs.

1. Introdução

O cultivo de camarão marinho consiste numa atividade mundialmente expressiva, com destaque especial para a espécie *Litopenaeus vannamei*, a qual apresenta boas taxas de reprodução em cativeiro, suportando altas densidades populacionais, diversos níveis de salinidade e baixas concentrações de oxigênio (Perez-Henriquez *et al.*, 2009; Fan *et al.*, 2016). Apesar de esse peneídeo ser um dos mais comercializados no mundo, a alta incidência de patógenos nos tanques de engorda é uma ameaça constante ao seu cultivo. Vírus como os causadores da Síndrome da Mancha Branca (*White Spot Syndrome Virus*; WSSV) (Rao *et al.*, 2016), Síndrome de Taura (*Taura Syndrome Virus*; TSV) (Sookruksawong *et al.*, 2013) e Cabeça Amarela (*Yellow Head Virus*; YHV) (Junkunlo *et al.*, 2011) se espalham muito rapidamente no ambiente aquático, com uma velocidade ainda maior em cativeiro (La Veja *et al.*, 2008; Leu *et al.*, 2013). Na tentativa de combater o patógeno, o organismo sofre um forte estresse metabólico devido à produção de radicais reativos de oxigênio (*Reactive Oxygen Species*; ROS), que podem causar danos à estrutura celular e do DNA,

levando a erros nos processos de transcrição, tradução e apoptose (Rewitz *et al.*, 2006; Qian *et al.*, 2014).

Nos invertebrados, as respostas imunes são totalmente dependentes do sistema inato (Wang *et al.*, 2009; Tassanakajon *et al.*, 2013). Apesar da vasta gama de proteínas atuantes na imunidade nos decápodes, como lectinas (Wang *et al.*, 2013), crustacianina (Fan *et al.*, 2016), hemocianina (Zheng *et al.*, 2016), entre outras, e as várias estratégias adotadas no combate a micro organismos, como endocitose (Liu *et al.*, 2014), reconhecimento de padrões específicos (Song *et al.*, 2010) e aglutinação da hemolinfa (Cheng *et al.* 2008), ainda existem alguns patógenos letais para os camarões. Um exemplo disso é o vírus da Mancha Branca, causando uma doença que chega a alcançar 100% de mortalidade nos tanques de cultivo e sem tratamento (Wang *et al.*, 2005; Gajula *et al.*, 2013). Os principais sintomas já aparecem poucos dias após a infecção e incluem principalmente cutícula frouxa, carapaça de cor avermelhada, letargia, manchas brancas e necrose, causando enormes prejuízos à carcinicultura a cada ano (FAO, 2016). A síndrome é causada por um vírus envelopado de dupla fita de DNA da família *Nimaviridae* e sabe-se que as proteínas contidas no envelope são as principais responsáveis pela invasão, reconhecimento celular e proliferação viral no hospedeiro, além de desencadear fatores imunes (Kulkarni *et al.*, 2014).

Nos crustáceos, a maioria das proteínas participantes da imunidade é produzida e armazenada nos hemócitos, que devido ao sistema circulatório aberto dos crustáceos, se espalham por todo o organismo, incluindo o hepatopâncreas. Ele é um dos órgãos mais atuantes na imunidade inata nos camarões e tem função semelhante ao fígado, pâncreas e intestino dos vertebrados, atuando na absorção de minerais e vitaminas, além de participar de outros processos metabólicos (Imjongjirak *et al.* 2011; Wang *et al.*, 2013). Como um órgão integrante da imunidade e de metabolismo muito ativo o hepatopâncreas está envolvido na detecção e eliminação de patógenos, produção de antígenos perante severa proliferação viral e mudanças metabólicas

induzidas pelo estágio infeccioso (Alday-Sanz *et al.*, 2002; Fan *et al.*, 2016). Sendo assim, sabendo quais mRNA são expressos em órgãos específicos sob condições de infecção possibilita uma maior compreensão da fisiologia imune dos camarões, viabilizando a aplicação desses dados na aquicultura. O ideal para a aquicultura é a seleção de camarões resistentes à infecção pelos principais patógenos letais, como o WSSV, sem a necessidade do uso de pesticidas (Rewitz *et al.*, 2006, Qian *et al.*, 2014).

Apesar de esforços relacionados ao melhor entendimento da imunidade contra o WSSV nos decápodes (Ye *et al.*, 2015) e também contra outros patógenos como o TSV (Zeng *et al.*, 2013), YHV (Soowannayan *et al.*, 2015), muitas proteínas imunológicas e seus respectivos mecanismos de ação ainda permanecem desconhecidos, tornando os estudos genéticos e fisiológicos relacionados a estas questões cada vez mais necessários. Com a falta de um genoma descrito para *L. vannamei* ou mesmo para espécies próximas, as bases de dados obtidos a partir de transcriptomas tornam-se essenciais na identificação de marcadores moleculares e função de novos genes e proteínas que possam levar ao preenchimento de lacunas no metabolismo imune nos peneídeos (Rao *et al.*, 2015; Sun *et al.*, 2015). O sequenciamento do transcriptoma, conhecido como RNA-seq, tem sido empregado para ilustrar uma fotografia de um momento específico, que serve de base para a anotação funcional dos transcritos (Young *et al.*, 2010). Essa abordagem é uma das maneiras mais práticas e eficientes de se identificar praticamente todos os genes e vias metabólicas envolvidas em uma questão biológica específica (Powell *et al.* 2015).

Nesse estudo, a metodologia de RNA-seq foi utilizada para identificar proteínas envolvidas na imunidade inata de crustáceos e testar a hipótese de que os padrões de expressão dos genes induzidos diferem entre o hepatopâncreas e músculo de camarões infectados e não infectados pelo vírus da Mancha Branca. Com esse propósito, duas comparações independentes foram conduzidas para identificação de genes DE: (i) no hepatopâncreas e músculo de indivíduos de *L. vannamei* e (ii)

nas condições de animal 'saudável' e 'doente', levando-se em consideração o efeito dos tecidos nos padrões de expressão.

2. Material e Métodos

2.1. Coleta de amostras e preparação das bibliotecas

A extração e sequenciamento de RNA, construção das bibliotecas e anotação dos transcritos estão descritas no Capítulo 3.

2.2. Genes Diferencialmente Expressos (DE)

Os *reads* das amostras utilizadas nas duas comparações foram mapeados contra a referência construída pelo Trinity, versão 2.1.1 (Grabherr *et al.*, 2011), utilizando-se o software Bowtie2, versão 2.2.6 (Langmead e Salzberg, 2012). Um maior detalhamento pode ser encontrado no Capítulo 2. Foram coletados o hepatopâncreas (HP) e o tecido muscular (MU) de 20 camarões de um tanque de cultivo com alta incidência da doença do vírus da Mancha Branca, 10 com sintomas e 10 assintomáticos. A presença/ausência do vírus foi verificada através de qPCR (*Quantitative Real-Time Polymerase Chain Reaction*). Após a qPCR, verificou-se que dentre os 10 animais que apresentavam sintomas, quatro apresentavam a doença (doentes), e dentre os 10 sem sintomas, todos os 10 estavam sem a doença (saudáveis). Assim, os quatro animais doentes e mais quatro saudáveis foram utilizados nas análises, totalizando 16 bibliotecas pareadas de cDNA, já que para cada um dos oito animais amostrados foram sequenciados os tecidos hepatopâncreas e músculo (Figura 1). As 16 bibliotecas seguiram para as análises de genes DE entre (i) os tecidos HP X MU e (ii) os tratamentos Doente X Saudável, considerando-se o efeito do tecido.

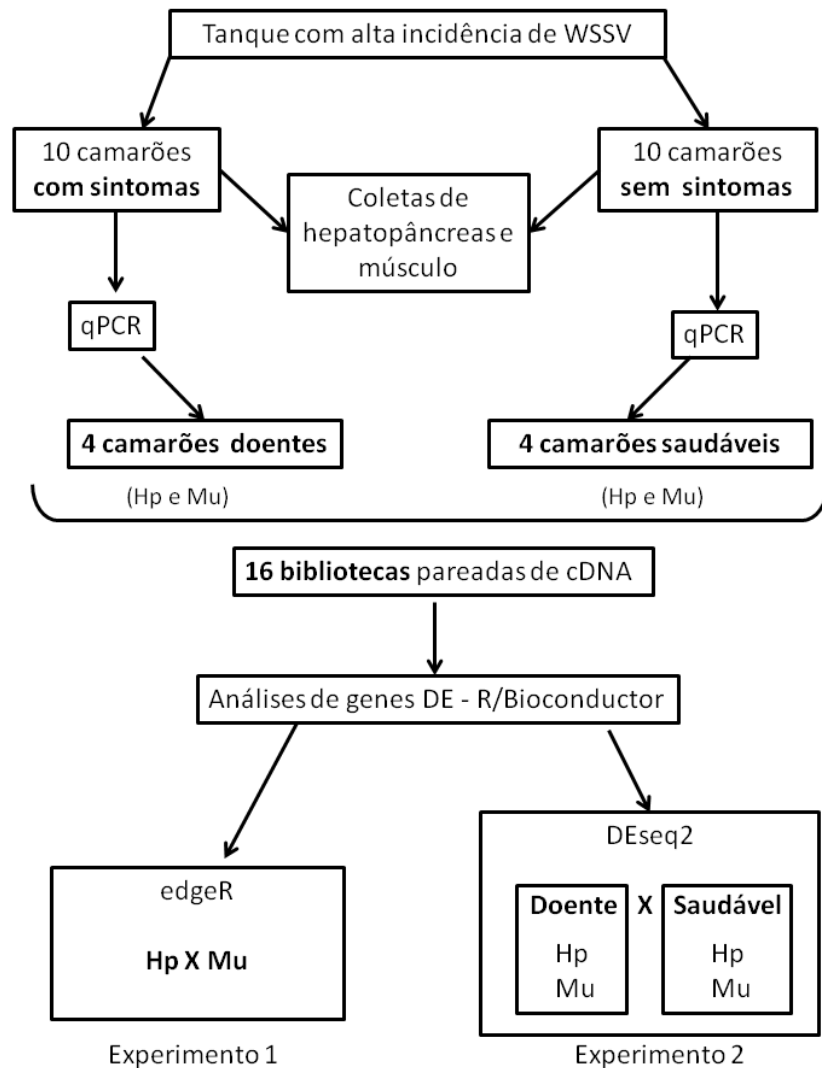


Figura 1: Esquema amostral dos oito animais utilizados no experimento de exposição ao WSSV. Foram sequenciados os tecidos hepatopâncreas e músculo de animais doentes e saudáveis, totalizando 16 bibliotecas de cDNA.

2.2.1. Experimento 1: Comparação pareada entre tecidos

Para identificação dos genes diferencialmente expressos entre o tecido muscular e o hepatopâncreas, foi realizada a comparação par-a-par na qual são analisados os dois tecidos para o mesmo indivíduo. Dessa forma, foi aplicada a abordagem *blocking factor* do pacote *edgeR* (Robinson *et al.*, 2010), do *R/Bioconductor* (v. 3.2, Gentleman *et al.*, 2004; Love *et al.*, 2015). No

edgeR foi aplicado um modelo aditivo que considera a dependência entre as amostras. Os dados foram ajustados de acordo com a distribuição binomial negativa do modelo linear generalizado (GLM). O GLM foi então ajustado aos dados com a função `glmFit()` e os *p-values* foram calculados utilizando-se o teste da razão de verossimilhança com a função `glmLRT()` (Nelder e Wedderburn, 1972). Os valores de \log_2 -FoldChange (\log_2FC) foram calculados juntamente com os *p-values*, os quais foram corrigidos pelo método de Benjamini-Hochberg (BH) (Benjamini e Hochberg, 1995). Os genes com $FDR < 0.05$ foram considerados significativos para expressão diferencial, sendo aqueles com \log_2FC negativo *up-regulated* para hepatopâncreas e positivos *up-regulated* para músculo. Um *heatmap* das amostras foi plotado para ilustrar como estas estão agrupadas baseando-se nas suas distâncias euclidianas e utilizando-se os dados das contagens transformados, sendo no caso do edgeR, o \log_2 das contagens por milhão (\log_2CPM).

2.2.2. Experimento 2: Comparação entre tratamentos considerando o efeito do tecido

Para a comparação entre os tratamentos doente x saudável foi utilizado o DEseq2 (Anders e Huber, 2010), outro pacote do R/Bioconductor. Para verificar o efeito das condições 'saudável' X 'doente', levando em consideração os tipos de tecidos 'hepatopâncreas' e 'músculo' amostrados, um modelo multifatorial com interação foi adotado, ao qual foram incorporados as condições e os tipos de tecido descritos anteriormente. A normalização foi realizada ajustando a distribuição dos dados segundo uma distribuição binomial negativa, seguida da filtragem dos *contigs* removendo aqueles com valores de `baseMean` < 5 . Os valores de \log_2FC e seus respectivos *p-values* foram calculados com base nas duas variáveis, tecido e condição. Os valores de *p-values* ajustados foram calculados através do método de BH e apenas os genes com $FDR < 0.05$ foram considerados significativos para expressão diferencial.

2.3. Enriquecimento funcional no Gene Ontology (GO)

A anotação funcional dos transcritos da referência montada pelo Trinity com regiões codantes de proteínas preditas pelo Transdecoder foi realizada no Trinotate (<https://trinotate.github.io/>), um pacote desenvolvido para anotação funcional de transcriptomas. No *pipeline* do Trinotate, homologias nas sequências foram procuradas no banco de dados do Uniprot e posteriormente os termos Processo Biológico, Função Molecular e Componente Celular do Gene Ontology (GO) (Ashburner *et al.*, 2000) foram atribuídos.

Para identificar as anotações do GO enriquecidas no conjunto de genes diferencialmente expressos nas duas análises realizadas quando comparado ao genoma de referência montado, foram realizadas análises no pacote GOSec, v.1.24.0 (Young *et al.*, 2010), implementado no software R. As categorias do GO enriquecidas foram calculadas utilizando a aproximação de Wallenius (Wallenius, 1963) e os *p-values* foram ajustados através do método de BH. As categorias com *p-value* ajustado (FDR) < 0.1 foram consideradas como enriquecidas.

3. Resultados

As 16 amostras utilizadas nos dois experimentos realizados totalizaram 786.662.168 *reads* após a filtragem, dos quais 755.946.418 (96%) foram mapeados contra a referência (Tabela S1 - Material Suplementar).

3.1. Experimento 1: comparação entre hepatopâncreas e músculo

Para as comparações entre tecidos (HP X MU) foram encontrados 5.816 genes DE significativos dos quais 3.096 (53%) são *up-regulated* para hepatopâncreas e 2.720 (47%) para músculo (Figura 2).

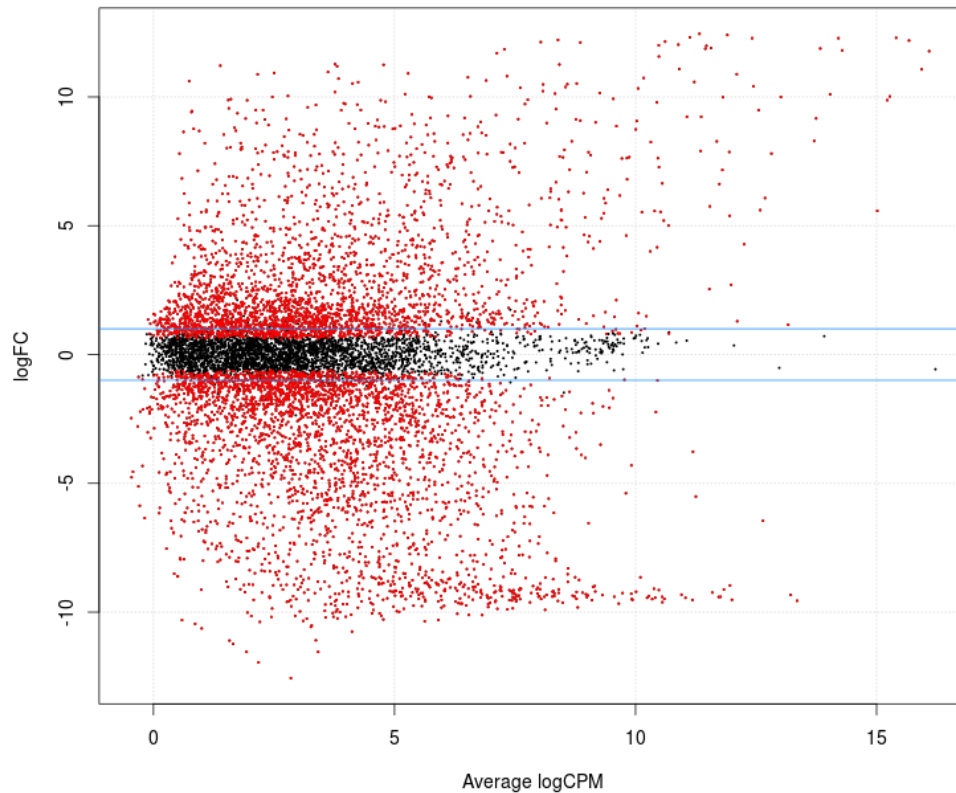


Figura 2: MAplot resultante da comparação entre hepatopâncreas e músculo. Os genes DE significativos são representados pelos pontos vermelhos ($FDR < 0.05$). Os pontos negativos são *up-regulated* para hepatopâncreas (\log_2FC negativos), os positivos para músculo (\log_2FC positivos). Média dos valores de $\log_2FC = 0.6$ (linhas azuis).

Dentre os principais genes DE identificados na comparação entre os tecidos hepatopâncreas e músculo (Tabela S2 – Material Suplementar), com funções relacionadas à imunidade nos crustáceos estão: peptídeos com atividade antimicrobiana (AMPs), proteínas de atividade antimicrobiana, proteases e inibidores de proteases, proteínas de reconhecimento padrão (PRPs) e outras relacionadas à coagulação de hemolinfa, estresse térmico e morte celular (Tabela 1).

Tabela 1: Genes candidatos envolvidos na resposta imune contra patógenos nos decápodes para a comparação entre os tecidos hepatopâncreas (HP) ($\log_2FC < 0.5$) e músculo (MU) ($\log_2FC > 0.5$) no experimento 1. (WSSV: White Spot Syndrome Virus).

Contig ID	No. acesso UniprotKB / Espécie <i>BlastX</i> hit	Proteína/Micro organismo alvo	Símbolo do gene	\log_2FC (FDR < 0.05)	Condição <i>up-regulated</i>
Peptídeos Antimicrobianos (AMP)					
Locus_23466	M4HXU5 <i>Fenneropenaeus chinensis</i>	Fator anti-lipo polissacarídeo isoforma 5 Fungos, WSSV	ALF5	-9,387	HP
Locus_28086	C0KJQ4 <i>Portunus trituberculatus</i>	Fator anti-lipopolissacarídeo Fungos, WSSV	ALF	3.108	MU
Locus_41665	B9A7T8 <i>Marsupenaeus japonicus</i>	Fator 2 anti-lipopolissacarídeo Fungos, WSSV	MjALF2	1.31	MU
Locus_17729	G1FET2 <i>Litopenaeus vannamei</i>	LITAF Gram-negativa bactéria, WSSV	-	-3,159	HP
Locus_24551	C3U0K7 <i>Penaeus monodon</i>	Crustina Pm5 Bactéria Gram-positiva	-	-8,195	HP
Locus_20875	C7B728 <i>Pandalopsis japonica</i>	Crustina 1 Bactéria Gram-positiva	PJC1	-9,328	HP
Locus_22671	C7B729 <i>Pandalopsis japonica</i>	Crustina 2 Bactérias Gram-positiva e negativa	PJC2	-4,969	HP
Locus_15620	A5A3L2 <i>Pacifastacus leniusculus</i>	PI-crustina 2	N/A	-4,762	HP
Locus_22939	C7B731 <i>Pandalopsis japonica</i>	Crustina 4	PJC4	-2,309	HP
Locus_14945	I7AFQ9 <i>Pandalopsis japonica</i>	Precursor IIb de peptídeo antimicrobiano tipo 2	crusIIa	-9.111	HP
Locus_24086	T1W2K4 <i>Pandalopsis japonica</i>	Precursor IIc de peptídeo antimicrobiano tipo 2	crusIIc	-3.895	HP
Locus_1836	T1W2K4 <i>Pandalopsis japonica</i>	Antimicrobial peptide type 2 precursor IIc	crusIIc	8.973	MU
Locus_14748	T1W2K0 <i>Pandalopsis japonica</i>	Precursor Ic de peptídeo antimicrobiano tipo 1	crusIc	7.627	MU
Locus_21300	A8VIQ7 <i>Litopenaeus stylirostris</i>	Stilicina 2 Fungos	N/A	8.536	MU
Locus_25051	Q95NT0 <i>Litopenaeus vannamei</i>	Peneidina-4a Bactéria Gram-positiva Fungos	N/A	0.965	MU
Proteínas Antimicrobianas					

Contig ID	No. acesso UniprotKB / Espécie <i>BlastX</i> hit	Proteína/Micro organismo alvo	Símbolo do gene	log2FC (FDR < 0.05)	Condição <i>up-regulated</i>
Locus_22579	P10787 <i>Panulirus interruptus</i>	Cadeia B de hemocianina Vírus	N/A	-11.539	HP
Locus_5575	P80096 <i>Panulirus interruptus</i>	Cadeia C de hemocianina Vírus	N/A	-10.517	HP
Locus_11646	P58989 <i>European lobster</i>	Subunidade A1 de crustacianina WSSV	N/A	-9.336	HP
Locus_30970	P80007 <i>European lobster</i>	Subunidade A2 de crustacianina WSSV	N/A	-7,905	HP
Locus_29427	P80029 <i>European lobster</i>	Subunidade C1 de crustacianina WSSV	N/A	5.282	MU
Locus_25022	X2FU51 <i>Penaeus monodon</i>	Proteína de resposta viral	VRP15	-2,569	HP
Locus_23798	B9U440 <i>Penaeus monodon</i>	Proteína ligante de fortilina 1 WSSV	N/A	4.25	MU
Cascatas de Proteases					
Locus_4067	V5WCC7 <i>Penaeus monodon</i>	Serpina 3 Bactéria Gram-negativa WSSV	N/A	-9,381	HP
Locus_39380	Q26058 <i>Pacifastacus leniusculus</i>	Inibidor de serina protease Bactéria Gram-negativa WSSV	PAPI 2	-7,452	HP
Locus_25241	CAA56043 <i>Pacifastacus leniusculus</i>	PAPI I Bactéria	PAPI I	-9,349	HP
Locus_6831	C5I7U2 <i>Fenneropenaeus chinensis</i>	Serina protease	N/A	-2.6259	HP
Locus_15233	C5I7U1 <i>Fenneropenaeus chinensis</i>	Serina protease	N/A	6.395	MU
Proteínas Reconhecimento Padrão (PRPs)					
Locus_21160	S4USJ7 <i>Fenneropenaeus chinensis</i>	Proteína com domínio de lectina tipo C	AV	-8.33	HP
Locus_21160	E2CV64 <i>Marsupenaeus japonicus</i>	Lectina B isoforma 2 WSSV	LecB	-6.38	HP
Locus_19724.	E2CV61 <i>Marsupenaeus japonicus</i>	Lectina E	LecE	-5,277	HP
Locus_23463	E2CV63 <i>Marsupenaeus japonicus</i>	Lectina B isoforma 3 WSSV	LecB	-5,077	HP
Locus_16621	E2CV66	Lectina A isoforma 4	LecA	-4,099	HP

Contig ID	No. acesso UniprotKB / Espécie <i>BlastX</i> hit	Proteína/Micro organismo alvo	Símbolo do gene	log2FC (FDR < 0.05)	Condição <i>up-</i> <i>regulated</i>
	<i>Marsupenaeus japonicus</i>	WSSV			
Locus_25035	S5TC87 <i>Fenneropenaeus merguensis</i>	Lectina 3 tipo C WSSV	LC3	-1,375	HP
Locus_21128	S5U2R4 <i>Fenneropenaeus merguensis</i>	Lectina 5 tipo C	LC5	7.918	MU
Locus_24257	B8R3L3 <i>Fenneropenaeus chinensis</i>	Lectina 3 tipo C WSSV	N/A	5.948	MU
Locus_23348	Q0IJY0 <i>Fenneropenaeus chinensis</i>	Lectina tipo C WSSV	N/A	4.322	MU
Locus_14675	F8UN26 <i>Litopenaeus vannamei</i>	Proteína com domínio de lectina tipo C WSSV	CTL	1.815	MU
Locus_23573	V9P9Q9 <i>Marsupenaeus japonicus</i>	Lectina C tipo 1	Lec1	1.272	MU
Locus_22833	L7UV69 <i>Marsupenaeus japonicus</i>	Proteína antiviral (PmAV) WSSV	N/A	3.591	MU
Locus_26651	Q8N0N3 <i>Penaeus monodon</i>	Proteína ligante de beta- 1.3 glicanos Bactéria Gram-negativa, Fungos	N/A	-9,243	HP
Locus_11178	K7X5F8 <i>Litopenaeus vannamei</i>	Fator de ativação da feniloxidase	ppaf	4.76	MU
Outras					
Locus_38840	X2D760 <i>Litopenaeus vannamei</i>	Fator 1 de estresse térmico	N/A	-2,466	HP
Locus_23131	Q25008	Cadeia alfa 1 da tubulina	N/A	-1,386	HP
Locus_29990	<i>Homarus americanus</i>	Bactéria, Fungos, WSSV	N/A	6.436	MU
Locus_22151	G5CJW5 <i>Litopenaeus vannamei</i>	Spz3 Bactéria, Fungos, WSSV	N/A	-1.113	HP
Locus_18737	M4M6F0 <i>Litopenaeus vannamei</i>	p53	N/A	6.172	MU
Locus_30850 Locus_7409	Q9U572 <i>Penaeus monodon</i>	Coagulante de hemolinfa	N/A	5.61	MU
Locus_11736	P21902 <i>Tachypleus tridentatus</i>	Enzima pró coagulante	N/A	5.179	MU

A distância euclidiana entre os valores de expressão das diferentes amostras nos tecidos hepatopâncreas e músculo foi ilustrada em um *heatmap* (Figura 3). Como se pode observar, as amostras de músculo e hepatopâncreas estão completamente separadas em clados distintos, indicando uma grande diferença entre os perfis de expressão dos genes desses tecidos.

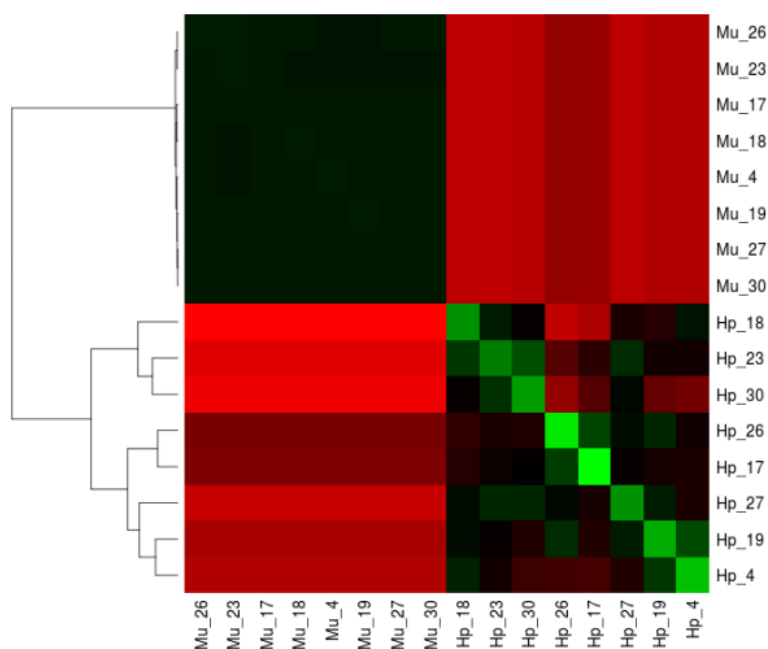


Figura 3: *Heatmap* das amostras de hepatopâncreas (HP) e músculo (MU), baseado nos valores das contagens por milhão normalizados (logCPM). Notar a grande diferença entre os perfis de expressão dos genes dessas amostras (dois quadrantes vermelhos). A intensidade das cores indica a heterogeneidade entre os valores dissimilares de expressão gênica (vermelho), percebendo-se que o tecido hepatopâncreas apresenta uma maior heterogeneidade de perfis de expressão.

Na análise de enriquecimento funcional no GO para os genes DE foram encontrados 323 e 340 termos do GO enriquecidos ($FDR < 0.1$), respectivamente para hepatopâncreas e músculo. A anotação nesse banco de dados fornece uma relação hierárquica para a representação das informações de Processo Biológico (BP), Função Molecular (MF) e Componente Celular (CC),

possibilitando a atribuição de detalhes mais precisos aos produtos gênicos identificados. No hepatopâncreas se destacam os termos lisossomo (GO:0005764) e capsídeo viral (GO:0019028), como CC; atividade de glutathione transferase (GO: 0004364) como MF e regulação positiva da resposta imune (GO: 0002827) e aprimoramento da imunidade adaptativa (GO: 0035397) como BP. Para músculo alguns dos principais termos enriquecidos foram: organização estrutural da actomiosina (GO: 0031032), organização do sarcômero (GO: 0045214) e contração muscular (GO: 0006936) como BP; estrutura muscular (GO: 0008307) como MF e filamento de miosina (GO: 0032982) como CC (Tabela 2).

Tabela 2: Resumo do resultado dos termos Processo Biológico (BP), Função Molecular (MF) e Componente Celular (CC) do GO dos genes DE da comparação entre hepatopâncreas e músculo e as vias enriquecidas.

GO ID	Categoria	Tecido	Ajusted <i>p-value</i>	No. Genes DE	Nome
GO:0005764	CC	HP	2.23-10	85	Lisossomo
GO:0019028	CC	HP	0.000107	12	Capsídeo viral
GO:0004364	MF	HP	0.003921	15	Atividade de glutathione transferase
GO:0002827	BP	HP	0.003921	10	Regulação positiva da resposta imune
GO:0035397	BP	HP	0.003921	10	Aprimoramento da imunidade adaptativa
GO:0031032	BP	MU	3.33E-27	70	Organização estrutural da actomiosina
GO:0045214	BP	MU	4.24E-22	48	Organização do sarcômero
GO:0032982	CC	MU	1.24E-22	42	Filamento de miosina
GO:0008307	MF	MU	2.24E-22	42	Estrutura muscular
GO:0006936	BP	MU	3.32E-18	51	Contração muscular

3.2. Experimento 2: comparação entre tratamentos considerando o efeito do tecido

Entre os tratamentos (doentes X saudáveis) foram encontrados 1.637 genes DE, sendo 1.017 (62%) *up-regulated* nos animais saudáveis e 620 (38%) nos doentes (Figura 4). Dentre os genes relacionados à resposta imune destacam-se crustina, AMPs, stílicina, serina protease, lectinas,

estresse térmico (Hsp), tubulinas e coagulantes de hemolinfa (Tabela S3 – Material Suplementar), todos induzidos em animais saudáveis (Tabela 3).

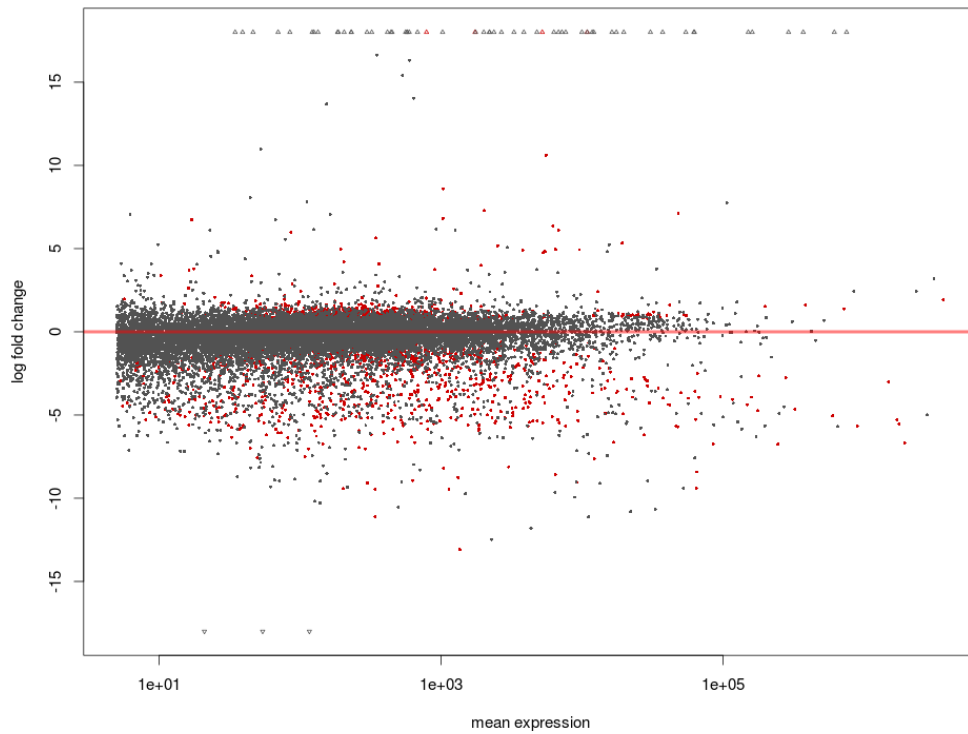


Figura 4: MAplot resultante da comparação entre as condições doente e saudável. Os genes DE significativos são representados pelos pontos vermelhos (FDR < 0.05). Os pontos negativos são genes *up-regulated* para animais saudáveis e os positivos para doentes.

Tabela 3: Genes candidatos envolvidos na resposta imune encontrados na comparação entre animais doentes e saudáveis no Experimento 2.

Contig ID	No. acesso UniprotKB /Espécie <i>BlastX</i> hit	Proteína/Micro-organismo alvo	Símbolo do gene	log2FC (FDR < 0.05)	Condição <i>up-regulated</i>
Peptídeos Antimicrobianos (AMP)					
Locus_22939	C7B731 <i>Panulirus japonicus</i>	Crustina 4	PJC4	-3.1466	Saudável
Locus_4717	AFO68120 <i>Macrobrachium rosenbergii</i>	Crustina Vírus (WSSV)	N/A	-1.4177	Saudável
Locus_1836	T1W2K4 <i>Pandalopsis japonica</i>	Precursor Ilc de peptídeo antimicrobiano tipo 2	crusIlc	-5.3593	Saudável
Locus_14748	T1W2K0 <i>Pandalopsis japonica</i>	Precursor Ic de peptídeo antimicrobiano tipo 1	crusIc	-4.3684	Saudável
Locus_21300	A8VIQ7 <i>Litopenaeus stylirostris</i>	Stilicina 2 Fungos	N/A	-4.6893	Saudável
Cascatas de Proteases					
Locus_6831	C517U2 <i>Fenneropenaeus chinensis</i>	Serina protease	N/A	-2.6259	Saudável
Proteínas Reconhecimento Padrão (PRPs)					
Locus_23348	Q0IJY0 <i>Fenneropenaeus chinensis</i>	Lectina tipo C WSSV	N/A	-2.7328	Saudável
Locus_22345	V5N901 <i>Marsupenaeus japonicus</i>	Lectina 2 tipo C	N/A	-4.6893	Saudável
Locus_22833	L7UV69 <i>Litopenaeus vannamei</i>	Proteína antiviral (PmAV) WSSV	N/A	-2.2747	Saudável
Outras					
Locus_7392	H8XYP5 <i>Macrobrachium rosenbergii</i>	Proteína de estresse térmico 21	N/A	-2.7389	Saudável
Locus_29990	Q25008 <i>Homarus americanus</i>	Cadeia alfa 1 da tubulina Bactéria, Fungos, WSSV	N/A	-3.4719	Saudável
Locus_7409	Q9U572 <i>Penaeus monodon</i>	Coagulante de hemolinfa	N/A	-4.4394	Saudável
Locus_30046	Q27081 <i>Tachypleus tridentatus</i>	Fator B de coagulação	N/A	-7.0415	Saudável
Locus_11736	P21902 <i>Tachypleus tridentatus</i>	Enzima pró coagulante	N/A	-11.0865	Saudável

WSSV: White Spot Syndrome Virus

O *heatmap* entre as amostras pertencentes a cada uma das condições ‘saudável’ e ‘doente’, está apresentado na Figura 5. Pode-se observar que os tecidos estão separados em grupos

diferentes e para o hepatopâncreas foi observada diferença nos padrões de expressão entre os animais doentes e saudáveis, o que não é perceptível no tecido muscular.

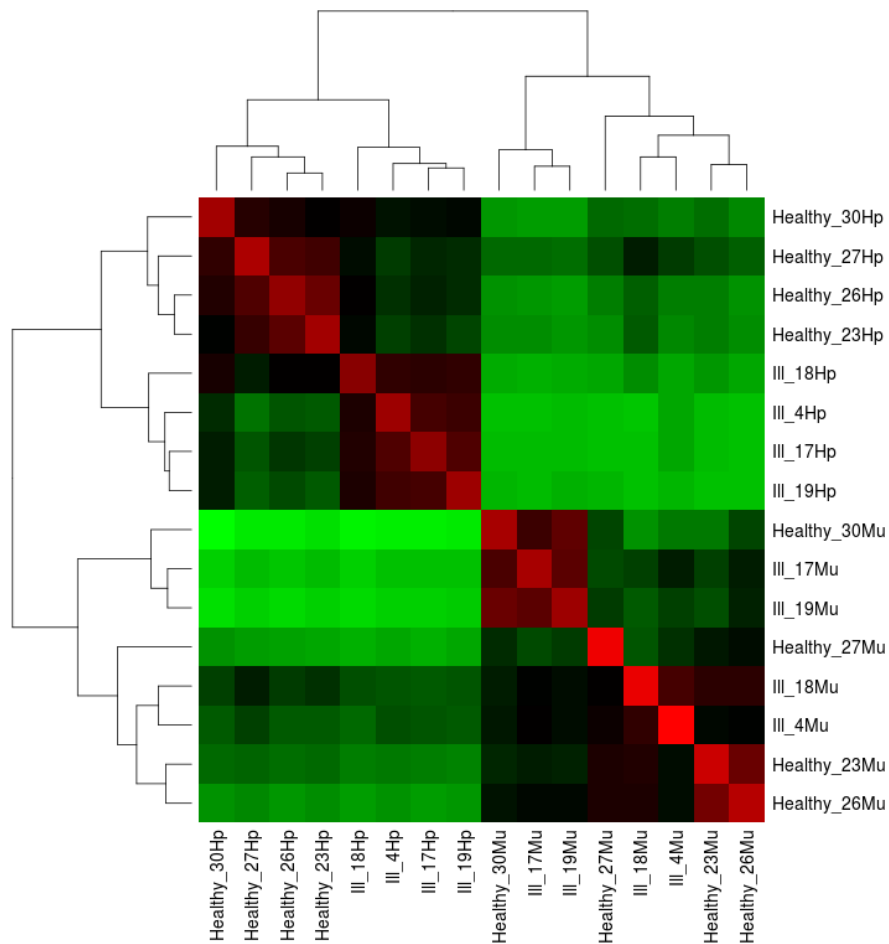


Figura 5: Distância entre as amostras dos animais saudáveis e doentes considerando-se a informação do tecido, com base nos dados das contagens por milhão (logCPM). Nota-se as amostras de hepatopâncreas e músculo separadas em diferentes grupos. Em relação à condição ‘doente’ (Ill) e ‘saudável’ (Healthy) da infecção por WSSV, as amostras são agrupadas separadamente apenas no hepatopâncreas. A cor verde indica a intensidade dos valores dissimilares de expressão gênica, com o hepatopâncreas apresentando uma maior heterogeneidade entre os perfis de expressão de animais doentes e saudáveis.

Após a análise de enriquecimento funcional do GO, 337 e 109 termos foram identificados como enriquecidos para os animais saudáveis e doentes, respectivamente. Nos animais não infectados pelo WSSV, a categoria Processo Biológico é a que aparece mais frequentemente apresentando *hits* principalmente para proteínas musculares, estando estas relacionadas à organização do citoesqueleto de actina (GO:0030036), filamentos de actina (GO:0030029), organização da estrutura de actomiosina (GO:0031032), organização do sarcômero (GO:0045214) e montagem do filamento fino de miosina no músculo (GO:0030241). Para Função Molecular os principais papéis proteicos encontrados foram ligantes (GO:0005515) e estrutural do músculo (GO:0008307). Como Componente Celular enriquecido merece destaque o filamento de miosina (GO:0032982). Já nos animais doentes, o Componente Celular foi a categoria mais frequente com termos relacionados a complexo ribonucleoprotéico (GO:0030529), nucléolo (GO:0005730), ribossomo (GO:0005840) e mitocôndria (GO:0005739). Como Função Molecular apareceram termos participantes da estrutura do ribossomo (GO:0003735) e ligante de RNA (GO:0003723) como enriquecidos. Para o Processo Biológico destacaram-se os relacionados a tradução (GO:0006412) (Tabela 4).

Tabela 4: Resumo dos termos Biológico (BP), Função Molecular (MF) e Componente Celular (CC) do GO dos genes diferencialmente expressos entre os animais saudáveis e doentes e as vias enriquecidas.

GO ID	Categoria	Tratamento	Ajusted <i>p</i> -value	No. Genes DE	Nome
GO:0005515	MF	Saudável	4.70E-16	360	Ligante de proteína
GO:0030036	BP	Saudável	2.94E-12	60	Organização do citoesqueleto de actina
GO:0044430	CC	Saudável	4.51E-12	118	Citoesqueleto
GO:0030029	BP	Saudável	5.06E-12	63	Filamento de actina
GO:0008307	MF	Saudável	5.27E-12	25	Estrutural do músculo
GO:0031032	BP	Saudável	4.64E-11	34	Organização estrutural da actomiosina
GO:0045214	BP	Saudável	1.94E-10	26	Organização do sarcômero
GO:0030241	BP	Saudável	2.27E-10	23	Montagem do filamento fino de miosina
GO:0032982	CC	Saudável	2.27E-10	23	Filamento de miosina
GO:0030529	CC	Doente	1.32E-39	113	Complexo ribonucleoproteico
GO:0003735	MF	Doente	1.09E-28	55	Constituinte do ribossomo
GO:0005730	CC	Doente	4.46E-27	83	Nucléolo
GO:0006412	BP	Doente	5.76E-27	56	Tradução
GO:0005840	CC	Doente	1.05E-26	55	Ribossomo
GO:0044428	CC	Doente	5.57E-24	163	Núcleo
GO:0003723	MF	Doente	9.25E-24	128	Ligante de RNA
GO:0005739	CC	Doente	1.69E-21	100	Mitocôndria

4. Discussão

Uma grande variedade de proteínas relacionadas à resposta imune de *L. vannamei* está sendo descrita neste trabalho, adicionando informações relevantes ao conjunto de dados já existente para o grupo dos camarões (Powell *et al.*, 2015; Nguyen *et al.*, 2016; Sun *et al.*, 2016). Foi observado um número expressivo de genes candidatos à resistência, sendo identificado um número maior de genes DE *up-regulated* para o hepatopâncreas quando comparado ao músculo nos camarões expostos ao WSSV amostrados (Tabela 1), corroborando estudos anteriormente realizados que destacam a relevância do hepatopâncreas, como órgão relacionado à imunidade dos crustáceos (Antony *et al.*, 2011; Guo *et al.*, 2013).

Na comparação entre as condições doente e saudável, foi encontrado um número maior de genes DE nos animais saudáveis (Tabela 3). Uma das possíveis justificativas para isso é o fato do metabolismo dos animais infectados ser menos ativo, provavelmente devido ao progresso dos sintomas da doença, resultando em um menor número de genes com alto padrão de expressão (Wang *et al.*, 2009; Tassanakajon *et al.*, 2013). Os genes DE identificados nas duas análises conduzidas são discutidos juntos, devido a participação significativa do hepatopâncreas na imunidade dos crustáceos e o diferente comportamento desse órgão em animais doentes e saudáveis.

4.1. Principais proteínas imunes encontradas em crustáceos

4.1.1. Peptídeos antimicrobianos

Uma grande quantidade de *hits* em crustáceos é observada para os peptídeos antimicrobianos ou *Antimicrobial peptides* (AMP) (Tabelas 2 e 3), como peneidinas, stilicinas, crustinas e *anti-lipopolysaccharide factors* (ALF), possuindo um papel essencial nas respostas imunes contra a invasão de patógenos nos invertebrados. São moléculas produzidas e estocadas nos hematócitos antes de serem liberadas na hemolinfa perante ativação por contato com lipopolissacarídeos (LPS) das paredes celulares de fungos e bactérias (Antony *et al.*, 2011; Imjongjirak *et al.*, 2011; Tassanakajon *et al.*, 2013; Tassanakajon *et al.*, 2014).

Estudos sugerem que alguns dos tipos de crustinas encontradas aqui atuem como AMPs contra bactérias gram-positivas, tais como os tipos 1 (PJC1) e 2 (PJC2) e gram-negativas, tipo 2 somente (Shockey *et al.*, 2009; Tassanakajon *et al.*, 2014). Aqui, os tipos 1 ($\log_2FC = -9,328$, $p = 0$), 2 ($\log_2FC = -4,969$, $p = 0$) e Pm5 ($\log_2FC = -8,195$, $p = 0$) aparecem induzidos exclusivamente no hepatopâncreas de *L. vannamei* (Tabela 1), enquanto o tipo 4 aparece nesse tecido ($\log_2FC = -2,309$, $p = 0.004$) e também nos animais saudáveis ($\log_2FC = -3.1466$, $p = 0.002$) (Tabela 3). Como

pode ser observado em Imjongjirak *et al.*, (2011), crustinas do tipo 1 mostraram atividade contra o crescimento de bactérias gram-positivas no caranguejo *Scylla paramamosai*. Já as crustinas do tipo 2 possuem uma resposta imune contra bactérias gram-positivas e negativas, dependendo da isoforma atuante (Tassanakajon *et al.*, 2014). Em *L. vannamei*, após o silenciamento do gene de crustina tipo 2 foi observado a proliferação da bactéria *Vibrio penaeicida*, sugerindo uma ligação dessa proteína com a inibição de bactérias gram-negativas (Shockey *et al.*, 2009). O tipo Pm5 encontrado aqui é exclusivo da espécie *Penaeus monodon*, mostrando atividade apenas contra gram-positivas, e pode ser considerado um indicador do estresse sofrido pelo animal devido à infecção (Vatanavicharn *et al.*, 2009). O tipo 4 já havia sido relacionado à resposta imune em camarões e caranguejos, sendo indicado como um possível inibidor de proteases nos crustáceos (Tassanakajon *et al.*, 2014). O que foi observado nos resultados do presente trabalho foi que as crustinas mostraram-se mais induzidas no hepatopâncreas, local onde muitas das proteínas produzidas pelos hemócitos são expressas e uma maior atividade das proteínas de defesa imunológica nos animais saudáveis, corroborando estudos prévios realizados (Brouwer *et al.*, 2008; Shockey *et al.*, 2009).

As peneidinas foram reportadas como sendo exclusivas dos camarões peneídeos, e estudos realizados nas espécies *Litopenaeus setiferus* e *L. vannamei* mostraram que estas proteínas são expressas principalmente nos hemócitos e possuem ação contra bactérias gram-positivas (Cuthbertson *et al.* 2004; Du e Jin, 2015). Estudos anteriores já haviam demonstrado a presença de peneidinas também na cutícula de peneídeos, relacionando sua função à atividade antifúngica e proteção sistêmica contra agentes externos (Destoumieux *et al.*, 2000). O presente trabalho demonstrou que estas proteínas podem ser encontradas expressas também no músculo, induzidas aqui exclusivamente no tecido muscular de *L. vannamei* (Tabela 1), adicionando uma informação relevante a esta categoria de peptídeos antimicrobianos.

Já as stilicinas são a família de AMPs mais recentemente descrita e estudos realizados na espécie *Litopenaeus stylirostris* demonstraram que esta proteína apresenta função antifúngica, mas seus mecanismos de ação ainda não são bem conhecidos (Rolland *et al.*, 2010; Tassanakajon *et al.*, 2013). Destoumieux *et al.* (2000) reportou que as stilicinas participam de processos de defesa mais generalizados, impedindo, por exemplo, a entrada de patógenos através da cutícula de animais no período de muda, em que o exoesqueleto dos crustáceos está mole ou mesmo ausente. Nesse trabalho, as stilicinas aparecem bastante induzidas no tecido muscular ($\log_2FC = 8.536$, $p = 0$) e nos animais saudáveis ($\log_2FC = -4.6893$, $p = 0.0001$), possivelmente como um reflexo da proteção sistêmica realizada por elas no organismo.

Em relação aos fatores lipopolissacarídeos (ALF), a resposta imune é observada quando há uma grande afinidade entre o ALF e o padrão molecular ao qual ele vai se ligar (Tassanakajon *et al.*, 2014). Essa família proteica facilita a fagocitose nos hemócitos, atuando como opsoninas ao ligar-se ao patógeno. Os AFLs possuem uma grande diversidade de isoformas, o que certamente aumenta a resposta imune a diferentes tipos infecção (Liu *et al.*, 2014). Nesse estudo foi encontrado um fator lipopolissacarídeo comum (ALF), a isoforma 5 (ALF5) e o fator 2 (MjALF2), todos com ação contra fungos e o vírus da Mancha Branca (WSSV) (La Veja *et al.*, 2008). O ALF5 aparece fortemente induzido no hepatopâncreas de *L. vannamei* ($\log_2FC = -9.387$, $p = 0$), mas os genes ALF ($\log_2FC = 3.108$, $p = 0$) e MjALF2 ($\log_2FC = 1.31$, $p = 0$) também são induzidos no músculo, embora em menor intensidade.

Outro exemplo da atuação dos AFLs na resposta imune contra o WSSV é o fator TNF- α induzível por lipopolissacarídeos ou *lipopolysaccharide-induced TNF- α factor* (LITAF) (Wang *et al.*, 2012). Aqui foi encontrada induzida no hepatopâncreas de *L. vannamei* (Tabela 1). Considerando que os animais amostrados no nosso estudo foram expostos ao WSSV, é esperado que respostas

imunes antivirais aconteçam mais ativamente no hepatopâncreas dos crustáceos (Imjongjirak *et al.* 2011; Wang *et al.*, 2013).

4.1.2. Proteínas antimicrobianas

Dentre as proteínas com ação antimicrobiana encontradas aqui estão as hemocianinas, crustacianinas, de resposta viral (VRP15), peritrofinas e fortilinas (Tabelas 1 e 3). As hemocianinas constituem uma das maiores famílias de proteínas relacionadas à defesa imune nos artrópodes e neste trabalho as cadeias B ($\log_2FC = -11.539$, $p = 0$) e C ($\log_2FC = -10.517$, $p = 0$) aparecem fortemente induzidas no hepatopâncreas de *L. vannamei*. Atua principalmente contra vírus culminando na aglutinação do patógeno na hemolinfa com consequente lise celular (Fan *et al.*, 2016; Zheng *et al.*, 2016), sendo encontrada aqui expressa majoritariamente no hepatopâncreas de *L. vannamei*.

As crustacianinas por sua vez, são exclusivas dos crustáceos se ligam a superfície da célula doente induzindo a apoptose e influenciando fortemente a regulação imune, principalmente mediante contato com o WSSV (Kulkarni *et al.*, 2014, Fan *et al.*, 2016). Neste trabalho foram encontradas as subunidades A1 ($\log_2FC = -9.336$, $p = 0$) e A2 ($\log_2FC = -7,905$, $p = 0$) ambas induzidas no hepatopâncreas e a subunidade C1 induzida no músculo ($\log_2FC = 5.282$, $p = 0$). No presente trabalho a crustacianina mostrou-se presente no hepatopâncreas e também no músculo de *L. vannamei*, auxiliando na regulação imune nesses animais e na proteção sistêmica contra patógenos sob situações de estresse.

As fortilinas parecem inibir consideravelmente o crescimento de uma ampla diversidade de vírus, incluindo o WSSV. Acredita-se que as fortilinas possam diminuir os níveis de expressão do WSSV, atuando desde o início do contágio na inibição da proliferação viral (Sinthujaroen *et al.*, 2015). As fortilinas estão relacionadas principalmente à resposta ao estresse e prevenção da

apoptose (Panrat *et al.*, 2012; Leu *et al.*, 2013). Aqui, esta classe de proteínas apareceu induzida no músculo de *L. vannamei* (Tabela 1), atuando provavelmente no controle da entrada de patógenos através da cutícula em *L. vannamei* (Destoumieux *et al.* 2000).

4.2. Cascatas de proteases

As serinas proteases (SPs) são uma classe de enzimas proteolíticas amplamente distribuídas entre os eucariotos e caracterizadas por participar da coagulação da hemolinfa e resposta imune (Homvises *et al.*, 2010; Amparyup *et al.*, 2013; Zhao *et al.*, 2014). A maioria das respostas imunes nos eucariotos é mediada por cascatas de proteinases e estas, reguladas por inibidores de serinas proteases (SERPINs ou serpins). Aqui, foram encontradas três serinas proteases induzidas no hepatopâncreas, músculo (Tabela 1) e nos animais saudáveis (Tabela 3). Dentre os inibidores de serinas proteases, todos os encontrados são fortemente induzidos no hepatopâncreas: serpina 3 ($\log_2FC = -9.38$, $p = 0$) e inibidor de serina protease (PAPI 2) ($\log_2FC = -7.45$, $p = 0$). Esses inibidores participam de processos como coagulação, apoptose e imunidade inata contra o WSSV (Homvises *et al.*, 2010; Somnuk *et al.*, 2012; Zhao *et al.*, 2014). Neste contexto, podemos sugerir que os inibidores de serina protease identificados aqui são induzidos majoritariamente no hepatopâncreas e apresentam atividade contra o WSSV (Tabela 1) (Homvises *et al.*, 2010; Somnuk *et al.*, 2012), tornando-se claro o envolvimento do hepatopâncreas e dessas proteínas na resposta imune desencadeada pela exposição ao WSSV também na espécie *L. vannamei*.

4.3. Lectinas e Proteínas de Reconhecimento Padrão

As proteínas de reconhecimento padrão ou *Pattern recognition proteins* (PRPs) são caracterizadas por reconhecerem padrões específicos nos patógenos, com destaque no caso dos artrópodes, para as lectinas tipo C (Wang *et al.*, 2009; Song *et al.*, 2010; Pathan *et al.*, 2013;

Tassanakajon *et al.*, 2013; Wang *et al.*, 2013). Esse reconhecimento é o primeiro passo no desencadeamento da resposta imune nos crustáceos e acontece majoritariamente com o domínio de reconhecimento de carboidratos (CRD) em moléculas de açúcar específicas.

Dentre as lectinas encontradas aqui induzidas no hepatopâncreas (Tabela 1), as que apresentam atividade contra o WSSV são: tipo B, isoformas 2 e 3 (LecB) de *Marsupenaeus japonicus*; tipo A isoforma 4 (LecA) de *M. japonicus* (Song *et al.*, 2010) e tipo C isoforma 3 (LC3) de *Fenneropenaeus merguensis* (Rattanaporn e Utarabhand, 2011). No músculo apareceram as isoformas 3 do tipo C de *Fenneropenaeus chinensis* (Wang *et al.*, 2009); tipo C de *F. chinensis* (Liu *et al.*, 2007) (também expressa nos animais saudáveis – Tabela 3) e proteína com domínio de lectina tipo C (CTLD) de *L. vannamei* (Junkunlo *et al.* (2011).

Neste trabalho, algumas lectinas com ação contra o WSSV que foram encontradas no músculo já haviam sido reportadas em outros tecidos de camarões peneídeos. Wang *et al.* (2009), observou a lectina tipo C isoforma 3 induzida apenas no hepatopâncreas de *F. chinensis*. Por sua vez, a lectina tipo C de *F. chinensis* foi observada primeiramente apenas em hemócitos (Liu *et al.*, 2007) e a proteína com domínio de lectina tipo C de *L. vannamei* apenas no intestino (Junkunlo *et al.* 2011). Aqui todas essas proteínas se mostraram induzidas no tecido muscular e a lectina tipo C também se mostrou induzida nos animais saudáveis, reforçando o metabolismo mais ativo dos camarões não infectados em comparação aos doentes e o papel das lectinas no combate a Mancha Branca em vários tecidos de *L. vannamei*.

Outra lectina do tipo C encontrada aqui, foi a proteína antiviral dos camarões ou *Shrimp Antiviral Protein* (PmAV). A PmAV atua neutralizando as estruturas moleculares virais na invasão das células alvo, impedindo que os vírus, como o da WSSV, utilizem a maquinaria das células hospedeiras para se replicar (Wang *et al.*, 2005). No nosso estudo, a PmAV se mostrou induzida no músculo e em animais saudáveis (Tabelas 1 e 3). Estes achados confirmam que algumas das

proteínas envolvidas na resposta imune dos crustáceos são expressas também no tecido muscular e reforçam o papel relevante dessa proteína no combate ao WSSV.

4.4. Outras proteínas relacionadas à imunidade

Proteínas de estresse térmico ou *Heat Shock Proteins* (HSPs) se mostraram induzidas no hepatopâncreas (Tabela 1) e em animais saudáveis (Tabela 3), estando esta condição possivelmente relacionada ao estresse ocasionado, por exemplo, pela alta densidade populacional e contato com o WSSV. As HSPs são chaperonas altamente conservadas que contribuem na detecção de proteínas não funcionais, além da manutenção da conformação e integridade proteicas (Feder e Hofmann, 1999; Li *et al.*, 2012). Embora as proteínas de estresse térmico nem sempre estejam envolvidas em respostas imunes, elas são bons indicadores da homeostase e adaptação do organismo, alterando seus níveis perante diferentes perturbações (Baringou *et al.*, 2016).

A proteína p53 foi encontrada aqui induzida no tecido muscular e está relacionada à detecção de estresse e consequente morte celular, com papel ainda pouco elucidado nos invertebrados. O gene da P53 é induzido em células sob estresse oxidativo severo, incluindo aquelas com infecção viral, e pode desencadear a morte celular, podendo caracterizar uma forma de eliminação do vírus (Qian *et al.*, 2014). Em ambas as condições do nosso experimento, tanto os camarões doentes quanto os saudáveis entraram em contato com o vírus da Mancha Branca, sendo, portanto, a morte celular uma possível alternativa encontrada contra a proliferação de um potencial patógeno (Rao *et al.*, 2015).

Dentre as proteínas responsáveis pela imunidade dos crustáceos através da hemolinfa foram encontradas aqui as coagulante de hemolinfa e a enzima pró-coagulante, induzidas no músculo e em animais saudáveis (Tabelas 1 e 3), além do fator B de coagulação, induzido em animais saudáveis (Tabela 3) apenas. O fator B de coagulação e a enzima pró-coagulante, aqui

fortemente induzida em animais saudáveis ($\log_2FC = -11.08$, $p = 0.002$), participam da imunidade das respostas imunes nos crustáceos através da coagulação da hemolinfa em contato com o patógeno e a consequente interrupção na proliferação do micro organismo (Wang *et al.*, 2001; Cheng *et al.*, 2008). A coagulação da hemolinfa é um processo fundamental nos organismos que possuem um sistema circulatório aberto, como os crustáceos, atuando na prevenção da perda de fluido corporal e invasão por micro organismos oportunistas (Yeh *et al.*, 1998). Esse resultado reforça a importância de proteínas que atuem no combate ao WSSV através da hemolinfa e sua relação com a parada ou diminuição da proliferação viral em peneídeos. Nos animais não infectados pela Mancha Branca, existe uma maior resposta metabólica no combate ao vírus, enquanto nos doentes, com o passar do tempo e o aumento da debilidade, a expressão de genes imunes tende a diminuir. Estes resultados foram observados no presente trabalho, indo de encontro a estudos anteriores, nos quais foi realizado desafio viral envolvendo o vírus da Síndrome de Taura em *L. vannamei* (Sookruksawong *et al.*, 2013; Zeng *et al.*, 2013) e observando-se um metabolismo menos ativo por parte dos animais doentes.

4.5. Conclusões sobre as comparações entre tecidos e tratamentos

Em relação ao perfil de expressão dos genes DE da comparação entre os tecidos realizada, existe um perfil de expressão distinto entre as amostras de hepatopâncreas e músculo, com os tecidos podendo ser observados em diferentes grupos (Figura 3). Ao se considerar a condição 'doente' e 'saudável' as amostras dos dois tecidos continuam sendo agrupadas em clados diferentes. Porém, as amostras de animais doentes e saudáveis apresentam-se separadas em grupos distintos apenas no hepatopâncreas (Figura 5). Perante os dados apresentados, a comparação entre os tecidos permitiu concluir o papel determinante do hepatopâncreas na imunidade dos crustáceos quando comparado ao tecido muscular. Além disso, a comparação entre

doentes e saudáveis levando-se em conta o tecido, mostrou de forma complementar, que os perfis de expressão dos conjuntos de genes induzidos no hepatopâncreas de animais doentes e saudáveis são diferentes entre si. Além da participação fundamental do hepatopâncreas para a imunidade dos decápodes, esse órgão se comporta de maneira distinta em animais doentes e saudáveis, ressaltando um comportamento diferencial. Diversas proteínas de imunidade produzidas pelos hemócitos migraram através da hemolinfa para o hepatopâncreas e lá estão sendo expressas, de acordo com o que pode ser observado nos resultados gerados aqui. Foram observados 191 genes DE *up-regulated* em comum para hepatopâncreas e animais doentes, enquanto para hepatopâncreas e animais saudáveis existem apenas 31 (Figura 6). Os animais utilizados neste trabalho estavam expostos ao estresse causado pelo contato com um patógeno letal, como o WSSV, demandando um metabolismo muito mais participante do hepatopâncreas na expressão proteica relacionada à defesa (Ye *et al.*, 2015; Fan *et al.* 2016; Rao *et al.*, 2016). Os genes do hepatopâncreas mostram-se fortemente induzidos pelo fato de ali serem expressas uma enorme variedade de proteínas relacionadas à imunidade nos crustáceos (Song *et al.*, 2010; Guo *et al.*, 2013; Rao *et al.*, 2015; Fan *et al.* 2016).

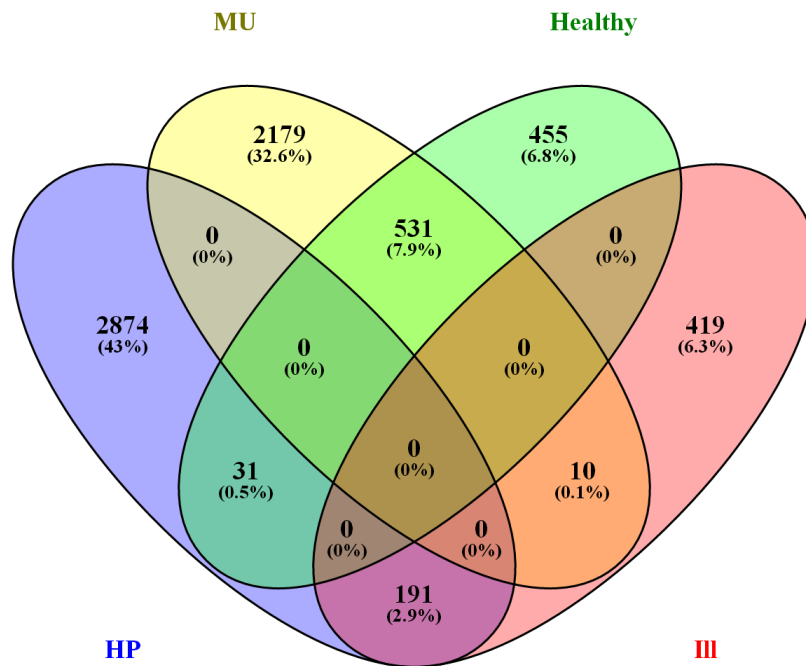


Figura 6: União dos genes DE *up-regulated* nos tecidos hepatopâncreas (HP) e músculo (MU) e nos tratamentos saudável (*Healthy*) e doente (*Ill*). Nota-se 531 genes DE comuns entre músculo e saudáveis, 191 entre hepatopâncreas e doentes, 31 entre hepatopâncreas e saudáveis e 10 entre músculo e doentes, com suas respectivas porcentagens.

4.6. Enriquecimento do GO

Na comparação entre os tecidos, o hepatopâncreas se destaca como um órgão com uma grande quantidade de proteínas imunes sendo expressas em comparação ao músculo nos crustáceos. Os principais termos enriquecidos encontrados no tecido muscular estão relacionados à organização de filamentos constituintes do músculo e contração. Por outro lado, dentre os principais termos enriquecidos encontrados no hepatopâncreas todos estão relacionados à resposta imune ou estresse. Os principais processos biológicos encontrados foram a regulação da resposta imune e da imunidade adaptativa (Tabela 2). Os lisossomos apareceram como um dos principais locais de ação na célula. São nessas organelas que acontecem eventos como a

endocitose, o que pode caracterizar a digestão de patógenos, compostos tóxicos ou sem utilidade para a célula (Liu *et al.*, 2014). Adicionalmente, diversas proteínas relacionadas à aglutinação da hemolinfa e fagocitose foram identificadas e induzidas no hepatopâncreas de *L. vannamei* conforme já descrito, reforçando a importância deste órgão no combate ao patógeno. Também foi encontrado como CC o capsídeo viral, sugerindo tratar-se de RNA viral, como o do WSSV, no qual o envelope (capsídeo) é a principal estrutura viral responsável pela proliferação do patógeno e desencadeamento de respostas imunes no hospedeiro (Kulkarni *et al.*, 2014). A atividade de glutathione transferase é característica de um organismo sob severo estresse, sendo esta uma das principais enzimas atuantes na captura de radicais de oxigênio e na proteção celular (Qian *et al.*, 2014).

Dentre os termos enriquecidos para os animais saudáveis são encontrados alguns relacionados a proteínas musculares, como actina e miosina (Tabela 4). No entanto, também são observados termos envolvidos em processos que ocorrem no citoesqueleto, podendo estar relacionados a algum processo de defesa, como a fagocitose (Wongpanya *et al.*, 2007). Por sua vez, os animais doentes apresentam termos considerados enriquecidos relacionados à síntese de proteínas. Essa ênfase dada à produção de proteínas pode estar relacionada a moléculas sinalizadoras de estresse, devido ao contágio com o vírus em uma tentativa de desencadear algumas respostas imunes para controlar a propagação viral. Nota-se também que a mitocôndria é um dos principais locais nos quais alguns processos ocorrem, ilustrando um cenário de estresse oxidativo e produção de radicais livres de oxigênio, provavelmente causado pela exposição ao WSSV (Tu *et al.*, 2010; Rao *et al.*, 2015).

Considerações finais

No presente trabalho foi realizada a análise de genes diferencialmente expressos de hepatopâncreas e músculo de indivíduos do camarão marinho *L. vannamei* expostos ao vírus da Mancha Branca, uma doença letal para os peneídeos. Foi encontrada uma grande variedade de proteínas responsáveis pelas respostas imunes nos crustáceos, como hemocianinas, crustacianinas, lectinas, crustinas, PmAV, coagulantes de hemolinfa, sendo todas igualmente importantes em um grupo dependente de um sistema imune inato, como os camarões. Foi possível concluir que a diferença entre os padrões de expressão dos dois tecidos analisados foi clara. Além disso, também foi possível a conclusão de que o comportamento do hepatopâncreas em animais doentes e saudáveis é diferente, sendo um dos possíveis fatores relacionados a maior indução de proteínas imunes nos animais saudáveis. Esse trabalho contribuiu para um maior entendimento da fisiologia dos peneídeos durante o contato com o WSSV, ilustrando alguns dos genes diferencialmente expressos em *L. vannamei*, as proteínas envolvidas, as respostas imunes desencadeadas na batalha contra o patógeno e a atuação diferencial do hepatopâncreas perante a infecção, embora mais estudos ainda sejam necessários para elucidar o mecanismo de ação em detalhes de algumas das proteínas contra a Mancha Branca.

Agradecimentos

Os autores agradecem a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) (Processo número 2012/17322-8), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro fornecido, além do Centro de Genômica Funcional do Laboratório de Biotecnologia Animal e ao Prof. Dr. Luiz Lehman Coutinho (ESALQ-USP), Piracicaba-SP, Brasil, pelo serviço de sequenciamento e uso do servidor para processamento dos dados.

Material Suplementar

Tabela S1: Total de fragmentos originados do sequenciamento Illumina, após a filtragem no Seqclean e após o mapeamento das 16 bibliotecas utilizadas. **(Pág. 156)**

Tabela S2: Genes DE significativos (FDR < 0.05) para a comparação entre hepatopâncreas e músculo. São mostrados os valores de log₂FC e seus respectivos p-values e p-values ajustados (FDR), além do log dos valores dos dados das contagens transformados por milhão (logCPM). **(Disponível no arquivo anexado)**

Tabela S3: Genes DE significativos (FDR < 0.05) para a comparação entre os animais doentes e saudáveis. São mostrados os valores de log₂FC e seus respectivos p-values ajustados (padj) e valores de base Mean. **(Disponível no arquivo anexado).**

Referências

- Alday-Sanz, V., Roque, A., & Turnbull, J. F. (2002). Clearing mechanisms of *Vibrio vulnificus* biotype I in the black tiger shrimp *Penaeus monodon*. *Diseases of aquatic organisms*, 48(2), 91-99.
- Anders, S., & Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biology*, 11(10), R106. <http://doi.org/10.1186/gb-2010-11-10-r106>
- Antony, S. P., Singh, I. S. B., Sudheer, N. S., Vrinda, S., Priyaja, P., & Philip, R. (2011). Molecular characterization of a crustin-like antimicrobial peptide in the giant tiger shrimp, *Penaeus monodon*, and its expression profile in response to various immunostimulants and challenge with WSSV. *Immunobiology*, 216(1-2), 184-194. <http://doi.org/10.1016/j.imbio.2010.05.030>
- Ashburner M., Ball C. A., Blake J.A., Botstein D., Butler H., Cherry J.M., Davis A.P., Dolinski K., Dwight S.S., Eppig J.T., Harris M.A, Hill D.P., Issel-Tarver L., Kasarskis A., Lewis S., M. J. C. (2000). Gene ontology: Tool for the identification of biology. *Natural Genetics*, 25(may), 25-29.
- Baringou, S., Rouault, J.-D., Koken, M., Hardivillier, Y., Hurtado, L., & Leignel, V. (2016). Diversity of cytosolic HSP70 Heat shock protein from decapods and their phylogenetic placement within Arthropoda. *Gene*. <http://doi.org/10.1016/j.gene.2016.06.061>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, 289-300.
- Brouwer, Marius, *et al.* "Changes in mitochondrial gene and protein expression in grass shrimp, *Palaemonetes pugio*, exposed to chronic hypoxia." *Marine environmental research* 66.1 (2008): 143-145.
- Burmester, T. (2001). Molecular Evolution of the Arthropod Hemocyanin Superfamily, 184-195.
- Cheng, W., Tsai, I. H., Huang, C. J., Chiang, P. C., Cheng, C. H., & Yeh, M. S. (2008). Cloning and characterization of hemolymph clottable proteins of kuruma prawn (*Marsupenaeus japonicus*) and white shrimp (*Litopenaeus vannamei*). *Developmental and Comparative Immunology*, 32(3), 265-274. <http://doi.org/10.1016/j.dci.2007.05.009>
- Cuthbertson, B. J., Büllsbach, E. E., Fievet, J., Bachère, E., & Gross, P. S. (2004). A new class (penaeidin class 4) of antimicrobial peptides from the Atlantic white shrimp (*Litopenaeus setiferus*) exhibits target specificity and an

- independent proline-rich-domain function. *The Biochemical Journal*, 381, 79–86.
<http://doi.org/10.1042/BJ20040330>
- de la Vega, E., O'Leary, N. A., Shockey, J. E., Robalino, J., Payne, C., Browdy, C. L., ... Gross, P. S. (2008). Anti-lipopolysaccharide factor in *Litopenaeus vannamei* (LvALF): A broad spectrum antimicrobial peptide essential for shrimp immunity against bacterial and fungal infection. *Molecular Immunology*, 45(7), 1916–1925.
<http://doi.org/10.1016/j.molimm.2007.10.039>
- Destoumieux, D., Muñoz, M., Cosseau, C., Rodriguez, J., Bulet, P., Comps, M., & Bachère, E. (2000). Penaeidins, antimicrobial peptides with chitin-binding activity, are produced and stored in shrimp granulocytes and released after microbial challenge. *Journal of Cell Science*, 113 (Pt 3, 461–469.
<http://doi.org/10.1016/j.bbagen.2015.12.010>
- Du, Z.-Q., & Jin, Y.-H. (2015). Molecular Characterization and Antibacterial Activity Analysis of Two Novel Penaeidin Isoforms from Pacific White Shrimp, *Litopenaeus vannamei*. *Applied Biochemistry and Biotechnology*, 177(8), 1607–1620. <http://doi.org/10.1007/s12010-015-1840-7>
- Fan, L., Wang, A., Miao, Y., Liao, S., Ye, C., & Lin, Q. (2016). Comparative proteomic identification of the hepatopancreas response to cold stress in white shrimp, *Litopenaeus vannamei*. *Aquaculture*, 454, 27–34.
<http://doi.org/10.1016/j.aquaculture.2015.10.016>
- Feder, M.E. (1999). Heat-shock proteins , molecular chaperones , and the stress response : *Annu. Rev. Physiol.* 61:243–82.
- Food and Agriculture Organization of the United Nations (FAO). Threats and risks of introducing alien shrimp species. Disponível em: <http://www.fao.org/docrep/009/a0086e/a0086e10.htm>. Acessado em 10 de setembro de 2016.
- Gajula M. P., Soni G., Babu G., Rai A., B. N. (2013). Journal of Applied Molecular Interaction Studies of Shrimp Antiviral Protein , PmAV with WSSV RING Finger. *Journal of Applied Bioinformatics & Computational Biology*, (2:1), 1–5.
- Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., ... & Hornik, K. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome biology*, 5(10), 1.
- Grabherr M., Brian J. Haas, Moran Yassour, Joshua Z. Levin, Dawn A. Thompson, Ido Amit, Xian Adiconis, Lin Fan, Raktima Raychowdhury, Qiandong Zeng, Zehua Chen, Evan Mauceli, Nir Hacohen, Andreas Gnirke, Nicholas Rhind, Federica di Palma, Bruce W., N., & Friedman, and A. R. (2013). Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature Biotechnology*, 29(7), 644–652.
<http://doi.org/10.1038/nbt.1883>.Trinity
- Guo, H., Xian, J.-A., Li, B., Ye, C.-X., Wang, A.-L., Miao, Y.-T., & Liao, S.-A. (2013). Gene expression of apoptosis-related genes, stress protein and antioxidant enzymes in hemocytes of white shrimp *Litopenaeus vannamei* under nitrite stress. *Comparative Biochemistry and Physiology. Toxicology & Pharmacology : CBP*, 157(4), 366–71.
<http://doi.org/10.1016/j.cbpc.2013.03.001>
- Homvise, T., Tassanakajon, A., & Somboonwivat, K. (2010). Penaeus monodon SERPIN, PmSERPIN6, is implicated in the shrimp innate immunity. *Fish and Shellfish Immunology*, 29(5), 890–898.
<http://doi.org/10.1016/j.fsi.2010.07.017>
- Huang, Y., Ma, F., Wang, W., & Ren, Q. (2015). Identification and molecular characterization of a peritrophin-like gene, involved in the antibacterial response in Chinese mitten crab, *Eriocheir sinensis*. *Developmental and*

- Comparative Immunology*, 50(2), 129–138. <http://doi.org/10.1016/j.dci.2015.01.002>
- Imjongjirak, C., Amparyup, P., & Tassanakajon, A. (2011). Molecular cloning, genomic organization and antibacterial activity of a second isoform of antilipopopolysaccharide factor (ALF) from the mud crab, *Scylla paramamosain*. *Fish and Shellfish Immunology*, 30(1), 58–66. <http://doi.org/10.1016/j.fsi.2010.09.011>
- Junkunlo, K., Prachumwat, A., Tangprasittipap, A., Senapin, S., Bavornphinyo, S., Flegel, T.W., Sritunyalucksana, K., 2011. A novel lectin domain-containing protein (LvCTLD) associated with response of the whiteleg shrimp *Penaeus (Litopenaeus) vannamei* to yellow head virus (YHV). *Dev. Comp. Immunol* 37, 334–341.
- Kulkarni, A. D., Kiron, V., Rombout, J. H. W. M., Brinchmann, M. F., Fernandes, J. M. O., Sudheer, N. S., & Singh, B. I. S. (2014). Protein profiling in the gut of *Penaeus monodon* gavaged with oral WSSV-vaccines and live white spot syndrome virus. *Proteomics*, 14(13–14), 1660–1673. <http://doi.org/10.1002/pmic.201300405>
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods*, 9(4), 357–359. <http://doi.org/10.1038/nmeth.1923>
- Leu, J. H., Lin, S. J., Huang, J. Y., Chen, T. C., & Lo, C. F. (2013). A model for apoptotic interaction between white spot syndrome virus and shrimp. *Fish and Shellfish Immunology*, 34(4), 1011–1017. <http://doi.org/10.1016/j.fsi.2012.05.030>
- Li, C., Weng, S., Chen, Y., Yu, X., Lü, L., Zhang, H., ... Xu, X. (2012). Analysis of *Litopenaeus vannamei* transcriptome using the next-generation DNA sequencing technique. *PLoS One*, 7(10), e47442. <http://doi.org/10.1371/journal.pone.0047442>
- Li, X. C., Wang, X. W., Wang, Z. H., Zhao, X. F., & Wang, J. X. (2009). A three-domain Kazal-type serine proteinase inhibitor exhibiting domain inhibitory and bacteriostatic activities from freshwater crayfish *Procambarus clarkii*. *Developmental and Comparative Immunology*, 33(12), 1229–1238. <http://doi.org/10.1016/j.dci.2009.07.001>
- Liu, Y.C., Li, F.H., Dong, B., Wang, B., Luan, W., Zhang, X.J., Zhang, L.S., Xiang, J.H., (2007). Molecular cloning, characterization and expression analysis of a putative C-type lectin (Fclectin) gene in Chinese shrimp *Fenneropenaeus chinensis*. *Mol. Immunol.* 44, 598–607.
- Liu, Y., Hou, F., He, S., Qian, Z., Wang, X., Mao, A., Liu, X. (2014). Identification, characterization and functional analysis of a serine protease inhibitor (Lvserpin) from the Pacific white shrimp, *Litopenaeus vannamei*. *Developmental and Comparative Immunology*, 43(1), 35–46. <http://doi.org/10.1016/j.dci.2013.10.012>
- Love, M. I., Anders, S., Kim, V., & Huber, W. (2015). RNA-Seq workflow: gene-level exploratory analysis and differential expression. *F1000Research*, 4, 1070. <http://doi.org/10.12688/f1000research.7035.1>
- Lv, J., Liu, P., Gao, B., Wang, Y., Wang, Z., Chen, P., & Li, J. (2014). Transcriptome analysis of the *Portunus trituberculatus*: De novo assembly, growth-related gene identification and marker discovery. *PLoS ONE*, 9(4). <http://doi.org/10.1371/journal.pone.0094055>
- Nelder, J. A, Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society. Series A (General)*, 135(3), 370-384.
- Panrat, T., Sinthujaroen, P., Nupan, B., Wanna, W., Tammi, M. T., & Phongdara, A. (2012). Characterization of a novel binding protein for Fortilin/TCTP - component of a defense mechanism against viral infection in *Penaeus monodon*. *PLoS ONE*, 7(3). <http://doi.org/10.1371/journal.pone.0033291>

- Pathan, M., Gireesh-Babu, P., Pavan-Kumar, A., Jeena, K., Sharma, R., Makesh, M., ... Krishna, G. (2013). In vivo therapeutic efficacy of recombinant *Penaeus monodon* antiviral protein (rPmAV) administered in three different forms to WSSV infected *Penaeus monodon*. *Aquaculture*, 376–379, 64–67. <http://doi.org/10.1016/j.aquaculture.2012.10.034>
- Perez-Henriquez, R., Hernández-Martínez, F., & Cruz, P. (2009). Genetic diversity status of White shrimp *Penaeus (Litopenaeus) vannamei* broodstock in Mexico. *Aquaculture*, 297(1–4), 44–50. <http://doi.org/10.1016/j.aquaculture.2009.08.038>
- Powell, D., Knibb, W., Remilton, C., & Elizur, A. (2015). De-novo transcriptome analysis of the banana shrimp (*Fenneropenaeus merguensis*) and identification of genes associated with reproduction and development. *Marine Genomics*, 22, 71–78. <http://doi.org/10.1016/j.margen.2015.04.006>
- Qian, Z., Liu, T., Liu, Q., He, S., Liu, Y., Hou, F., ... Liu, X. (2014). P53 is involved in shrimp survival via its regulation roles on MnSOD and GPx in response to acute environmental stresses. *Comparative Biochemistry and Physiology - C Toxicology and Pharmacology*, 159(1), 38–51. <http://doi.org/10.1016/j.cbpc.2013.09.009>
- Rao, R., Bing Zhu, Y., Alinejad, T., Tiruvayipati, S., Lin Thong, K., Wang, J., & Bhasu, S. (2015). RNA-seq analysis of *Macrobrachium rosenbergii* hepatopancreas in response to *Vibrio parahaemolyticus* infection. *Gut Pathogens*, 7(April 2016), 6. <http://doi.org/10.1186/s13099-015-0052-6>
- Rattanaporn, O., & Utarabhand, P. (2011). Molecular cloning of a C-type lectin with two CRD domains from the banana shrimp *Fenneropenaeus merguensis*: Early gene up-regulation after *Vibrio harveyi* infection. *Journal of Invertebrate Pathology*, 106(2), 196–204. <http://doi.org/10.1016/j.jip.2010.09.002>
- Rewitz, K. F., Styriahave, B., L?bner-Olesen, A., & Andersen, O. (2006). Marine invertebrate cytochrome P450: Emerging insights from vertebrate and insect analogies. *Comparative Biochemistry and Physiology - C Toxicology and Pharmacology*, 143(4), 363–381. <http://doi.org/10.1016/j.cbpc.2006.04.001>
- Robinson, M. D., & Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology*, 11(3), R25. <http://doi.org/10.1186/gb-2010-11-3-r25>
- Rolland, J. L., Abdelouahab, M., Dupont, J., Lefevre, F., Bachère, E., & Romestand, B. (2010). Stylicins, a new family of antimicrobial peptides from the Pacific blue shrimp *Litopenaeus stylirostris*. *Molecular Immunology*, 47(6), 1269–1277. <http://doi.org/10.1016/j.molimm.2009.12.007>
- Shockey, J. E., O'Leary, N. A., de la Vega, E., Browdy, C. L., Baatz, J. E., & Gross, P. S. (2009). The role of crustins in *Litopenaeus vannamei* in response to infection with shrimp pathogens: An in vivo approach. *Developmental and Comparative Immunology*, 33(5), 668–673. <http://doi.org/10.1016/j.dci.2008.11.010>
- Sinthujaroen, P., Tonganunt-Srithaworn, M., Eurwilaichitr, L., & Phongdara, A. (2015). Protection of *Litopenaeus vannamei* against the white spot syndrome virus using recombinant Pm-fortilin expressed in *Pichia pastoris*. *Aquaculture*, 435, 450–457. <http://doi.org/10.1016/j.aquaculture.2014.10.024>
- Somnuk, S., Tassanakajon, A., & Rimphanitchayakit, V. (2012). Gene expression and characterization of a serine proteinase inhibitor PmSERPIN8 from the black tiger shrimp *Penaeus monodon*. *Fish & Shellfish Immunology*, 33(2), 332–341. <http://doi.org/10.1016/j.fsi.2012.05.005>
- Sookruksawong, S., Sun, F., Liu, Z., & Tassanakajon, A. (2013). RNA-Seq analysis reveals genes associated with resistance to Taura syndrome Virus (TSV) in the Pacific white shrimp *Litopenaeus vannamei*. *Developmental*

- and Comparative Immunology*, 41(4), 523–33. <http://doi.org/10.1016/j.dci.2013.07.020>
- Soowannayan, Chumporn, *et al.* "Australian red claw crayfish (*Cherax quadricarinatus*) is susceptible to yellow head virus (YHV) infection and can transmit it to the black tiger shrimp (*Penaeus monodon*)." *Aquaculture* 445 (2015): 63-69.
- Song, K. K., Li, D. F., Zhang, M. C., Yang, H. J., Ruan, L. W., & Xu, X. (2010). Cloning and characterization of three novel WSSV recognizing lectins from shrimp *Marsupenaeus japonicus*. *Fish and Shellfish Immunology*, 28(4), 596–603. <http://doi.org/10.1016/j.fsi.2009.12.015>
- Tassanakajon, A., Somboonwiwat, K., & Amparyup, P. (2014). Sequence diversity and evolution of antimicrobial peptides in invertebrates. *Developmental and Comparative Immunology*, 48(2), 324–341. <http://doi.org/10.1016/j.dci.2014.05.020>
- Tassanakajon, A., Somboonwiwat, K., Supungul, P., & Tang, S. (2013). Discovery of immune molecules and their crucial functions in shrimp immunity. *Fish and Shellfish Immunology*, 34(4), 954–967. <http://doi.org/10.1016/j.fsi.2012.09.021>
- Tu, H. T., Silvestre, F., Phuong, N. T., & Kestemont, P. (2010). Effects of pesticides and antibiotics on penaeid shrimp with special emphases on behavioral and biomarker responses. *Environmental Toxicology and Chemistry / SETAC*, 29(4), 929–38. <http://doi.org/10.1002/etc.99>
- Vatanavicharn, T., Supungul, P., Puanglarp, N., Yingvilasprasert, W., & Tassanakajon, A. (2009). Genomic structure, expression pattern and functional characterization of crustinPm5, a unique isoform of crustin from *Penaeus monodon*. *Comparative Biochemistry and Physiology - B Biochemistry and Molecular Biology*, 153(3), 244–252. <http://doi.org/10.1016/j.cbpb.2009.03.004>
- Wade, N. M., Tollenaere, A., Hall, M. R., & Degnan, B. M. (2009). Evolution of a novel carotenoid-binding protein responsible for crustacean shell color. *Molecular biology and evolution*, 26(8), 1851-1864.
- Wallenius, K.T. (1963). Biased sampling: the non-central hypergeometric probability distribution. PhD: Stanford University
- Wang, R., Lee, S. Y., Cerenius, L., & Söderhäll, K. (2001). Properties of the prophenoloxidase activating enzyme of the freshwater crayfish, *Pacifastacus leniusculus*. *European Journal of Biochemistry*, 268(4), 895–902. <http://doi.org/10.1046/j.1432-1327.2001.01945.x>
- Wang, Z., Chua, H. K., Gusti, A. a R. a, He, F., Fenner, B., Manopo, I., Kwang, J. (2005). RING-H2 protein WSSV249 from white spot syndrome virus sequesters a shrimp ubiquitin-conjugating enzyme, PvUbc, for viral pathogenesis. *Journal of Virology*, 79(14), 8764–8772. <http://doi.org/10.1128/JVI.79.14.8764-8772.2005>
- Wang, X. W., Xu, W. T., Zhang, X. W., Zhao, X. F., Yu, X. Q., & Wang, J. X. (2009). A C-type lectin is involved in the innate immune response of Chinese white shrimp. *Fish and Shellfish Immunology*, 27(4), 556–562. <http://doi.org/10.1016/j.fsi.2009.07.011>
- Wang, X. W., & Wang, J. X. (2013). Pattern recognition receptors acting in innate immune system of shrimp against
- Ye, Ting, *et al.* "Ferritin protect shrimp *Litopenaeus vannamei* from WSSV infection by inhibiting virus replication." *Fish & shellfish immunology* 42.1 (2015): 138-143.
- Yeh, M. S., Chen, Y. L., & Tsai, I. H. (1998). The hemolymph clottable proteins of tiger shrimp, *Penaeus monodon*, and related species. *Comparative Biochemistry and Physiology - B Biochemistry and Molecular Biology*, 121(2),

169–176. [http://doi.org/10.1016/S0305-0491\(98\)10085-8](http://doi.org/10.1016/S0305-0491(98)10085-8)

- Young, M. D., Wakefield, M. J., Smyth, G. K., & Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biology*, *11*(2), R14. <http://doi.org/10.1186/gb-2010-11-2-r14>
- Zhao, Y. R., Xu, Y. H., Jiang, H. S., Xu, S., Zhao, X. F., & Wang, J. X. (2014). Antibacterial activity of serine protease inhibitor 1 from kuruma shrimp *Marsupenaeus japonicus*. *Developmental and Comparative Immunology*, *44*(2), 261–269. <http://doi.org/10.1016/j.dci.2014.01.002>
- Zeng, D., Chen, X., Xie, D., Zhao, Y., Yang, C., Li, Y., ... Chen, X. (2013). Transcriptome analysis of Pacific white shrimp (*Litopenaeus vannamei*) hepatopancreas in response to Taura syndrome Virus (TSV) experimental infection. *PloS One*, *8*(2), e57515. <http://doi.org/10.1371/journal.pone.0057515>
- Zheng, L., Zhao, X., Zhang, P., Chen, C., Liu, S., Huang, R., ... Zhang, Y. (2016). Hemocyanin from Shrimp *Litopenaeus vannamei* Has Antiproliferative Effect against HeLa Cell In Vitro. *Plos One*, *11*, e0151801. <http://doi.org/10.1371/journal.pone.0151801>

5. Conclusões

- Os resultados obtidos através das análises de mineração e validação das ESTs demonstraram que os locos ESTs contendo SNPs e SSRs minerados inseridos na cadeia principal de proteínas podem ser polimórficos, uma vez que os genes 60S ribossomal e crustacianina apresentaram-se polimórficos. Dessa forma, a Hipótese I deve ser aceita.
- Polimorfismos presentes nos genes codificadores das proteínas 60S ribossomal e crustacianina não demonstraram associação com características de desempenho nos animais avaliados para altas taxas de crescimento e sobrevivência na amostragem utilizada, rejeitando-se, portanto, a **Hipótese II**.
- Existe diferença no perfil de expressão entre os genes de indivíduos de *L. vannamei* de famílias apresentando baixo e alto crescimento, sendo possível identificar alguns genes induzidos exclusivamente nos animais maiores, os quais se constituem em potenciais genes candidatos para ganho de peso. Portanto, a **Hipótese III** pode ser aceita.
- Foi possível identificar genes diferencialmente expressos relacionados à resposta imune no hepatopâncreas e músculo de indivíduos de *L. vannamei* expostos a Mancha Branca. Também foi observada diferença nos padrões de expressão entre os genes dos animais não infectados (saudáveis) e infectados (doentes). Portanto, a **Hipótese IV** pode ser aceita.

6. Referências citadas na Introdução

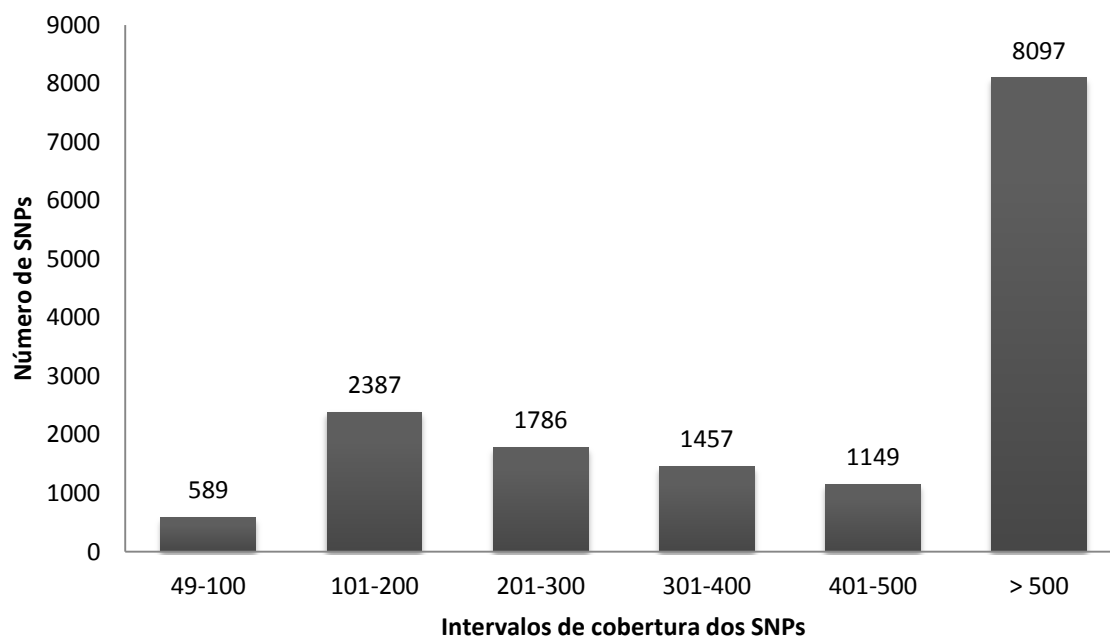
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., & Harris, M. A. (2000). Gene Ontology: tool for the unification of biology. *Nature genetics*, 25(1), 25-29.
- Baldwin J. D., Bass A. L., Bowen B. W., Clark W. H. (1998). Molecular phylogeny and biogeography of the marine shrimp *Penaeus*. *Molecular phylogenetics and evolution*, 10(3), 399-407. doi: 10.1006/mpev.1998.0537.

- Chen, K., Li, E., Li, T., Xu, C., Wang, X., Lin, H., ... Chen, L. (2015). Transcriptome and molecular pathway analysis of the hepatopancreas in the Pacific white shrimp *Litopenaeus vannamei* under chronic low-salinity stress. *PLoS ONE*, *10*(7), 1–22. <http://doi.org/10.1371/journal.pone.0131503>
- Jin, S., Fu, H., Zhou, Q., Sun, S., Jiang, S., Xiong, Y., ... Zhang, W. (2013). Transcriptome Analysis of Androgenic Gland for Discovery of Novel Genes from the Oriental River Prawn, *Macrobrachium nipponense*, Using Illumina Hiseq 2000. *PLoS ONE*, *8*(10). <http://doi.org/10.1371/journal.pone.0076840>
- Jung, H., Lyons, R. E., Hurwood, D. a., & Mather, P. B. (2013). Genes and growth performance in crustacean species: A review of relevant genomic studies in crustaceans and other taxa. *Reviews in Aquaculture*, *5*(2), 77–110. <http://doi.org/10.1111/raq.12005>
- Lv, J., Liu, P., Gao, B., Wang, Y., Wang, Z., Chen, P., & Li, J. (2014). Transcriptome analysis of the *Portunus trituberculatus*: De novo assembly, growth-related gene identification and marker discovery. *PLoS ONE*, *9*(4). <http://doi.org/10.1371/journal.pone.0094055>
- Ma, K. Y., Chan, T.-Y., & Chu, K. H. (2009). Phylogeny of penaeoid shrimps (Decapoda: Penaeoidea) inferred from nuclear protein-coding genes. *Molecular Phylogenetics and Evolution*, *53*(1), 45–55. <http://doi.org/10.1016/j.ympev.2009.05.019>
- McCarthy, D. J., Chen, Y., & Smyth, G. K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research*, *40*(10), 4288–4297. <http://doi.org/10.1093/nar/gks042>
- Oshlack, A., Robinson, M. D., & Young, M. D. (2010). From RNA-seq reads to differential expression results, 1–10.
- Perez, F., Ortiz, J., Zhinaula, M., Gonzabay, C., Caldero, J., Pe, F., & Volckaert, F. A. M. J. (2005). Development of EST-SSR Markers by Data Mining in Three Species of Shrimp : *Litopenaeus vannamei* , *Litopenaeus stylirostris* , and *Trachypenaeus birdy*, *7*, 554–569. <http://doi.org/10.1007/s10126-004-5099-1>
- Perez-Henriquez, R., Hernández-martínez, F., & Cruz, P. (2009). Genetic diversity status of White shrimp *Penaeus Litopenaeus*) *vannamei* broodstock in Mexico. *Aquaculture*, *297*(1–4), 44–50. <http://doi.org/10.1016/j.aquaculture.2009.08.038>
- UniProt Consortium. (2014). UniProt: a hub for protein information. *Nucleic acids research*, gku989.
- Visscher, P. M., Hill, W. G., & Wray, N. R. (2008). Heritability in the genomics era — concepts and misconceptions. *Nature Reviews Genetics*, *9*(4), 255–266. <http://doi.org/10.1038/nrg2322>
- Zhang, L., Li, L., Zhu, Y., Zhang, G., & Guo, X. (2014). Transcriptome analysis reveals a rich gene set related to innate immunity in the Eastern oyster (*Crassostrea virginica*). *Marine Biotechnology (New York, N.Y.)*, *16*(1), 17–33. <http://doi.org/10.1007/s10126-013-9526-z>

Material Suplementar

Manuscrito 4

Figura S1



Manuscrito 5

Tabela S1

Fragmentos Illumina – Alto X Baixo Crescimento		
Músculo (48 bibliotecas)		
	Alto crescimento	Baixo crescimento
Fragmentos paired-end (desvio padrão)	1.138.673.850	1.109.682.380
Média do número de fragmentos paired-end/biblioteca (milhões)	47.4	46.2
Desvio padrão	4.734.381	5.472.493
Total de fragmentos	2.248.356.230	
Após filtragem no SeqyClean		
Paired-end remanescentes (desvio padrão)	859.886.382	837.515.610
Média do número de fragmentos paired-end/biblioteca (milhões)	35.8	34.8
Desvio Padrão	3.595.355	4.244.876
Total fragmentos (%)	1.697.401.992 (75.4%)	
Após mapeamento		
	831.930.442	796.126.113
Total fragmentos (%)	1.628.056.555 (96%)	

Tabela S2

Loco	logFC	logCPM	FDR
Alto Crescimento			
Locus_12483.0_hypothetical_protein_X777_10321_[Cerapachys_biroi]	-14.3615	7.2904	0.0000
Locus_26459.0_membrane_glycoprotein_LIG-1_putative_[Ixodes_scapularis]	-13.4736	4.3451	0.0000
Locus_33175.0_hypothetical_protein_L798_13875_[Zootermopsis_nevadensis]	-12.8390	6.3004	0.0000
Locus_2465.0_PREDICTED_nucleobindin-1-like_[Metaseiulus_occidentalis]	-12.7379	3.6113	0.0000
Locus_2428.0_nicotinic_acetylcholine_receptor_subunit_alpha_10_[Pandalopsis_japonica]	-12.6400	3.5143	0.0000
Locus_2430.0_hypothetical_protein_L798_02724_[Zootermopsis_nevadensis]	-12.5588	3.4332	0.0000
Locus_5524.0_PREDICTED_uncharacterized_protein_LOC659947_isoform_X1_[Tribolium_castaneum]	-12.2928	3.1682	0.0000
Locus_23910.0 BlastHit gi 47605414 sp Q7M4F4.1 CUD1_SCHGR	-12.2685	3.1437	0.0000
Locus_9518.0_PREDICTED_uncharacterized_protein_LOC659018_[Tribolium_castaneum]	-12.2020	3.4770	0.0000
Locus_31558.0_JHE-like_carboxylesterase_1_[Pandalopsis_japonica]	-11.7558	2.6358	0.0000
Locus_2157.0_hypothetical_protein_L798_11131_partial_[Zootermopsis_nevadensis]	-11.7429	2.6219	0.0000
Locus_13786.0_adenylate_cyclase_[Danaus_plexippus]	-11.6040	2.4819	0.0000
Locus_41117.0_PREDICTED_uncharacterized_protein_LOC102805443_[Saccoglossus_kowalevskii]	-11.4931	2.3757	0.0000
Locus_21570.0_PREDICTED_uncharacterized_protein_LOC101847791_isoform_X1_[Aplysia_californica]	-11.4538	2.3377	0.0000
Locus_2173.0_topoisomerase_TOP1-interacting_protein_BTBD1_putative_[Ixodes_scapularis]	-11.4235	2.3071	0.0000
Locus_22711.0 BlastHit gi 544013 sp P36362.1 CHIT_MANSE	-11.3377	2.2219	0.0000
Locus_16948.0 BlastHit gi 113750 sp P15684.2 AMPN_RAT	-11.2660	2.1513	0.0000
Locus_3367.0_unnamed_protein_product_[Oncorhynchus_mykiss]	-11.1262	2.0154	0.0000
Locus_19780.0 BlastHit gi 341940516 sp Q9JLB4.3 CUBN_MOUSE	-10.9004	1.7908	0.0000
Locus_16512.0 BlastHit gi 75026432 sp Q9V7U0.1 RESIL_DROME	-10.8468	1.7347	0.0000
Locus_17093.0_antimicrobial_peptide_type_1_precursor_lf_[Pandalopsis_japonica]	-10.8047	1.6924	0.0000
Locus_17125.0_down_syndrome_cell_adhesion_molecule_[Pacifastacus_leniusculus]	-10.7924	1.6815	0.0000
Locus_17128.0_PREDICTED_arylsulfatase_B-like_[Aplysia_californica]	-10.7127	1.6018	0.0000
Locus_22952.1_hypothetical_protein_CRE_23800_[Caenorhabditis_remanei]	-10.7076	1.6030	0.0000
Locus_22996.0_unnamed_protein_product_[Oikopleura_dioica]	-10.5716	1.4642	0.0000
Locus_34246.0 BlastHit gi 300681185 sp B0WC46.1 TRET1_CULQU	-10.5516	1.4501	0.0000
Locus_11705.0 BlastHit gi 68566103 sp Q8VHJ4.1 TM11D_RAT	-10.4707	1.3678	0.0000
Locus_23192.0_hypothetical_protein_DAPPUDRAFT_305518_[Daphnia_pulex]	-10.2725	1.1780	0.0000
Locus_17114.0_PREDICTED_WAP_four-disulfide_core_domain_protein_2_[Trichechus_manatus_latiostris]	-10.2128	1.1213	0.0000
Locus_23314.0_structural_constituent_of_cuticle_putative_[Pediculus_humanus_corporis]	-10.1454	1.0523	0.0000
Locus_23169.0_C-type_lectin_2_[Fenneropenaeus_chinensis]	-10.1299	1.0375	0.0000
Locus_22949.0_PREDICTED_histone-lysine_N-methyltransferase_H3_lysine-79_specific-like_isoform_X9_[Apis_dorsata]	-10.0366	0.9491	0.0000

Locus_30562.0_Ran_GTPase-activating_protein_1_[Zootermopsis_nevadensis]	-10.0091	0.9131	0.0000
Locus_3996.0 BlastHit gi 544013 sp P36362.1 CHIT_MANSE	-9.9491	0.8555	0.0000
Locus_30497.0_Ninein-like_protein_[Zootermopsis_nevadensis]	-9.9210	0.8334	0.0000
Locus_28077.0 BlastHit gi 11387125 sp Q10836.1 TRHDE_RAT	-9.8852	0.7969	0.0000
Locus_16353.0_hypothetical_protein_BRAFLDRAFT_130664_[Branchiostoma_floridae]	-9.8734	0.7892	0.0000
Locus_39158.0_Selenocysteine_lyase_[Crassostrea_gigas]	-9.8709	0.7901	0.0000
Locus_9969.0 BlastHit gi 75026432 sp Q9V7U0.1 RESIL_DROME	-9.8302	0.7503	0.0000
Locus_26643.0_PREDICTED_zinc_finger_MYM-type_protein_1-like_[Loxodonta_africana]	-9.7419	0.6572	0.0000
Locus_30636.0 BlastHit gi 172044825 sp Q9ULH0.3 KDIS_HUMAN	-9.6957	0.6159	0.0000
Locus_8860.0_Sodium/hydrogen_exchanger_2_[Pteropus_alecto]	-9.6773	0.5940	0.0000
Locus_42217.0_hypothetical_protein_CAPTEDRAFT_201987_[Capitella_teleta]	-9.5919	0.5147	0.0000
Locus_4900.0 BlastHit gi 8928107 sp Q9VLK4.1 DIUX_DROME	-9.4749	0.4066	0.0000
Locus_39598.0_PREDICTED_uncharacterized_protein_LOC100207753_[Hydra_vulgaris]	-9.4547	0.3874	0.0000
Locus_23155.0_hypothetical_protein_L798_01553_[Zootermopsis_nevadensis]	-9.4370	0.3736	0.0000
Locus_33304.0 BlastHit gi 83286899 sp P69478.1 CHSTB_RAT	-9.4118	0.3325	0.0000
Locus_16502.0 BlastHit gi 118572923 sp Q3SZX4.3 CAH3_BOVIN	-9.3595	0.2906	0.0000
Locus_23144.0_reverse_transcriptase,_partial_[Panaeus_monodon]	-9.3571	0.2970	0.0000
Locus_23114.0_gastrolith_protein_[Cherax_quadricarinatus]	-9.2800	0.2100	0.0000
Locus_38772.0 BlastHit gi 728859 sp Q06527.1 ANKH_ALLVD	-9.2772	0.2120	0.0000
Locus_23043.0_Mitochondrial_tumor_suppressor_1-like_protein_[Camponotus_floridanus]	-9.2160	0.1613	0.0000
Locus_11935.0_hypothetical_protein_L798_01417,_partial_[Zootermopsis_nevadensis]	-9.1931	0.1328	0.0000
Locus_23143.0_G117733_[Drosophila_mojavensis]	-9.1330	0.0842	0.0000
Locus_23145.0_hypothetical_protein_Phum_PHUM213790_[Pediculus_humanus_corporis]	-9.0560	0.0118	0.0000
Locus_23285.0_PREDICTED_uncharacterized_protein_LOC101893139_[Musca_domestica]	-9.0151	-0.0269	0.0000
Locus_23012.0_splicing_factor_45_[Danaus_plexippus]	-8.9900	-0.0492	0.0000
Locus_23005.0_hypothetical_protein_L798_15326_[Zootermopsis_nevadensis]	-8.8525	-0.1771	0.0000
Locus_23064.0_conserved_hypothetical_protein_[Pediculus_humanus_corporis]	-8.8390	-0.1940	0.0000
Locus_16502.0 BlastHit gi 31076619 sp Q8UWA5.3 CAH2_TRIHK	-8.7411	-0.3006	0.0000
Locus_23076.0_PREDICTED_uncharacterized_protein_LOC103307708_[Acyrtosiphon_pisum]	-8.7240	-0.2982	0.0000
Locus_19192.0_Lethal(3)malignant_brain_tumor-like_3_protein_[Zootermopsis_nevadensis]	-8.6323	-0.3976	0.0000
Locus_23006.0_hypothetical_protein_DAPPUDRAFT_329602_[Daphnia_pulex]	-8.6081	-0.4061	0.0000
Locus_21528.0 BlastHit gi 121949 sp P06144.1 H1_LYTP1	-8.5395	-0.4968	0.0000
Locus_32542.0_PREDICTED_uncharacterized_protein_LOC102208367_isoform_X1_[Pundamilia_nyererei]	-8.5175	-0.4902	0.0000
Locus_23285.0_GJ17388_[Drosophila_virilis]	-8.4992	-0.5089	0.0000
Locus_23011.0_hypothetical_protein_EAG_07008_[Camponotus_floridanus]	-8.4908	-0.5130	0.0000
Locus_23019.0_PREDICTED_egl_nine_homolog_1-like_[Saccoglossus_kowalevskii]	-8.4849	-0.5335	0.0000
Locus_22963.0_PREDICTED_uncharacterized_protein_LOC100901383_[Metaseiulus_occidentalis]	-8.4410	-0.5600	0.0000
Locus_22949.0_PREDICTED_serine/arginine_repetitive_matrix_protein_2_[Nasonia_vitripennis]	-8.4350	-0.5618	0.0000
Locus_22963.0_hypothetical_protein_DAPPUDRAFT_62084_[Daphnia_pulex]	-8.4289	-0.5689	0.0000

Locus_23050.0_hypothetical_protein_[Scylla_paramamosain]	-8.4084	-0.5912	0.0000
Locus_22992.0_prophenoloxidase-activating_enzyme_2a_[Penaeus_monodon]	-8.3963	-0.5991	0.0000
Locus_18704.0_PREDICTED_nucleolin_[Anolis_carolinensis]	-8.3659	-0.2734	0.0000
Locus_24802.0 BlastHit gi 48428798 sp Q9I993.1 BCDO1_CHICK	-8.3347	-0.6732	0.0000
Locus_23279.0_predicted_protein_[Nematostella_vectensis]	-8.3208	-0.6688	0.0000
Locus_39723.0 BlastHit gi 81895664 sp Q812A3.1 ANR23_MOUSE	-8.3073	-0.6888	0.0000
Locus_30928.0_PREDICTED_protein_unc-13_homolog_D_[Ochotona_princeps]	-8.2817	1.6240	0.0000
Locus_5243.0 BlastHit gi 117612 sp P04813.1 CTR2_CANFA	-8.2797	-0.7226	0.0000
Locus_44285.0_abdominal-B_isoform_I_[Parhyale_hawaiensis]	-8.2547	-0.7553	0.0000
Locus_19270.1 BlastHit gi 81876135 sp Q8C255.1 DPEP2_MOUSE	-8.2217	-0.7717	0.0000
Locus_755.0 BlastHit gi 75026432 sp Q9V7U0.1 RESIL_DROME	-8.1978	-0.7915	0.0000
Locus_27467.0_Tripartite_motif-containing_protein_2_[Crassostrea_gigas]	-8.1916	-0.8119	0.0000
Locus_31059.1 BlastHit gi 57012917 sp Q19269.2 NAS14_CAEEL	-8.1902	-0.8178	0.0000
Locus_23176.0_hypothetical_protein_DAPPUDRAFT_319198_[Daphnia_pulex]	-8.1617	-0.8153	0.0000
Locus_22973.0_lectin_3_[Macrobrachium_rosenbergii]	-8.1098	-0.8567	0.0000
Locus_22960.0_PREDICTED_uncharacterized_protein_LOC662936_[Tribolium_castaneum]	-8.0119	-0.9516	0.0000
Locus_39927.0_putative_reverse_transcriptase_[Marsupenaeus_japonicus]	-7.9776	5.5218	0.0000
Locus_23263.0_hypothetical_protein_BRAFLDRAFT_86771_[Branchiostoma_floridae]	-7.9657	-0.9918	0.0000
Locus_718.0_unnamed_protein_product_[Tetraodon_nigroviridis]	-7.9481	-1.0173	0.0000
Locus_34656.0_PREDICTED_uncharacterized_protein_LOC101240642_[Hydra_vulgaris]	-7.9299	-1.0306	0.0000
Locus_31385.0_PREDICTED_afadin_isoform_X2_[Tribolium_castaneum]	-7.9194	-1.0747	0.0000
Locus_23310.0_transposase_[Portunus_granulatus]	-7.9081	-1.0551	0.0000
Locus_23248.0_predicted_protein_[Trichoplax_adhaerens]	-7.9004	-1.0439	0.0000
Locus_31894.0 BlastHit gi 75026432 sp Q9V7U0.1 RESIL_DROME	-7.8168	-1.1444	0.0000
Locus_28077.0 BlastHit gi 82123133 sp O57579.1 AMPN_CHICK	-7.8008	-1.1677	0.0000
Locus_21456.0 BlastHit gi 182705256 sp P18172.4 DHGL_DROPS	-7.7656	-1.1707	0.0000
Locus_10746.0 BlastHit gi 67462083 sp Q9Y115.1 UN93L_DROME	-7.6700	1.1206	0.0000
Locus_30636.0 BlastHit gi 189029946 sp Q9HCD6.3 TANC2_HUMAN	-7.6600	-1.2685	0.0000
Locus_33872.0_RAB6-interacting_golgin_[Zootermopsis_nevadensis]	-7.6179	-1.3092	0.0000
Locus_22949.0_PREDICTED_MAP7_domain-containing_protein_1-like_isoform_X1_[Tribolium_castaneum]	-7.5632	-1.3566	0.0000
Locus_27322.0 BlastHit gi 115881 sp P04069.1 CBPB_ASTFL	-7.4559	3.6207	0.0000
Locus_23103.0_hypothetical_protein_SINV_10203_[Solenopsis_invicta]	-7.4264	-1.4541	0.0000
Locus_23188.0_hypothetical_protein_[Macrobrachium_nipponense]	-7.2501	-1.6069	0.0000
Locus_36760.0 BlastHit gi 74717091 sp Q96PH1.1 NOX5_HUMAN	-6.0048	0.3701	0.0000
Locus_14008.0 BlastHit gi 416927 sp P25331.2 ECLH_BOMMO	-5.9077	2.6464	0.0000
Locus_10825.0 BlastHit gi 1352370 sp P98074.1 ENTK_PIG	-5.7611	1.6457	0.0000
Locus_26180.0 BlastHit gi 123582 sp P22978.1 HS6B_DROME	-5.4704	-1.0579	0.0000
Locus_9057.0 BlastHit gi 118572922 sp Q1LZA1.3 CAH1_BOVIN	-5.2561	0.5887	0.0000
Locus_16207.0_hypothetical_protein_LOTGIDRAFT_239733_[Lottia_gigantea]	-5.1870	1.2854	0.0000

Locus_20827.0 BlastHit gi 374110746 sp B3A0Q8.1 PLSP2_LOTGI	-5.0750	0.1079	0.0000
Locus_41122.0_Zinc_finger_protein_ubi-d4_A_[Harpegnathos_saltator]	-4.4612	1.2000	0.0000
Locus_3714.0_conserved_hypothetical_protein_[Pediculus_humanus_corporis]	-4.4228	-0.1007	0.0000
Locus_17874.0 BlastHit gi 137111 sp P15120.1 UROK_CHICK	-4.3403	2.9845	0.0000
Locus_29395.0 BlastHit gi 25167323 sp O60079.1 UBP12_SCHPO	-4.1810	-1.4912	0.0000
Locus_10790.0 BlastHit gi 20137194 sp Q9JJ22.2 ERAP1_RAT	-3.9287	0.1747	0.0004
Locus_34398.0_hypothetical_protein_DAPPUDRAFT_122319_[Daphnia_pulex]	-3.6204	-0.4156	0.0019
Locus_30711.0 BlastHit gi 81908906 sp Q505D1.1 ANR28_MOUSE	-3.6089	-1.2875	0.0000
Locus_29079.0 BlastHit gi 123907332 sp Q28DV3.1 HDAC3_XENTR	-3.4192	3.4934	0.0000
Locus_28014.0 BlastHit gi 109892506 sp Q3ZBF8.1 CERS2_BOVIN	-3.2046	-1.1557	0.0000
Locus_30778.0 BlastHit gi 387912917 sp Q01484.4 ANK2_HUMAN	-3.1413	-0.8787	0.0000
Locus_19000.0 BlastHit gi 17368331 sp P82968.1 MCPI_MELCP	-3.0859	-0.3598	0.0000
Locus_8361.0_hypothetical_protein_CAPTEDRAFT_200095_[Capitella_teleta]	-3.0140	0.2371	0.0036
Locus_28664.0_cln3/battenin_[Culex_quinquefasciatus]	-2.9053	-0.7217	0.0006
Locus_36845.0_endonuclease-reverse_transcriptase_[Bombyx_mori]	-2.8488	0.3172	0.0427
Locus_11507.0 BlastHit gi 81896794 sp Q8BN59.1 LARP6_MOUSE	-2.8075	7.4118	0.0034
Locus_29902.1 BlastHit gi 189082520 sp Q9UQC9.2 CLCA2_HUMAN	-2.6074	0.6960	0.0162
Locus_38555.0_PREDICTED_uncharacterized_protein_LOC101737182_[Bombyx_mori]	-2.5797	2.1686	0.0014
Locus_38563.0_Hypothetical_protein_CBG13367_[Caenorhabditis_briggsae]	-2.5590	3.8567	0.0031
Locus_30863.1 BlastHit gi 13124197 sp P82147.1 L2EFL_DROME	-2.5380	4.0354	0.0022
Locus_23376.0 BlastHit gi 30315951 sp Q9N0Z4.2 AT11B_RABIT	-2.5222	-0.9268	0.0347
Locus_17042.0 BlastHit gi 82228760 sp Q502K3.1 ANR52_DANRE	-2.5143	-1.0854	0.0169
Locus_22040.0_unnamed_protein_product_[Oncorhynchus_mykiss]	-2.5128	0.3897	0.0021
Locus_36770.0 BlastHit gi 47117286 sp Q98931.1 LRP8_CHICK	-2.5010	-0.8159	0.0152
Locus_30970.0 BlastHit gi 117330 sp P80007.1 CRA2_HOMGA	-2.4706	-1.1170	0.0074
Locus_24921.0 BlastHit gi 334350829 sp A2AGA4.1 RHBL2_MOUSE	-2.4525	-0.5256	0.0296
Locus_16255.0 BlastHit gi 462150 sp P34904.1 GBRG4_CHICK	-2.3988	-1.2281	0.0110
Locus_29753.0 BlastHit gi 50401204 sp Q9VCA2.1 ORCT_DROME	-2.3553	1.7556	0.0081
Locus_8385.0 BlastHit gi 74644329 sp Q8TGM6.1 TAR1_YEAST	-2.3237	2.4376	0.0025
Locus_5248.0_PREDICTED_uncharacterized_protein_LOC101848009_[Aplysia_californica]	-2.2825	-0.9074	0.0212
Locus_30855.0 BlastHit gi 10719883 sp P57078.1 RIPK4_HUMAN	-2.2627	-0.8399	0.0053
Locus_23503.0 BlastHit gi 342187367 sp Q8R151.3 ZNF1_MOUSE	-2.1950	-0.8986	0.0003
Locus_1788.0_PREDICTED_protein_unzipped_[Nasonia_vitripennis]	-2.1467	-0.8714	0.0274
Locus_25317.0 BlastHit gi 300681185 sp B0WC46.1 TRET1_CULQU	-2.1403	1.8071	0.0152
Locus_27155.0 BlastHit gi 81868592 sp Q9JJ06.1 C1GLT_MOUSE	-2.1305	-0.9200	0.0354
Locus_25626.0 BlastHit gi 11386708 sp Q9VE00.2 C12A4_DROME	-2.1060	-0.6622	0.0353
Locus_8599.0 BlastHit gi 1707896 sp P55040.1 GEM_HUMAN	-2.0681	-1.0059	0.0251
Locus_27023.0_GA15113_[Drosophila_pseudoobscura_pseudoobscura]	-2.0314	-0.3978	0.0074
Locus_13219.0_Tectonic-2_[Crassostrea_gigas]	-1.9767	0.6861	0.0001

Locus_21425.0 BlastHit gi 338817662 sp C0KJQ4.2 ALPS_PORTR	-1.9766	-1.1091	0.0053
Locus_30488.1 BlastHit gi 75061728 sp Q5R7D3.1 HSP71_PONAB	-1.9324	-0.5447	0.0263
Locus_41733.0_reverse_transcriptase_partial_[Penaeus_monodon]	-1.9027	4.2642	0.0377
Locus_25082.0 BlastHit gi 20455517 sp P17789.2 TTKB_DROME	-1.8991	-0.5582	0.0036
Locus_19520.0_hypothetical_protein_L798_13582_[Zootermopsis_nevadensis]	-1.7702	2.7598	0.0003
Locus_23974.0 BlastHit gi 45644947 sp P11536.2 E74EB_DROME	-1.5497	2.0979	0.0377
Locus_28939.2 BlastHit gi 261263180 sp A4II20.2 EGR1_XENTR	-1.3953	1.0434	0.0139
Locus_30951.0 BlastHit gi 290463239 sp C6KFA3.1 GP126_DANRE	-1.3343	1.9256	0.0039
Locus_28236.0 BlastHit gi 166225158 sp A2BGA0.1 RFX4_DANRE	-1.2307	2.7856	0.0097
Locus_26066.0 BlastHit gi 74871454 sp Q9VX32.2 RG190_DROME	-0.9629	2.7257	0.0211
Locus_21918.0 BlastHit gi 82233871 sp Q5ZL16.1 WIP13_CHICK	-0.9316	0.7394	0.0137
Locus_29467.1 BlastHit gi 51702257 sp P63036.1 DNJA1_RAT	-0.8939	3.8434	0.0073
Locus_21498.0_Organic_cation_transporter_protein_[Zootermopsis_nevadensis]	-0.8825	2.6286	0.0320
Baixo Crescimento			
Locus_25494.0 BlastHit gi 125950361 sp Q86U28.2 ISCA2_HUMAN	0.5975	2.5356	0.0376
Locus_28792.0 BlastHit gi 82178112 sp Q567E6.1 CQ10B_DANRE	0.6669	0.8392	0.0249
Locus_21674.0 BlastHit gi 226709091 sp P10586.2 PTPRF_HUMAN	0.7186	0.5397	0.0446
Locus_24884.0_PREDICTED_uncharacterized_protein_LOC580526_[Strongylocentrotus_purpuratus]	0.7997	-0.0104	0.0324
Locus_21142.0 BlastHit gi 20137671 sp Q9UBX3.2 DIC_HUMAN	0.9464	3.4219	0.0067
Locus_29786.0 BlastHit gi 543859 sp P37064.1 ASO_CUCPM	0.9581	2.4704	0.0253
Locus_22749.0 BlastHit gi 82181849 sp Q68F13.1 DOPDA_XENLA	1.0246	1.4721	0.0177
Locus_29919.0 BlastHit gi 257051061 sp Q12955.3 ANK3_HUMAN	1.0579	1.7194	0.0361
Locus_24539.0 BlastHit gi 48428477 sp Q94527.1 NFKB1_DROME	1.0694	0.1947	0.0478
Locus_27160.0 BlastHit gi 1708509 sp P10184.2 IOV7_CHICK	1.1556	1.2073	0.0269
Locus_26222.0 BlastHit gi 338817662 sp C0KJQ4.2 ALPS_PORTR	1.1680	4.8270	0.0471
Locus_22851.0 BlastHit gi 338817662 sp C0KJQ4.2 ALPS_PORTR	1.1784	3.7642	0.0249
Locus_28953.0 BlastHit gi 81906126 sp Q9DC37.1 MFSD1_MOUSE	1.2303	1.7098	0.0249
Locus_28928.0 BlastHit gi 82180006 sp Q5U4K5.1 TM41B_XENLA	1.2369	2.2006	0.0008
Locus_25503.0_PREDICTED_uncharacterized_protein_LOC101863835_isoform_X1_[Aplysia_californica]	1.2520	3.0955	0.0008
Locus_11852.0 BlastHit gi 126599 sp P17897.1 LYZ1_MOUSE	1.2771	5.0084	0.0195
Locus_31128.0 BlastHit gi 152031623 sp P02515.4 HSP22_DROME	1.2858	1.8831	0.0299
Locus_37969.0_hypothetical_protein_WUBG_04557_[Wuchereria_bancrofti]	1.3026	4.4957	0.0002
Locus_21476.0 BlastHit gi 34921426 sp O96790.2 DPGN_DIPMA	1.3273	3.8286	0.0028
Locus_15894.1 BlastHit gi 338817968 sp Q3UU35.2 OVOS_MOUSE	1.3343	-0.2819	0.0051
Locus_25617.0_hypothetical_protein_L798_10755_[Zootermopsis_nevadensis]	1.3950	3.0283	0.0160
Locus_26595.0 BlastHit gi 160358734 sp A6QP84.2 SOAT_BOVIN	1.4634	1.4416	0.0139
Locus_29736.0 BlastHit gi 73620982 sp P91622.2 PDK_DROME	1.4757	3.3586	0.0004
Locus_19148.0_GD19304_[Drosophila_simulans]	1.6563	0.8516	0.0074
Locus_28620.0 BlastHit gi 75059261 sp Q29042.1 FCN1_PIG	1.7088	-0.4629	0.0027

Locus_20351.0 BlastHit gi 190360177 sp Q80ZD3.2 S2611_MOUSE	1.8588	1.1631	0.0210
Locus_21563.0 BlastHit gi 6016173 sp P78417.2 GSTO1_HUMAN	2.0760	0.2527	0.0249
Locus_15931.1 BlastHit gi 5921938 sp P81578.1 CUPA4_CANPG	2.1296	4.3220	0.0422
Locus_25519.0 BlastHit gi 338817662 sp C0KJQ4.2 ALPS_PORTR	2.1564	2.6918	0.0012
Locus_27719.0 BlastHit gi 76363507 sp Q9D3D0.3 TTPAL_MOUSE	2.2022	3.0793	0.0031
Locus_20802.0_Segmentation_protein_cap'n'collar_[Zootermopsis_nevadensis]	2.2190	3.8199	0.0233
Locus_39260.0_Prototo-oncogene_tyrosine-protein_kinase_ROS_[Zootermopsis_nevadensis]	2.3831	1.6920	0.0129
Locus_27377.0_PREDICTED_uncharacterized_protein_LOC100869681_[Apis_florea]	2.4647	2.3487	0.0053
Locus_27922.0 BlastHit gi 61211707 sp O93403.1 CHST3_TORCA	3.0472	4.5974	0.0001
Locus_29902.1 BlastHit gi 81892746 sp Q6Q473.2 CLCA4_MOUSE	3.2047	1.8528	0.0139
Locus_37905.0_hypothetical_protein_DAPPUDRAFT_63008_[Daphnia_pulex]	3.3249	6.5734	0.0069
Locus_11821.0_ENSANGP00000021035-like_[Litopenaeus_vannamei]	3.4640	1.9205	0.0055
Locus_23346.0_Juvenile_hormone-inducible_protein_putative_[Aedes_aegypti]	3.8216	2.6354	0.0004
Locus_16240.0_PREDICTED_methyltransferase-like_protein_22-like_[Musca_domestica]	4.0024	2.9818	0.0000
Locus_19698.0_PREDICTED_MD-2-related_lipid-recognition_protein-like_partial_[Nasonia_vitripennis]	4.5513	9.7545	0.0072
ocus_29637.0 BlastHit gi 5921939 sp P81579.1 CUPA5_CANPG	4.6158	6.9594	0.0044
Locus_42808.0_hypothetical_protein_L798_08670_[Zootermopsis_nevadensis]	4.9162	2.4590	0.0039
Locus_27571.0 BlastHit gi 5921939 sp P81579.1 CUPA5_CANPG	5.0517	6.4026	0.0012
Locus_16326.0_hypothetical_protein_EAG_06277_[Camponotus_floridanus]	7.1235	4.8323	0.0001
Locus_34877.0_PREDICTED_receptor-type_tyrosine-protein_phosphatase_N2_[Tribolium_castaneum]	7.8057	5.2067	0.0000

Manuscrito 6

Tabela S1

Fragmentos Illumina				
Animais expostos ao WSSV				
Tecido	Músculo (8 bibliotecas)		Hepatopâncreas (8 bibliotecas)	
Tratamento	Saudável	Doente	Saudável	Doente
Fragmentos <i>paired-end</i>	248.795.146	235.537.380	266.191.030	233.223.996
Média dos fragmentos <i>paired-end</i> /biblioteca (milhões)	62.1	58.8	66.5	58.3
Desvio padrão	5.585.147	4.580.309	6.611.724	6.768.318
Total por tecido	484.332.526		499.415.026	
Total	983.747.552			
Após filtragem no SeqyClean				
Fragmentos <i>paired-end</i>	195.463.320	187.180.664	213.313.414	190.704.770
Média dos fragmentos <i>paired-end</i> /biblioteca (milhões)	48.8	46.7	53.3	47.6
Desvio padrão	4.296.012	3.496.056	5.188.692	5.647.694
Total por tecido	382.643.984		404.018.184	
Total	786.662.168 (80%)			
Após o mapeamento				
Fragmentos <i>paired-end</i>	191.577.540	180.169.952	200.513.272	183.685.654
Total por tecido	371.747.492		384.198.926	
Total mapeado	755.946.418 (96%)			