

BICM

**UNIVERSIDADE FEDERAL DO CEARÁ  
CENTRO DE CIÊNCIAS AGRÁRIAS  
DEPARTAMENTO DE ENGENHARIA DE PESCA  
CURSO DE ENGENHARIA DE PESCA**

**RENATA PINHEIRO CHAVES**

**CARACTERIZAÇÃO DE GENES EXPRESSOS A PARTIR DA GERAÇÃO DE ESTS  
(EXPRESSED SEQUENCE TAGS) DO GENOMA DO CAMARÃO MARINHO**

*Litopenaeus vannamei*

**FORTALEZA**

**2011**

**RENATA PINHEIRO CHAVES**

**CARACTERIZAÇÃO DE GENES EXPRESSOS A PARTIR DA GERAÇÃO DE ESTS  
(EXPRESSED SEQUENCE TAGS) DO GENOMA DO CAMARÃO MARINHO**

*Litopenaeus vannamei*

Monografia submetida à Coordenação do Curso de Engenharia de Pesca, da Universidade Federal do Ceará, como requisito parcial para obtenção do grau de Bacharel em Engenharia de Pesca.

Área de atuação: Biologia Aquática

**Orientador: Prof. Dr. Thalles Barbosa Grangeiro.**

**FORTALEZA**

**2011**

Dados Internacionais de Catalogação na Publicação  
Universidade Federal do Ceará  
Biblioteca de Ciências e Tecnologia

- 
- C439c Chaves, Renata Pinheiro.  
Caracterização de genes expressos a partir da geração de ESTs (Expressed Sequence Tags) do genoma do camarão marinho *Litopenaeus vannamei* / Renata Pinheiro Chaves. – 2011.  
75 f. : il. color., enc. ; 30 cm.
- Monografia (graduação) – Universidade Federal do Ceará, Centro de Ciências Agrárias, Departamento de Engenharia de Pesca, Curso de Engenharia de Pesca, Fortaleza, 2011.  
Orientação: Prof. Dr. Thalles Barbosa Grangeiro.
1. Carcinicultura. 2. *Litopenaeus vannamei*. 3. Projeto Genoma EST. I. Título.

---

CDD 639.2

**RENATA PINHEIRO CHAVES**

**CARACTERIZAÇÃO DE GENES EXPRESSOS A PARTIR DA GERAÇÃO DE ESTS (EXPRESSED SEQUENCE TAGS) DO GENOMA DO CAMARÃO MARINHO *Litopenaeus vannamei***

Monografia submetida à Coordenação do Curso de Engenharia de Pesca em 22 de novembro de 2011, da Universidade Federal do Ceará, como requisito parcial para obtenção do grau de Bacharel em Engenharia de Pesca. Área de atuação: Biologia Aquática.

Aprovado em: \_\_\_\_/\_\_\_\_/\_\_\_\_.

**BANCA EXAMINADORA**



---

Prof. Dr. Thalles Barbosa Grangeiro (Orientador)  
Universidade Federal do Ceará – UFC

---

Prof. Dr. José Renato de Oliveira César  
Universidade Federal do Ceará – UFC

---

Prof. Dr. Manuel Antonio Furtado Neto  
Universidade Federal do Ceará – UFC

Dedico este trabalho a minha mãe, por ter me colocado no caminho certo e pela força e apoio durante toda a minha jornada acadêmica.

Ainda por todas as alegrias e paz que ela e meu irmão me proporcionaram e ainda me darão.

A toda a minha família por ter acreditado em mim.

“Um amor mais forte que tudo, mais obstinado que tudo, mais duradouro que tudo, é somente o amor de mãe”

Paul Raynal

## AGRADECIMENTOS

A Deus e minha querida mãe, que me ensinaram tudo que eu sei e traçaram o melhor caminho para mim.

A toda a minha família, primos (Anarda Pinheiro, Arilo Pinheiro, Vitor Pinheiro, Diego Pinheiro, Matheus Pinheiro, Ádamo Pinheiro, Aron Pinheiro, Cazuza Pinheiro e Sávio Pinheiro) e meu irmão Rafael Pinheiro pela amizade, confiança, carinho e companherismo nessa jornada.

Aos meus amigos da faculdade, Arthur Melo, Ana Caroline, Rômulo Farias, João Henrique, Nara Rúbia, Guilherme Lima, Fernando Edson, Mirgon Conde e muitos outros, por toda alegria, estudos, conselhos, brigas e amizade nessa jornada academia, que ainda haja muitas histórias por vir. Aos meus grandes amigos de sempre, Marilce Lopes, Sandna Sanny, Nayla Uchoa e Mayara Uchoa. Aos meus amigos de balada e aos que se encontram distantes, mas que permanecem na lembrança.

A oportunidade dada pelo professor Thalles Barbosa Grangeiro, pelos ensinamentos e confiança depositado em minha pessoa. A minha segunda casa, laboratório de genética molecular, com meus amigos Patricia Gadelha, Edvar Junior, Raquel Sombra, Lucas Takase, Juliana, Bruno, Priscila e Marcel, Eder e Ana Beatriz. Em especial a Suelen Carneiro e Denise Rocha, por toda ajuda, ensinamentos e amizade. A minha co-orientadora e amiga Marina Lobo, por toda a amizade, conselhos, puxões de orelhas, paciência e ensinamentos.

Só tenho a agradecer por tudo e todos!

"Quando recebemos um ensinamento  
Devemos receber como um valioso presente  
E não como uma dura tarefa.  
Eis aqui a diferença que transcende."

Albert Einstein

## RESUMO

A carcinicultura é um dos cultivos que vêm crescendo após a estabilidade da produtividade pesqueira, devido a sobrepesca. No entanto seu manejo inadequado, principalmente com relação aos reprodutores, em cativeiro e o isolamento das populações cultivadas têm ocasionado uma diminuição na variabilidade genética devido à deriva genética. O projeto genoma EST do camarão *Litopenaeus vannamei* (ShEST) teve por objetivo seqüenciar 300.000 Seqüências de Genes Expressos (ESTs) do *L. vannamei*, servindo como base para inúmeros estudos. O alinhamento de seqüências explora o grau de similaridades entre cadeias de DNA, RNA e proteínas. Um alto grau de similaridade indica uma alta probabilidade das funções executadas pelas moléculas comparadas serem semelhantes. O objetivo desse trabalho tem por identificar e caracterizar genes expressos, pelo sequenciamento de clones de bibliotecas de cDNA e geração de ESTs (etiquetas de seqüências expressas), a partir da comparação das ESTs produzidas contra o banco de dados NCBI. Clones de *E. coli* contendo plasmídeos recombinantes foram cultivados e os plasmídeos isolados pelo método de lise alcalina. Seqüências parciais dos insertos foram então determinadas em um seqüenciador MegaBACE™ 1000. A identificação dos genes expressos foi realizada através da análise de similaridade das ESTs com seqüências já depositadas no GenBank utilizando o programa BLAST. A eficiência do seqüenciamento foi de 88,25% e foi obtido das 3643 ESTs um total de 245 *contigs* e 609 *singletons*. Destas ESTs validadas 52,7% correspondem a cDNAs que possuem similaridade com seqüências codificando proteínas com função já conhecida. Dentre elas foram encontradas Histona H2A-VI, Catepsina L e Lectina tipo C, Ferritina, Proteína QM, Selenoproteína, Hemocianina e Tiorredoxina, que estão relacionadas com o sistema imunológico, em diferentes tecidos. Também foi encontrado Proteína sarcoplasmática de ligação a cálcio, Actina A1, Ubiquitina b, Crustacianina subunidade A2, Beta-1 tubulina, Alfa-2 tubulina, peptídeos codificados com conhecida ou presumida função imunológicas contra a infecção WSSV. Além de proteínas relacionadas com o estresse, tais como: Proteínas de choque térmico 90, Beta-adaptina e Metalotioneína, que foram encontradas em tecidos de hepatopâncreas, pedúnculo ocular e em ovos. Sabe-se que existem respostas diversificadas baseado em peptídeo antimicrobiano existente no camarão, e que as respostas antivirais nestes animais dependem, em parte, do reconhecimento de RNA viral, e que a expressão de genes com funções antibacterianas é modificada durante a resposta a infecções virais. Estes recursos funcionais do genoma poderão fornecer bases para a geração de hipóteses para orientar futuras pesquisas em crustáceo sobre interações hospedeiro-vírus.

**Palavras-Chave:** Transcriptoma, Etiquetas de Sequências Expressas, camarão *L. vannamei*.



## LISTA DE FIGURAS

<b>FIGURA 1</b>	<i>Litopenaeus vannamei</i> .....	13
<b>FIGURA 2</b>	Anatomia geral externa dos camarões da família Penaeidae.....	16
<b>FIGURA 3</b>	Anatomia geral interna dos camarões da família Penaeidae.....	16
<b>FIGURA 4</b>	Estágios larvais do camarão <i>Litopenaeus vannamei</i> : náuplio, zoea, mísis e pós-larva.....	19
<b>FIGURA 5</b>	Esquema da clonagem de cDNA e sequenciamento de ESTs (etiquetas de sequências expressas).....	23
<b>FIGURA 6</b>	Representação esquemática do sequenciamento <i>shotgun</i> por clonagens em plasmídeos e processo de montagem genômico.....	23
<b>FIGURA 7</b>	Eletroforese em gel de agarose 0,8% dos plasmídeos purificados das bibliotecas MC2078 (A), MC2079 (B), MC2080 (C), MC2081 (D) e MC2082 (E), MC2084 (F).....	33
<b>FIGURA 8</b>	Número de sobreposições de ESTs para formar <i>contigs</i> .....	35
<b>FIGURA 9</b>	Porcentagem da similaridade das sequências unigenes com sequências já depositadas no <i>GenBank</i> , por meio da ferramenta <i>BlastX</i> .....	37
<b>FIGURA 10</b>	Porcentagem da similaridade das sequências <i>contigs</i> das bibliotecas com sequências já depositadas no <i>GenBank</i> , por meio da ferramenta <i>BlastX</i> .....	38
<b>FIGURA 11</b>	Porcentagem da similaridade das sequências <i>singletons</i> das bibliotecas com sequências já depositadas no <i>GenBank</i> , por meio da ferramenta <i>BlastX</i> .....	39
<b>FIGURA 12</b>	Classificação de algumas funções das ESTs de acordo com o programa “ <i>GOanna</i> ”.....	40

## LISTA DE TABELAS

<b>TABELA 1</b>	Principais funções dos órgãos e tecidos dos camarões.....	17
<b>TABELA 2</b>	Resultados da análise individual e conjunta da qualidade das bases e montagem das ESTs das bibliotecas.....	35
<b>TABELA 3</b>	Porcentagem da similaridade das sequências unigenes das bibliotecas com sequências já depositadas no <i>GenBank</i> , por meio da ferramenta <i>BlastX</i> .....	37
<b>TABELA 4</b>	Porcentagem da similaridade das sequências <i>contigs</i> das bibliotecas com sequências já depositadas no <i>GenBank</i> , por meio da ferramenta <i>BlastX</i> .....	38
<b>TABELA 5</b>	Porcentagem da similaridade das sequências <i>singletons</i> das bibliotecas com sequências já depositadas no <i>GenBank</i> , por meio da ferramenta <i>BlastX</i> .....	39
<b>TABELA 6</b>	Definições de acordo com o programa on line “ <i>AmiGO</i> ” do site “ <i>The Gene Ontology</i> ” .....	41
<b>TABELA 7</b>	Algumas proteínas codificadas com conhecida ou presumida função imunológicas no camarão marinho <i>L. vannamei</i> .....	45
<b>TABELA 8</b>	Alguns peptídeos codificados com conhecida ou presumida função imunológicas contra a infecção WSSV.....	47
<b>TABELA 9</b>	Outras proteínas codificadas de função imunológica.....	47
<b>TABELA 10</b>	Proteínas codificadas com função relacionada em resposta ao estresse.....	50

## SUMÁRIO

<b>1</b>	<b>Introdução.....</b>	<b>11</b>
<b>1.2</b>	<b>Revisão de Literatura.....</b>	<b>13</b>
1.2.1	Aquicultura.....	13
1.2.2	<i>Litopenaeus vannamei</i> .....	14
1.2.2.1	Taxonomia e Morfologia.....	14
1.2.2.2	Cultivo.....	18
1.2.3	Projeto genoma EST do camarão marinho <i>L. vanammei</i> .....	21
1.2.4	Etiquetas de sequências expressas (ESTs).....	22
1.2.4.1	Bibliotecas ETSs para Crustáceos.....	24
1.2.5	Bioinformática.....	25
<b>2</b>	<b>Metodologia.....</b>	<b>27</b>
<b>2.1</b>	<b>Extração de plasmídeo.....</b>	<b>27</b>
<b>2.2</b>	<b>Eletroforese em gel agarose.....</b>	<b>28</b>
<b>2.3</b>	<b>Sequenciamento.....</b>	<b>28</b>
<b>2.4</b>	<b>Bioinformática.....</b>	<b>29</b>
2.4.1	Análise de Qualidade.....	30
2.4.2	Identificação das ESTs.....	30
2.4.3	Caracterização das ESTs.....	30
<b>3</b>	<b>Resultados e Discussão.....</b>	<b>32</b>
<b>3.1</b>	<b>Extração de Plasmídeo.....</b>	<b>32</b>
<b>3.2</b>	<b>Análise de qualidade.....</b>	<b>34</b>
<b>3.3</b>	<b>Identificação das ESTs.....</b>	<b>36</b>
<b>3.4</b>	<b>Classificação das ESTs.....</b>	<b>40</b>
<b>3.5</b>	<b>Caracterização das ESTs.....</b>	<b>44</b>
3.5.1	Proteínas codificadas relacionadas com a defesa inata.....	44

3.5.2	Proteínas codificadas relacionadas com o estresse.....	50
3.5.3	Codificação de outras proteínas.....	52
<b>4</b>	<b>Conclusão.....</b>	<b>54</b>
	<b>Referencias.....</b>	<b>55</b>
	<b>Anexos.....</b>	<b>61</b>

## 1 INTRODUÇÃO

Aqüicultura é o processo de produção de organismos aquáticos em cultivo controlado ou semi-controlado em qualquer estágio de desenvolvimento, não sendo um bem coletivo como as populações de recursos pesqueiros (CAMARGO; POUHEY, 2005). Este tipo de cultivo vem se expandindo e, cada vez mais, sustentando um crescimento de oferta global de alimentos diante a estabilização da pesca (ORMOND *et al.*, 2004).

Dentre o filo Arthropoda, o subfilo Crustacea contém muitas espécies marinhas e dulcícolas de interesse comercial para a pesca e aquicultura. As espécies comestíveis são as mais familiares às pessoas, como por exemplo, lagostas, lagostins, caranguejos e camarões (HICKMAN; ROBERTS; LARSON, 1997). Um dos mais importantes produtos da aqüicultura brasileira é o camarão marinho, sendo de destaque do atual desenvolvimento da carcinicultura brasileira a espécie *Litopenaeus vannamei*, originária do Oceano Pacífico (COELHO, 2005). Problemas oriundos do manejo em cativeiro e o isolamento das populações cultivadas têm ocasionado uma diminuição na variabilidade genética devido à deriva genética, isto é, à alteração aleatória da presença de alelos, modificando a sua frequência e característica de uma geração para outra (GANEM, 2011), acarretando um aumento à susceptibilidade a doenças causadas por patógenos virais, bacterianos e fúngicos (SILVA, 2007).

O projeto genoma EST do camarão *Litopenaeus vannamei* (ShEST), teve seu núcleo na Universidade Federal de São Carlos (UFSCar) e participação de 14 laboratórios de universidades brasileiras, dentre eles, o Laboratório de Genética Molecular do Departamento de Biologia da UFC. O projeto teve por objetivo seqüenciar 300.000 Seqüências de Genes Expressos (ESTs) do *L. vannamei*, servindo como base para inúmeros estudos (SANTOS; FREITAS, 2008).

Muitos bancos de dados sobre informações genômicas estão disponíveis para acesso, contudo todos trocam dados em um intervalo de 24 horas com o NCBI (Centro Nacional para Informações Biotecnológicas). O *GenBank* do NCBI é considerado o principal banco de dados armazenando seqüências disponíveis de DNA, RNA e proteínas (SANTOS; ORTEGA, 2003).

O objetivo geral desse trabalho teve por identificar e caracterizar genes expressos produzidos pelo sequenciamento de clones de bibliotecas de cDNA e geração de ESTs

(etiquetas de sequências expressas) do Projeto Genoma EST do camarão marinho *L. vannamei*, a partir da comparação das ESTs produzidas contra o banco de dados NCBI. Como objetivos específicos, analisar e verificar a qualidade das 3643 ESTs de diferentes tecidos (pedúnculo ocular, músculo, hepatopâncreas, fases larvais e ovo) obtidas pelo seqüenciador automático MegaBACE1000 DNA Sequencing System do Projeto Genoma EST do camarão marinho *L. vannamei* e, assim, caracterizar as 245 contigs e 609 singletons utilizando as ferramentas BLAST do banco de dados do NCBI na modalidade BLASTx.

## 1.2 Revisão de literatura

### 1.2.1 Aquicultura

A produção pesqueira mundial vem se estabilizando frente ao estancamento das capturas, observado desde meados dos anos 1980, devido à sobrepesca indiscriminada dos estoques pesqueiros naturais. Atualmente, a aquicultura se tornou uma alternativa para produção de alimento de alto valor protéico com o aumento da demanda de consumo humano (CAMARGO; POUEY, 2005), mantendo uma taxa média de crescimento anual de 8,3% mundialmente (ou 6,5% excluindo a China), entre 1970 e 2008 (FAO, 2010).

A aquicultura, tanto marítima como continental, tem atingido níveis de tecnologia, produção e produtividade cada vez mais representativos em todo mundo. A produção global de pescado oriunda da aqüicultura atingiu 52,3 milhões de toneladas em 2008, tendo contribuição de 45,7% na produção de pescado, incluindo peixes de barbatanas, crustáceos, moluscos e outros animais aquáticos para consumo humano (FAO, 2010). Aquicultura marinha está se tornando cada vez mais, um importante componente no setor de produção de pescado (FAO, 2010). Na maricultura, o cultivo de camarão é o que mais tem se desenvolvido (MARTINS, 2003). No Brasil, a espécie de camarão mais cultivado é o *Litopenaeus vannamei* (Figura 1), originária do Oceano Pacífico (COELHO, 2005).



**Figura 1** – *Litopenaeus vannamei* (Fonte: Wikipédia, Penaeidae, 2011)

## 1.2.2 *Litopenaeus vannamei*

### 1.2.2.1 Taxonomia e Morfologia

A classificação taxonômica do camarão marinho mais cultivado no Brasil é a seguinte:

Reino Animalia

Filo Arthropoda

Subfilo Crustacea **Pennant, 1777**

Classe Malacostraca **Latreille, 1806**

Subclasse Eumalacostraca **Grobben, 1892**

Super Ordem Eucarida **Calman, 1904**

Ordem Decapoda **Latreille, 1803**

Subordem Dendodranchiata **Bate, 1888**

Superfamília Penaeoidea **Rofinesque, 1815**

Família Penaeidea **Rofinesque, 1815**

Subfamília Penaeinae **Dana, 1852**

Gênero *Penaeus* **Fabricius, 1798**

*Litopenaeus*<sup>1</sup> **Peréz Farfante e Kenley, 1997**

Espécie *Litopenaeus vannamei* **Boone, 1931**

<sup>1</sup>Segundo recente revisão taxonômica realizada por Peréz Farfante e Kenley (1997)

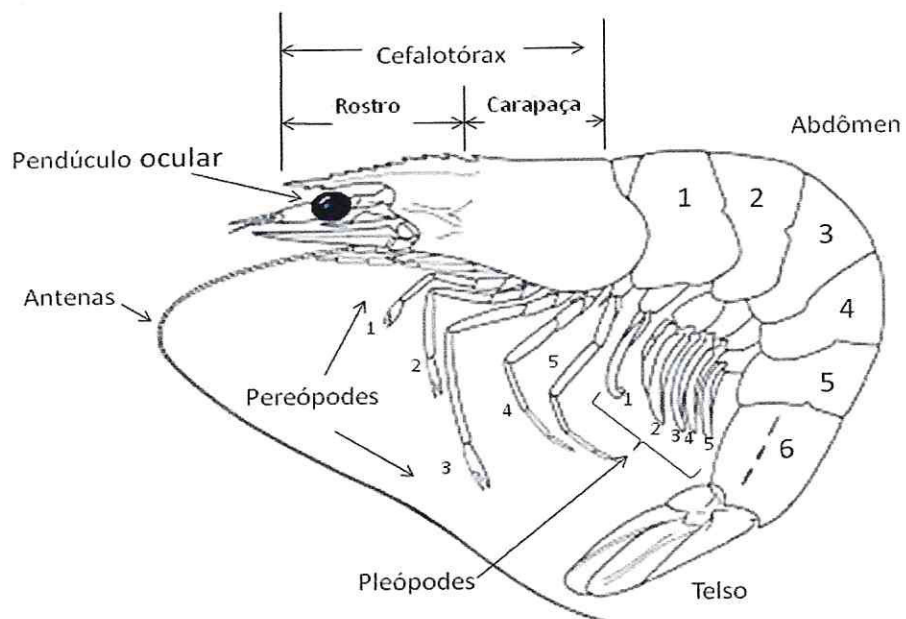
O filo Arthropoda é o maior filo do reino animal. Insetos e crustáceos somados compõem mais de 80% das espécies de animais descritas. Os crustáceos se diferenciam em vários aspectos dos demais artrópodes, sendo a presença de dois pares de antenas a principal característica. O maior grupo de crustáceos pertence à classe Malacostraca, que inclui lagostas, camarões, caranguejos, tatuzinhos-de-quintal e muitos outros (HICKMAN; ROBERTS; LARSON, 1997).



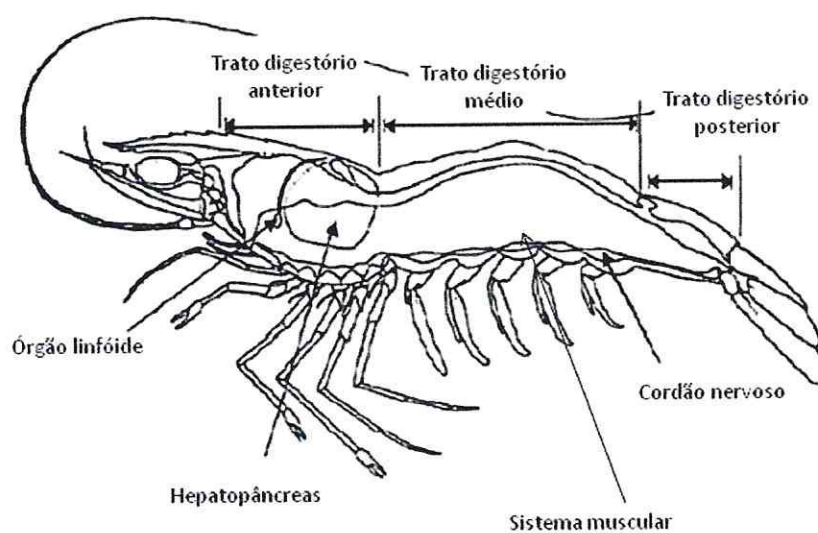
Os camarões são animais predominantemente aquáticos, que possuem carapaça externa e rígida (exoesqueleto calcário), cabeça e tórax fundido (cefalotórax), apêndices com dois ramos (birremes) (POLI *et al.*, 2004). O exoesqueleto dos camarões é constituído por quitina, proteínas e carbonato de cálcio, tendo seu corpo dividido em duas regiões: cefalotórax e abdômen (Figura 2). No cefalotórax se encontrma as estruturas de grande importância funcional, tais como: pedúnculo ocular, rostro, cérebro, coração, hepatopâncreas, estômago e as gônadas (Figura 3). No abdômen está localizado o ânus, um cordão nervoso e a maior parte da musculatura estriada (HICKMAN; ROBERTS; LARSON, 1997).

Os decápodes possuem três pares de maxilípodés, cinco pares de apêndices ambulatório (pereópodes) na região do cefalotórax e cinco pares de apêndices natatórios (pleópodes) na região do abdômen (HICKMAN; ROBERTS; LARSON, 1997). Os apêndices são caracterizados por apresentar dois ramos articulados (birremes) em cada somito e apresentam diversas funções, como a de alimentação, escavação, limpeza e captação de estímulos externos (POLI *et al.*, 2004).

As trocas de gases respiratórios ocorrem nas brânquias, que são encontradas nas laterais da carapaça, denominada de cavidade branquial. O sistema circulatório é caracterizado por ser “aberto” ou lacunar, ou seja, não havendo separação do fluido sanguíneo (HICKMAN; ROBERTS; LARSON, 1997). As principais funções dos órgãos e tecidos dos camarões estão descritas na Tabela 1.



**Figura 2** – Anatomia geral externa dos camarões da família Penaeidae. (Fonte: Adaptado do Departamento de Pesca y Acuicultura, FAO, 2011).



**Figura 3** – Anatomia geral interna dos camarões da família Penaeidae (Fonte: Adaptado de MARTINS, 2003).

**Tabela 1** – Principais funções dos órgãos e tecidos dos camarões

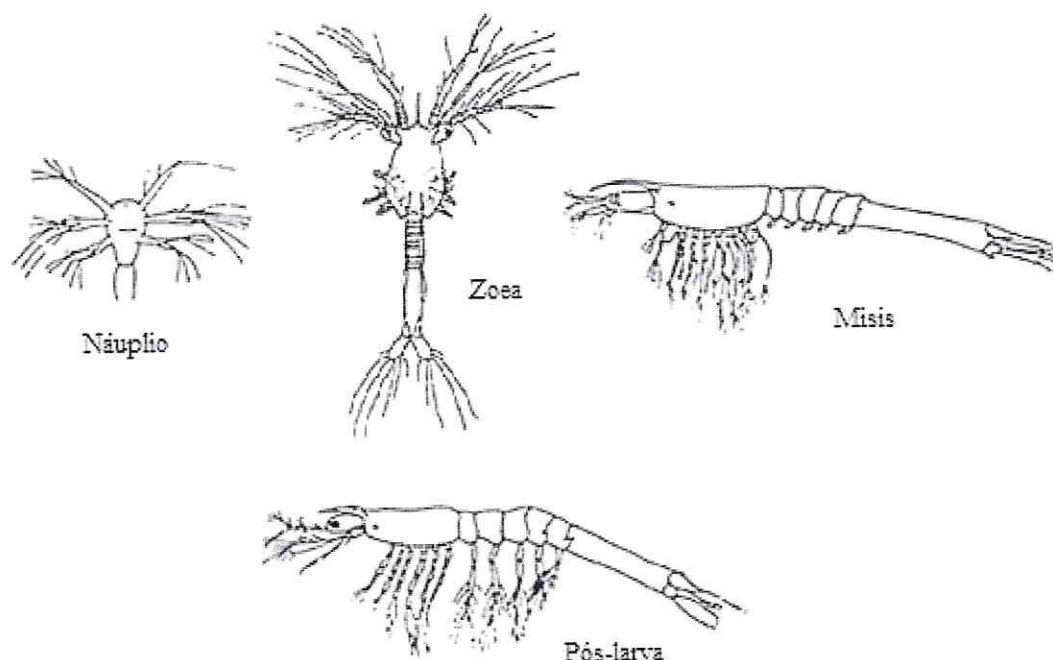
Órgão / Tecido	Principal Função
Rostro	Proteger contra predadores
Mandíbulas e maxilas	Órgãos sensitivos e desintegração do alimento
Maxílula e maxilípodés	Tato, paladar e manipulação dos alimentos
Antenas	Orgão sensitivo (detecção de predadores)
Gládula antenal	Excreção e balanço osmótico
Antênula	Sensorial (Quimiorecepção, táctil, equilíbrio)
Pendúculo Ocular	Produção de hormônios reguladores da muda e gametogênese
Exoesqueleto	Estrutura de suporte e barreira protetora
Brânquias	Respiração, excreção, osmorregulação e fagocitose
Hepatopâncreas	Digestão, absorção e armazenagem de nutrientes
Músculo abdominal	Movimentação rápida dos apêndices com objetivo de movimentação e fuga
Órgão linfóides	Possivelmente para captura de antígenos e fagocitose
Sistema digestório anterior	Ingestão, trituração e armazenamento temporário de alimento
Intestino	Absorção de nutrientes e excreção
Periópodés e pleópodes	Locomoção, captura e manipulação de alimentos
Urópodo e telso	Direcionamento da locomoção durante a natação

Fonte: Adaptado de BROCK & MAIN (1994)

### 1.2.2.2 Cultivo

O cultivo de camarões marinhos da espécie *L. vannamei* é realizado basicamente em duas etapas: a da larvicultura (produção de larvas) e a da engorda (crescimento do camarão até estar pronto para ser comercializado) (ORMOND *et al.*, 2004). As larvas são geralmente compradas de laboratórios especializados na reprodução dessa espécie ou se dá na própria fazenda, que comporta um laboratório para a reprodução. Nestas fazendas geralmente encontra-se um galpão totalmente fechado com os reprodutores, pois esta espécie possui características reprodutivas noturnas. Estes passam por um setor de quarentena antes do cruzamento e são separados de acordo com as famílias para que não ocorra endocruzamentos. Os reprodutores que não possuem deformidades são selecionados de acordo com suas características zootécnicas, crescimento rápido e resistência a doenças. As matrizes sofrem ablação, retirada de um único pedúnculo ocular, para facilitar a maturação ovariana (FRANCISCO, 2003). Após a fase de quarentena, os reprodutores são levados para os tanques de reprodução e as matrizes que realizaram a cópula são levadas para tanques de desova. Os ovos são recolhidos e levados para tanques de eclosão, onde permanecem até que os náuplios apresentem fototropismo positivo e assim serem levados para tanques de larvicultura ou serem comercializados para outras fazendas (ORMOND *et al.*, 2004).

A etapa de larvicultura no cultivo de camarões é dividida em duas fases: fase 1 e 2. A larvicultura fase 1 ocorre em ambiente fechado, onde os náuplios permanecem até se constituírem integralmente em pós-larvas, cerca de 13 dias (Figura 4). Já na fase 2, as pós-larvas permanecem nos tanques em ambiente aberto até a fase final de pós-larva, em média 12 dias, sendo levados então para a etapa da engorda. No período de larvicultura, a oferta de alimento é de microalgas e náuplios de artêmia. A segunda etapa do cultivo é a engorda, onde as pós-larvas passam aproximadamente 20 dias em tanques berçários para uma adaptação antes de povoar os viveiros. O ciclo total de engorda pode durar de 90 a 150 dias, dependendo das condições de cultivo e do peso que se deseja alcançar (ORMOND *et al.*, 2004).



**Figura 4** - Estágios larvais do camarão *Litopenaeus vannamei*: náuplio, zoea, mísis e pós-larva. (Fonte: Adaptado do Depósito de documentos, FAO, 2011).

O cultivo do camarão de camarões teve sua origem na costa oeste do continente Americano. O cultivo no nível de rentabilidade, onde foi capaz de atrair investidores de pequenos, médios e grandes produtores, firmou-se entre 1975 e 1985 (SEBRAE, 2008). A espécie *L. vannamei* possui excelentes condições de manejo, tais como: rápido crescimento, rusticidade, conversão alimentar eficiente obtendo grandes ganhos de peso e, com isso, passou a ser cultivada em 100% das fazendas brasileiras (MARTINS, 2003).

Por ser uma espécie exótica, os primeiros cultivos se deram pela adesão de reprodutores selvagens através de importações da costa do Pacífico, contudo uma lei decretada em 1997 pelo Instituto Brasileiro do Meio Ambiente (IBAMA) restringia a importação de crustáceos, moluscos, macrófitas e peixes, salvo aqueles destinados a aquarofilia, para reduzir os impactos sobre a biodiversidade nativa. O IBAMA juntamente com o Ministério da Agricultura e Abastecimento (MMA) possibilitou a eventual importação de alguns espécimes para o cultivo. Entretanto, a Associação Brasileira de Criadores de Camarão (ABCC) decidiu por não importar reprodutores para evitar a introdução de patógenos exógenos na carcinicultura brasileira (FREITAS, 2003).

A reprodução para o cultivo de espécies de camarão *L. vannamei* se dá agora pelo cruzamento entres indivíduos nascidos muitas vezes do mesmo cativeiro. Esta conduta pode, a médio e longo prazo, promover uma queda significativa dos níveis de variabilidade genética, uma vez que o cruzamento entre indivíduos pertencentes ao mesmo plantel tende a aumentar o grau de consangüinidade (FREITAS,2003). A redução da variabilidade genética pode possibilitar a perda de muitas características como a de matrizes com elevada eficiência de produção, a de adaptação a diferentes condições ambientais e o aumento à susceptibilidade a doenças causadas por patógenos virais, bacterianos e fúngicos (SILVA, 2007).

Um dos desafios encontrados no cultivo de camarão marinho é a infecção dos animais cultivados por doenças virais. A maioria das doenças virais identificada para camarão marinho foi descrita primeiramente em sistemas de cultivo, e a prevalência desses vírus em populações nativas parece estar diretamente relacionada ao aumento da atividade carcinicultura. O vírus da síndrome da mancha branca (WSSV), por exemplo, é a causa de uma doença viral que afeta principalmente camarões cultivados. Em novembro de 2004 camarões da espécie *L. vannamei* cultivados em fazendas na região de Laguna, sul do Brasil, foram infectados por este vírus, causando grande mortalidade e enormes perdas econômicas (CAVALLI *et al.*, 2008).

### 1.2.3 Projeto genoma EST do camarão marinho *L. vannamei*

O Projeto Genoma EST do camarão marinho *L. vannamei* (ShEST) teve seu núcleo na Universidade Federal de São Carlos (UFSCar) e foi concluído no início de 2006. O objetivo geral do projeto foi seqüenciar regiões expressas do genoma de *L. vannamei* para que estas servissem como base para inúmeros estudos, incluindo a identificação de um conjunto expressivo de marcadores moleculares e de genes de interesse que pudessem auxiliar em estudos genéticos, como caracterização de marcas SSRs e os diferentes genes existentes, suas proteínas regulatórias e suas funções (SANTOS; FREITAS, 2008).

Devido ao manejo nas condições de cativeiro, surgiram alguns problemas relacionados à redução do nível da variabilidade genética e do fitness, condições zootécnicas, dos estoques cativos, ocasionados pela elevada taxa de endocruzamentos. Assim, estudos para descrição de SSRs (Seqüências Simples Repetidas) em ESTs (*Expressed Sequence Tags*) foram realizados para auxiliar na identificação de QTLs (*Quantitative Trait Loci*), no desenvolvimento de programas de Seleção Assistida por Marcadores (MAS) e em estudos de mapeamento genético (SANTOS; FREITAS, 2008).

#### 1.2.4 Etiquetas de sequências expressas (ESTs)

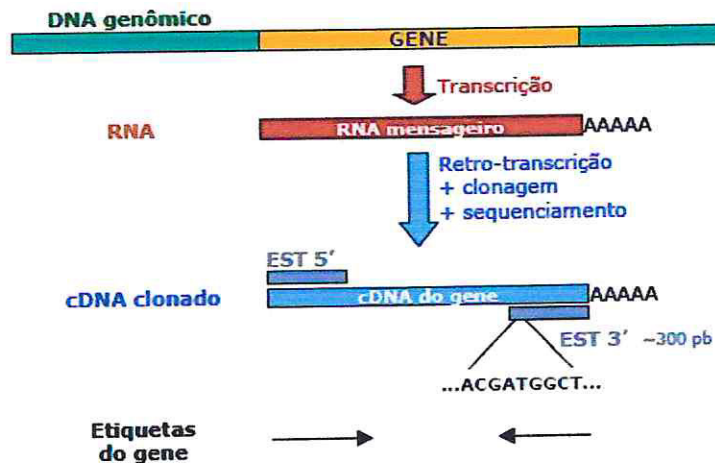
A transcrição reversa de um mRNA tem como produto um DNA complementar (cDNA), e o seu seqüenciamento se tornou uma técnica bem estabelecida e aceita (Figura 5). A partir da década de 1980, a estratégia de seqüenciamento *shotgun* foi introduzida e durante os anos seguintes, a aplicação das seqüências de DNA se deu por clonagens em plasmídeos (poucos kilobases (kb)) (Figura 6), cosmídeos (40 kb), cromossomos artificiais clonados em bactérias (BAC- *bacterial artificial chromosome*) e levedura (YAC- *yeast artificial chromosome*), com inserções de 100-500 kb (SCHEIBYE-ALSING *et al.*, 2009). O seqüenciamento *High throughput* (HTS), alta taxa de eficiência, de cDNA foi iniciado em 1991 por Adams (ADAMS *et al.*, 1991), que estabeleceu o termo "*Expressed Tag Sequence*" (EST) para se referir a este novo tipo de seqüenciamento. Para Hately *et al.* (1998), etiquetas de sequências expressas (ESTs) são seqüências parciais do cDNA resultante de passagem única, a partir das extremidades, de seqüenciamento de clones de bibliotecas de cDNA e diferentes ESTs podem ser obtidas a partir de um gene.

Bibliotecas de ESTs têm proporcionado uma diversidade de todos os genes codificadores de proteínas em um tecido, se tornando uma ferramenta importante e com várias aplicações, principalmente em relação à análise genética e descoberta do gene. Juntamente com o número crescente de genomas concluídos, o desenvolvimento de métodos computacionais para comparar genomas também é constantemente inovado, para se tornar, cada vez mais, possível a montagem de um grande número de fragmentos e, assim, aumentar confiabilidade dos resultados (SCHEIBYE-ALSING *et al.*, 2009).

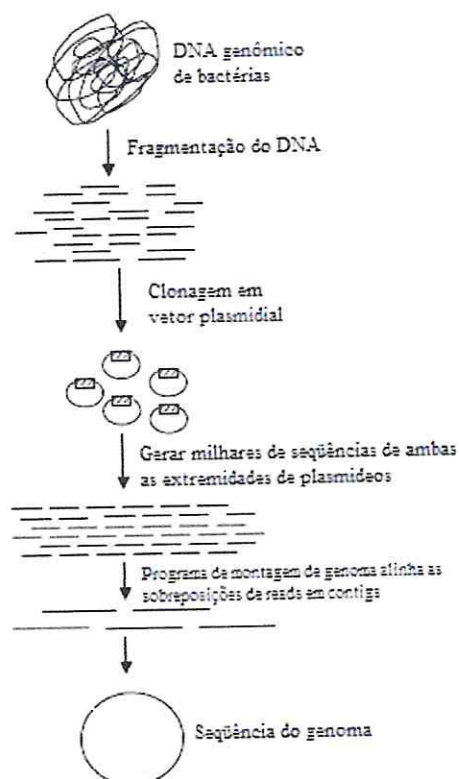
Com a descoberta da homologia de seqüências, onde uma proteína conhecida ou família de proteínas que pode fornecer as primeiras pistas sobre a possível função de um gene seqüenciado, e o crescente aumento de dados de seqüências de aminoácidos, que torna cada vez mais possível de encontrar tais homologias, torna-se útil a análise de genes seqüenciados (ALTSCHUL *et al.*, 1990).

Atualmente, com mais de 2000 projetos genoma diferentes, a partir de vírus até eucariontes, já concluído ou em andamento, estamos agora na era do genoma (MCGRATH, 2007).





**Figura 5** - Esquema da clonagem de cDNA e sequenciamento de ESTs (etiquetas de seqüências expressas). As ESTs são seqüências parciais do cDNA resultante de passagem única, a partir das extremidades, de sequenciamento de clones de bibliotecas de cDNA (Fonte: Adaptado de SANTOS; ORTEGA, 2003).



**Figura 6** - Representação esquemática do sequenciamento *shotgun* por clonagens em plasmídeos e processo de montagem genômico (Fonte: Adaptado de MCGRATH, 2007).

#### 1.2.4.1 Bibliotecas ESTs para Crustáceos

A construção de bibliotecas de cDNA de crustáceos, o desenvolvimento de etiquetas de seqüências expressas (ESTs) e sua deposição em bancos de dados acessíveis ao público gerou uma rica fonte para a descoberta de genes neste subfilo artrópode comercialmente e ecologicamente importantes (CHRISTIE *et al.*, 1998).

Gross *et al.* (2001) fizeram bibliotecas a partir de RNA de hematócitos e de hepatopâncreas dos camarões *Litopenaeus vannamei* e *Litopenaeus setiferus* para obtenção de gene imune e das 268 ESTs foram encontrados 44 genes que codificam para proteínas com funções imunológicas.

Freitas *et al.* (2005) relataram sobre a mineração de dados publicamente disponíveis de ESTs de *L. vannamei* para gerar marcadores de sequencias simples repetidas (SSRs) e sobre a sua transmissão entre as espécies relacionadas a camarões Peneídeos. Uma alta porcentagem (69%) das EST-SSRs (Sequências Repetidas Simples) foi transferida dentro do gênero *Litopenaeus*.

Maneeruttanarungroj *et al.* (2006) desenvolveram marcadores microssatélites para o mapeamento de ligação genética e programa de melhoramento genético do camarão tigre preto *Penaeus monodon*. A pesquisa por homologia representou, para as 997 ESTs, 8,6% de produtos de proteínas com função conhecida, 27,8% de proteínas hipotéticas e 63,6% de produtos de proteínas desconhecidas.

Andrew (2011) fez estudos filogenéticos, baseados em etiquetas de seqüências expressas (ESTs), sobre as relações dos grupos de artrópodes mais importantes: Chelicerata, Myriapoda, Crustacea, e Hexapoda. Segundo o mesmo autor o grupo Crustacea é o mais intimamente relacionado o Hexapoda, mas eles diferem em aspectos filogenéticos importantes, recentemente publicados, empregando sequencias de proteínas codificadas e, assim, afirma que ESTs sozinhas são incapazes de resolver convergências morfológicas ou evolutivas, e que são necessárias outras abordagens para resolver tais cenários evolutivas.

### 1.2.5 Bioinformática

A bioinformática é uma ciência recente que utiliza dados biológicos e conhecimento armazenado em bancos de dados de computador para obter novos conhecimentos biológicos. Atualmente, a bioinformática contribui com projetos de seqüenciamento do genoma a partir de organização e análise da enorme quantidade de dados codificados. A transferência de informações baseia-se na noção que as proteínas similares em seqüência também são semelhantes na função. O sistema de software realiza uma série de pesquisas do banco de dados e análises utilizando bases de dados atualizados diariamente, e fornece ao usuário final um relatório de síntese sobre a função da proteína consulta (ANDRADE; SANDER, 1997).

Muitos bancos de dados estão disponíveis em domínio público fornecidos por institutos acadêmicos e não acadêmicos (WHITFIELD; PRUESS; APWEILER, 2006).

Os bancos de dados envolvendo seqüências de nucleotídeos, de aminoácidos ou estruturas de proteínas podem ser classificados em bancos de seqüências primários e secundárias. Quando são deposições diretas de seqüências de nucleotídeos, aminoácidos ou estruturas protéicas, sem qualquer processamento e/ou análise são denominadas de sequencias primárias e quando são aqueles que derivam dos primários, ou seja, foram formados usando as informações depositadas nos bancos primários são denominados de sequencias secundárias. Os principais bancos de dados primários são o *GenBank*, o EBI (*European Bioinformatics Institute*), o DDBJ (*DNA Data Bank of Japan*) e o PDB (*Protein Data Bank*). Os três primeiros bancos são membros do INSDC (*International Nucleotide Sequence Database Collaboration*) e cada um desses centros possibilita a submissão individual de seqüências de DNA. Eles trocam informações entre si diariamente, de modo que todos os três possuem informações atualizadas de todas as seqüências de DNA depositadas em todo o mundo. Bancos de dados que atuam como secundários: o PIR (*Protein Information Resource*) ou o SWISS-PROT (PROSDOCIMI *et al.*, 2003).

O BLAST ou *Basic Local Alignment Search Tool* é a ferramenta mais popular de comparação de seqüências de DNA com os bancos de dados genômicos, através deste algoritmo podemos comparar uma seqüência de DNA ou proteína qualquer com todas as seqüências genômicas de domínio público. O resultado desta busca, que é feita no *GenBank* ou em várias de suas subdivisões, retorna aquelas seqüências depositadas com maior

homologia. Assim, várias regiões de DNA podem ser anotadas através do *BLAST*, atribuindo uma função a qualquer segmento de DNA sequenciado que apresenta homologia significativa a outras seqüências previamente depositadas no *GenBank* com função conhecida (SANTOS; ORTEGA, 2003).

Os métodos de bioinformática estão sendo utilizados para fixar o conhecimento biológico, comparação dos conjuntos de genes de diferentes espécies, atribuição de genes de vias biológicas, identificação de fatores de especificidade, e descrevendo conjuntos de proteínas altamente conservadas comum a todos os domínios da vida e, assim, podendo obter respostas para perguntas detalhadas sobre o conjunto completo de genes de um organismo (ANDRADE; SANDER, 1997).

Nos bancos de dados há também uma grande variedade de informações sobre estruturas moleculares, expressão gênica diferencial, diversidade genética, evolução, etc. que podem ser extraídas pela bioinformática (SANTOS; ORTEGA, 2003).

## 2 METODOLOGIA

Os experimentos foram desenvolvidos no Laboratório de Genética Molecular do Departamento de Biologia da Universidade Federal do Ceará. Para a geração das ESTs, foram utilizados clones de bibliotecas de cDNA, construídas no vetor pCMV-SPORT 6 (Invitrogen) cedidas pelo Laboratório de Biodiversidade Molecular e Citogenética do Departamento de Genética e Evolução da Universidade Federal de São Carlos (São Carlos, SP), onde se utilizou 50 amostras de *L. vannamei*, pertencentes a diversas populações cativas de diferentes centros de cultivo do país.

A nomeação das ESTs foi feita de acordo com os tecidos a partir do qual cada cDNA foi extraído: biblioteca EEG para os ovos, NP para diferentes estágios larvais, MC para o músculo, PO para o pendúculo ocular e HP para o hepatopâncreas.

Os materiais e equipamentos utilizados para a execução dos experimentos, de acordo com as exigências de cada protocolo utilizado, foram os seguintes: meio Circle Grow (MP Biomedicals, Saln, OH, USA), *Shaker* CERTOMAT IS (Sartorius AG, Goettingen, Germany), placas de 96 poços de fundo V e de fundo U (Axygen Scientific, Union City, CA, USA), placas de filtro (Millipore, Billerica, MA, USA), microplacas de 96 poços para PCR (Axygen Scientific), centrífuga refrigerada com rotor para placas (Eppendorf Centrifuge 5804R), termocicladores PTC200 (DNA *Thermal Cycler*, MJ Research, Waltham, MA, USA) e *Eppendorf Mastercycler gradient* (Eppendorf, Hamburg, USA), DYEnamic ET Dye *Terminator Cycle Sequencing Kit* (GE Healthcare Bio-Sciences, Uppsala, Suécia) e Seqüenciador MEGABACE1000 DNA *Sequencing System* (Amersham Bioscience, Uppsala, Suécia).

### 2.1 Extração de plasmídeo

Os inóculos foram feitos a partir de 12 µL de cultura estoque em glicerol de *Escherichia coli* contendo plasmídeos recombinantes pCMV-SPORT 6 (Invitrogen), com o auxílio de uma pipeta multicanal (P100), em blocos de cultura de 96 poços contendo 1,0 mL/poço de meio *Circle Grow* (MP Biomedicals, Saln, OH, USA) suplementado com

ampicilina 100 µg/mL. As culturas foram incubadas em *Shaker* sob agitação constante de 220 rpm a uma temperatura de 37 °C por um período de 24 horas. As extrações de plasmídeo foram realizadas usando-se o método de lise alcalina (SAMBROOK *et al.*, 1989), adaptado para o uso de placas de 96 poços.

Os *pelletes* de DNA, ao final de cada preparação, foram resuspendidos em 50 µL de água ultra-pura (MilliQ) estéril. As amostras foram cobertas com adesivos devidamente etiquetados, ficaram a temperatura ambiente “*overnight*” e, posteriormente, foram estocadas a -20 °C para uso posterior.

## 2.2 Eletroforese em gel agarose

A eletroforese em gel de agarose foi utilizada para confirmar a extração do DNA e sua qualidade. Os géis de agarose foram preparados de acordo com Sambrook *et al.* (1989) para 0,8% de agarose em solução-tampão (TBE). As amostras de DNA foram misturadas com Tampão de Amostra (azul de bromofenol 0,25%, glicerol 30% em TE pH 8,0) e as corridas eletroforéticas foram realizadas em cuba horizontal sob voltagem constante de 100 V, com o gel submerso em tampão de corrida TBE (Tris-Borato 45 mM, pH 8,0, contendo EDTA 1 mM). Para cada placa foram escolhidas aleatoriamente 24 amostras para análise eletroforética. O marcador utilizado foi ( $\lambda$ -DNA/*HIND* III *Fragments*, Invitrogen) que possui um alto peso molecular.

## 2.3 Sequenciamento

As reações de sequenciamento foram realizadas a partir de uma solução final de 10 µL, contendo 1 µL de primer universal M13 forward -20 (5'-GTTTCCCAGTCACGACGTTG-3') de uma solução 5 µM para uma concentração final de 5 pmoles, 4 µL do pré-mix de sequenciamento (DYEnamic ET Dye *Terminator Cycle Sequencing Kit*, GE Healthcare Bio-Sciences, Uppsala, Suécia) e 5 µL de DNA plasmidial purificado. O programa de amplificação possuía 24 ciclos, cada ciclo consistindo de uma

etapa de desnaturação (95°C por 20 segundos), uma de anelamento (50°C por 15 segundos) e uma de extensão (60°C durante 1 min), permanecendo armazenada, após o último ciclo, a 4°C até serem purificadas. As reações de sequenciamento foram realizadas em termocicladores Eppendorf e PTC200.

A precipitação dos produtos das reações foi realizada com adição de 1 uL de acetato de amônio 7,5 M a cada poço, em seguida a placa foi centrifugada a 4.000 rpm por 50 minutos a 4°C. Posteriormente, os pellets foram lavados com 30 µL de isopropanol 80% e uma nova centrifugação foi realizada. Após lavagem dos pellets com 45 uL de etanol 70%, a placa foi novamente centrifugada a 4.000 rpm por 10 minutos a 4°C. Finalmente, o DNA foi ressuspensionado em 10 µL de loading solution, uma solução de formamida 70% contendo EDTA 1 mM fornecida pelo fabricante. As placas foram levemente homogeneizadas em vortex por 10 segundos e, em seguida foi dada uma rápida centrifugação, para que os *pelletes* precipitassem no fundo das placas, antes de ser levado ao seqüenciamento.

As placas foram submetidas à eletroforese capilar no seqüenciador automático MegaBACE1000 DNA *Sequencing System* (Amersham Biosciences), utilizando matriz de poliacrilamida (MegaBACE Long Read Matrix, GE Healthcare), foram utilizados vários parâmetros de injeção da amostra e parâmetros de corrida até a otimização da voltagem e tempo adequados para o seqüenciamento.

## 2.4 Bioinformática

Após o seqüenciamento os arquivos gerados possuem extensão “.rsd”. Estes arquivos foram convertidos para a extensão “.scf” e “.esd” pelo programa *Sequence Analyzer*, fornecido pelo fabricante juntamente com seqüenciador automático MegaBACE1000.

### 2.4.1 Análise de qualidade

Os eletroferogramas gerados no sequenciamento dos clones de cDNA no formato “.scf” foram analisados pelo programa *CodonCode* (MCCULLOUGH *et al.*, 2007). As ESTs

analisadas foram selecionadas de acordo com a qualidade apresentada pelo programa para a montagem das placas. Todas as sequências das 5 bibliotecas diferentes foram analisadas. O programa também foi utilizado para identificar as sequências de vetor que foram eliminadas manualmente. Após as montagens das placas pelo programa *CodonCode* as sequências foram salvas em formato FASTA.

Após as montagens das placas as *contigs* foram construídas usando o programa *CAP3 Sequence Assembly* (HUANG; MADAN, 1999) para 3643 ESTs a partir de tecidos de *L. vannamei*.

#### 2.4.2 Identificação das ESTs

As sequências foram comparadas com outras sequências depositadas no *GenBank* (<http://www.ncbi.nlm.nih.gov>) para a identificação das ESTs através do BLAST (*Basic Local Alignment Search Tool*) (ALTSCHUL *et al.*, 1990). O BLAST compara as sequências de interesse com todas as sequências existentes no banco de dados, buscando similaridades e realizando alinhamentos. A ferramenta utilizada foi o *BlastX* que compara as sequências de nucleotídeos com sequências de proteínas depositadas no banco de dados.

#### 2.4.3 Caracterização das ESTs

As ESTs foram classificadas de acordo com as seguintes categorias: proteínas de função conhecida, proteínas hipotéticas, ESTs sem similaridade e proteínas pouco caracterizadas. As ESTs foram identificadas levando em consideração o seu *e-value*, que é a probabilidade da qual uma sequência possa ou não ter obtido similaridade com uma sequência existente no banco de dados, indicando assim que houve um alinhamento real ou foi por mero acaso. O valor de *e-value* próximo a zero significa maior confiabilidade de o alinhamento ter sido real. A identidade, que se referir a um valor máximo de igualdade que as sequências apresentam, também foi avaliada para a caracterização de proteínas. Outro programa de caracterização foi abordado para classificar as ESTs, o programa *on line* utilizado foi o “*GOanna*” do site “*The Gene Ontology*” (<http://www.geneontology.org/>). As ESTs foram



submetidas ao *BlastX* do programa, que enviou por email todas as classificações possíveis para cada EST, sendo utilizado a de maior confiabilidade, ou seja, a primeira classificação da lista, para a caracterização.

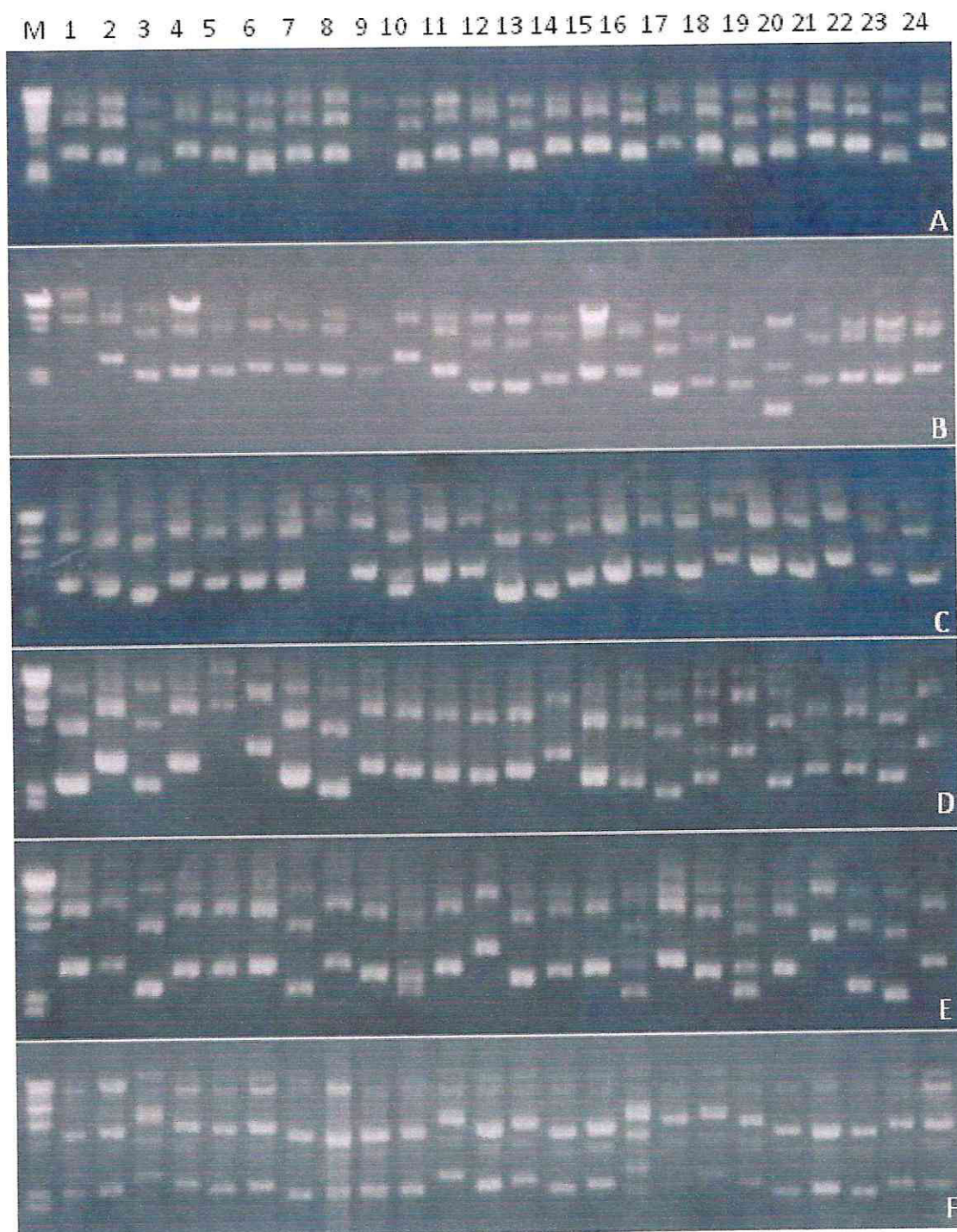
### 3 RESULTADOS E DISCUSSÃO

#### 3.1 Extração de plasmídeo

As culturas estoque em glicerol de *Escherichia coli* contendo plasmídeos recombinantes pCMV·SPORT 6 (Invitrogen) obtiveram confirmação da extração de plasmídeo através de eletroforese em gel de agarose 0,8%, onde se verifica a presença de fragmentos de três tamanhos diferentes (Figuras 7). Estes fragmentos correspondem às três formas em que o plasmídeo pode ser encontrado, na forma linear, helicoidal ou circular. Devido as suas conformações diferenciadas cada uma dessas formas obtém uma disposição individual durante a migração do gel de agarose. Com o marcador  $\lambda$ -DNA/ *Hind* III *Fragments* o número de pares de bases pode ser analisado.

As bandas no gel se apresentaram de maneira uniforme, indicando que a extração pode ser considerada foi bem sucedida, e que a quantidade e qualidade de DNA extraído em cada clone foram satisfatórias. A concentração de DNA obtido da extração pode ser estimada comparando a banda referente ao plasmídeo com as bandas do marcador, de tamanhos previamente conhecidos.

Após a confirmação da extração foram feitas várias reações de seqüenciamento até a otimização do mesmo e, em seguida, as sequências obtidas foram submetidas à análise de qualidade.



**Figura 7** - Eletroforese em gel de agarose 0,8% dos plasmídeos purificados das bibliotecas MC2078 (A), MC2079 (B), MC2080 (C), MC2081 (D) e MC2082 (E), MC2084 (F). O poço M contém 450ng do marcador  $\lambda$ -DNA/ *Hind* III Fragments (Invitrogen). Os demais contém 3  $\mu$ L de plasmídeo purificado selecionados aleatoriamente. Fonte: Protocolos do Projeto Genoma EST do camarão marinho *L. Vannamei* (2006).

### 3.2 Análise de qualidade

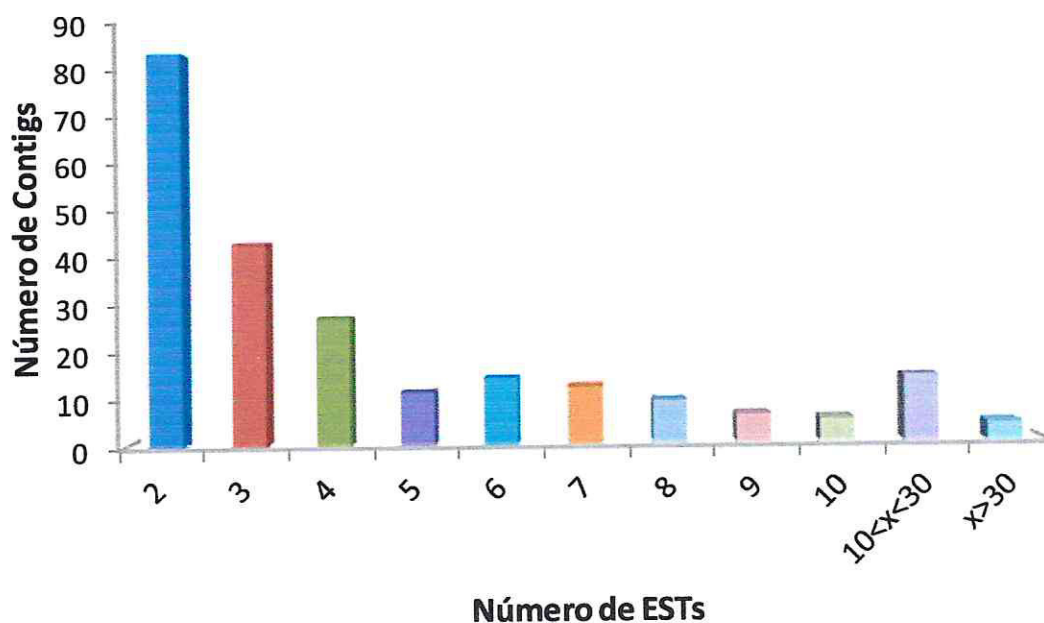
Todas as sequências foram submetidas à análise de qualidade de suas bases (*base calling*), efetuada por meio do programa *CodonCode Aligner*, que combina, entre outros, os algoritmos *Phred*, *Phrap* e *ClustalW* e que avaliou a qualidade dos picos correspondentes a cada base seqüenciada. A qualidade é dada em valor de bases de maior confiabilidade em relação ao número de bases seqüenciadas. Das repetições de sequenciamento utilizadas de cada placa foi selecionada a que apresentava maior qualidade, sendo as demais descartadas. O programa também foi utilizado para identificar as sequências de vetor que foram eliminadas manualmente.

Foi obtido um total de 4128 ESTs das quais 88,25% apresentaram sequências de boa qualidade (Tabela 2). Após o alinhamento pelo programa *Sequence Assembly* foram obtidas 245 *contigs* (Figura 10) e 609 *singletons* válidas para a caracterização.

**Tabela 2** – Resultados da análise individual e conjunta da qualidade das bases e montagem das ESTs das bibliotecas

	EGG	NP	MC	HP	PO	Total*
<b>Total de clones submetidos ao sequenciamento</b>	1152	1248	1248	384	96	4128
<b>Total de ESTs válidas</b>	1065	1164	997	324	93	3643
<b>Eficiência do sequenciamento (%)</b>	92,45	93,27	79,89	84,38	96,88	88,25
<b>Contigs formadas</b>	75	78	67	19	06	245
<b>Singletons</b>	115	143	182	101	68	609

\* Número total de todas a bibliotecas



**Figura 8** - Número de sobreposições de ESTs para formar *contigs*. Ao total foram 245 *contigs* formadas, sendo que maioria foi composta por sobreposição de duas ESTs.

### 3.3 Identificação das ESTs

Dos 854 unigenes (*contigs* + *singletons*) avaliadas pelo programa *GenBank* do NCBI, 190 ESTs pertenciam à biblioteca EGG, 221 à biblioteca NP, 249 à biblioteca MC, 120 à biblioteca HP e 74 eram provenientes à biblioteca PO. A ferramenta *BlastX* do NCBI foi utilizada para a identificação das 854 ESTs, comparando as sequências de nucleotídeos com sequências de proteínas identificadas no banco de dados de proteínas (<http://www.ncbi.nlm.nih.gov>). Para a confiabilidade de o alinhamento ter sido real o valor de *e-value* máximo utilizado foi de -5.

As ESTs de *contigs* e *singletons*, em todas as bibliotecas, se concentram, em maior porcentagem, entre as proteínas de função conhecida, as que não apresentaram similaridade com nenhuma sequência do banco de dados e as que apresentaram similaridade acima do limite (*e-value* > -5) (Figuras 11 e 12).

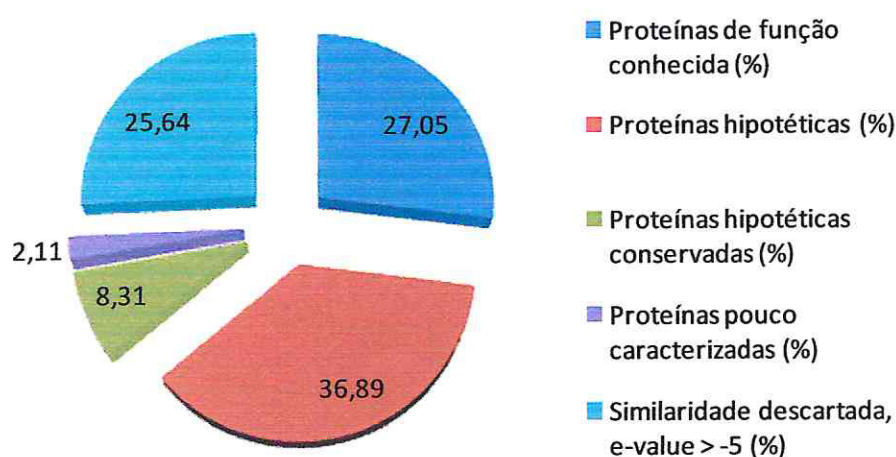
Em relação às demais características avaliadas, as *contigs* da biblioteca HP foram as que se apresentaram com maior porcentagem (10,5 %) de similaridade com proteínas hipotéticas, genes codificando para proteínas cuja atribuição funcional é inexistente, tipicamente derivados de projetos de sequenciamento em larga escala, necessitando de um aprofundamento de suas características funcionais (Tabela 3). Já para as ESTs de *singletons*, a que apresentou com maior porcentagem, dentro as demais características avaliadas, foi à biblioteca EGG para proteínas pouco caracterizadas com 31,4% (Tabela 4).

**Tabela 3.-.**Porcentagem da similaridade das sequências unigenes das bibliotecas com sequências já depositadas no *GenBank* (<http://www.ncbi.nlm.nih.gov/>), por meio da ferramenta *BlastX*

Unigenes	EGG	NP	MC	HP	PO	Total*
Unigenes	190	221	249	120	74	854
Proteínas de função conhecida (%)	33,68	26,24	34,94	8,33	14,86	27,05
Proteínas hipotéticas (%)	24,74	31,22	32,53	69,17	45,95	36,89
Proteínas hipotéticas conservadas (%)	16,84	9,50	5,62	1,67	9,46	8,31
Proteínas pouco caracterizadas (%)	4,74	1,81	2,81	0,00	4,05	2,11
Similaridade descartada, e-value > -5 (%)	20	31,22	24,10	20,83	25,68	25,64

\*Porcetagem total em relação a todas as 245 *contigs* das bibliotecas

**Porcentagem total da similaridade (*BlastX*) dos 854 unigenes**



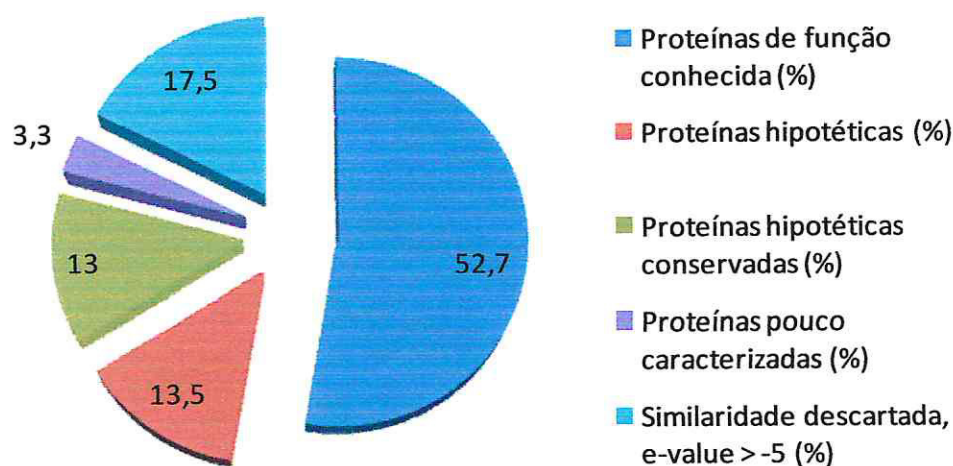
**Figura 9 -.**Porcentagem da similaridade das sequências unigenes com sequências já depositadas no *GenBank* (<http://www.ncbi.nlm.nih.gov/>), por meio da ferramenta *BlastX*. Um total de 315 unigenes codificaram proteínas hipotéticas.

**Tabela 4** - Porcentagem da similaridade das sequências *contigs* das bibliotecas com sequências já depositadas no *GenBank* (<http://www.ncbi.nlm.nih.gov/>), por meio da ferramenta *BlastX*

Contigs	EGG	NP	MC	HP	PO	Total*
<b>Contigs formados</b>	75	78	67	19	06	245
<b>Proteínas de função conhecida (%)</b>	45,3	51,3	65,6	52,6	16,7	52,7
<b>Proteínas hipotéticas (%)</b>	13,3	14,1	10,4	15,8	33,3	13,5
<b>Proteínas hipotéticas conservadas (%)</b>	18,8	15,4	6,0	10,5	0,0	13,0
<b>Proteínas pouco caracterizadas (%)</b>	2,6	2,6	3,0	0,0	33,3	3,3
<b>Similaridade descartada, e-value &gt; -5 (%)</b>	20,0	16,6	15,0	21,1	16,7	17,5

\*Porcentagem total em relação a todas as 245 *contigs* das bibliotecas

**Porcentagem total da similaridade (BlastX) das 245 *contigs***



**Figura 10** - Porcentagem da similaridade das sequências *contigs* das bibliotecas com sequências já depositadas no *GenBank* (<http://www.ncbi.nlm.nih.gov/>), por meio da ferramenta *BlastX*. Um total de 129 unigenes codificaram proteínas de função conhecida.

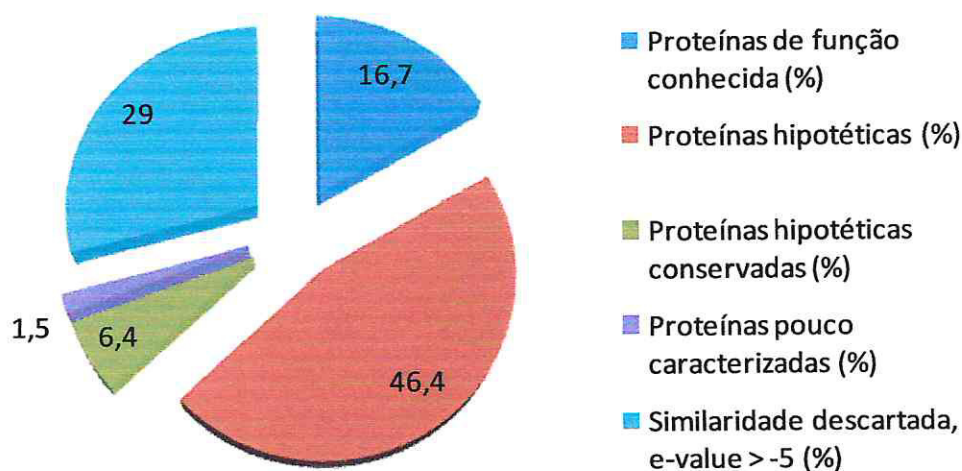


**Tabela 5** - Porcentagem da similaridade das sequências *singletons* das bibliotecas com sequências já depositadas no *GenBank* (<http://www.ncbi.nlm.nih.gov/>), por meio da ferramenta *BlastX*

Singletons	EGG	NP	MC	HP	PO	Total*
Singletons	115	143	182	101	68	609
Proteínas de função conhecida (%)	26,0	12,6	23,7	0,0	14,7	16,7
Proteínas hipotéticas (%)	32,15	40,5	40,7	79,2	47,1	46,4
Proteínas hipotéticas conservadas (%)	15,7	6,3	5,4	0,0	10,3	6,4
Proteínas pouco caracterizadas (%)	6,15	1,4	2,7	0,0	1,5	1,5
Similaridade descartada, e-value > -5 (%)	20	39,2	27,5	20,8	26,4	29,0

\*Porcetagem total em relação a todos os 605 *singletons* das bibliotecas

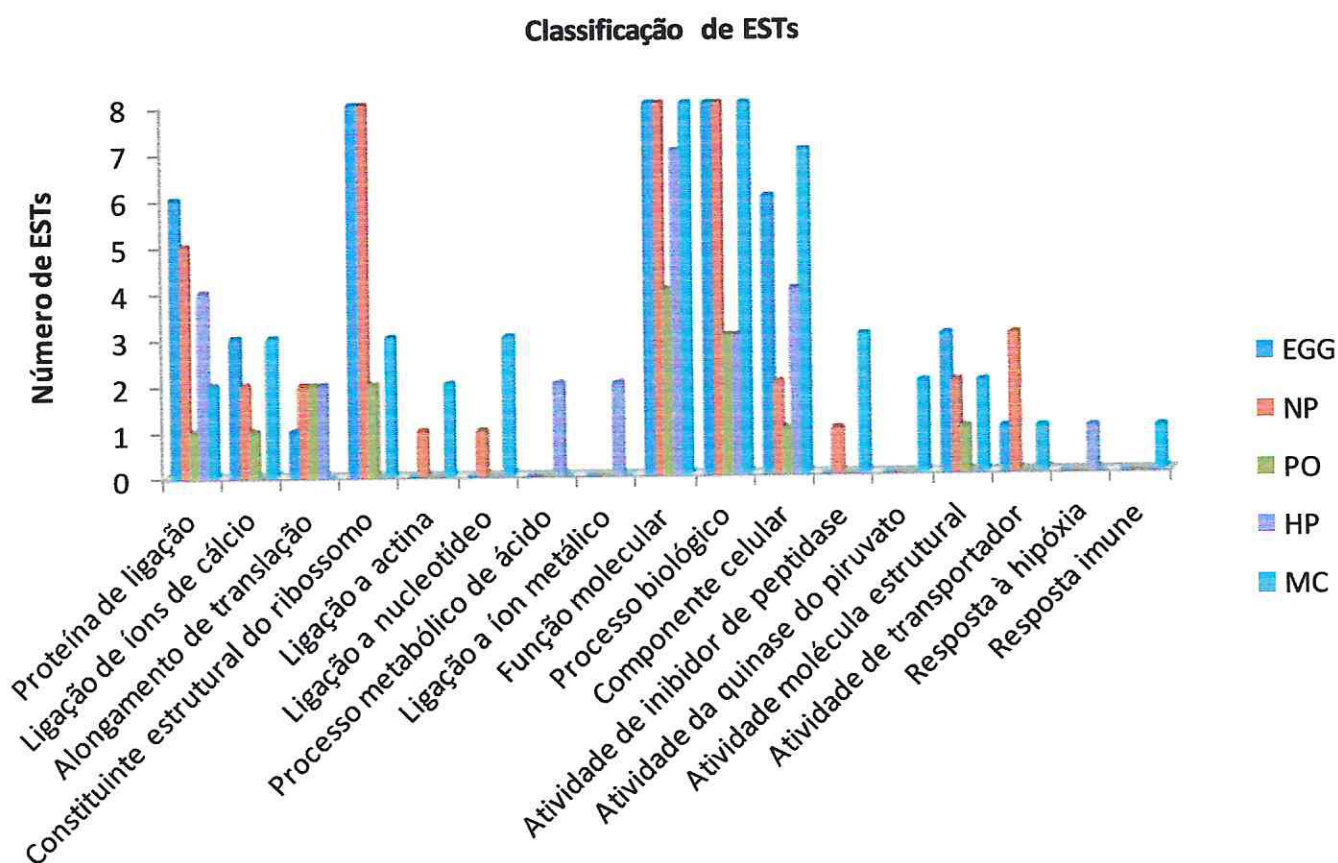
**Porcentagem total da similaridade (BlastX) dos 609 *singletons***



**Figura 11** - Porcentagem da similaridade das sequências *singletons* das bibliotecas com sequências já depositadas no *GenBank* (<http://www.ncbi.nlm.nih.gov/>), por meio da ferramenta *BlastX*. Um total de 282 unigenes codificaram proteínas hipotéticas.

### 3.4 Classificação das ESTs

A classificação das ESTs pelo programa “GOanna” (Figura 13) e as definições pelo programa “AmiGO” (Tabela 5) resultaram de acordo com as funções e atividades de cada biblioteca. As principais características e classificações das proteínas submetidas ao programa foram agrupadas para o estudo de acordo com cada tecido.



**Figura 12.-**Classificação de algumas funções das ESTs de acordo com o programa “GOanna” do site “The Gene Ontology”.

Tabela 6 – Definições de acordo com o programa on line “AmiGO” do site “The Gene Ontology”

Classificação	Definição
Proteína de ligação	Interagindo de forma seletiva e não-covalentemente com qualquer complexo protéico ou proteína.
Ligação de íons de cálcio	Interagindo de forma seletiva e não-covalentemente com íons cálcio (Ca <sup>2+</sup> ).
Alongamento de translação	A adição sucessiva de resíduos de aminoácidos de uma cadeia nascente de polipeptídeos durante a síntese protéica.
Constituinte estrutural do ribossomo	A ação de uma molécula que contribui para a integridade estrutural do ribossomo.
Ligação a actina	Interagindo de forma seletiva e não-covalentemente com formas monomérica ou multimérica de actina, incluindo filamentos de actina.
Ligação a nucleotídeo	Interagindo de forma seletiva e não-covalentemente com um nucleotídeo.
Processo metabólico de ácido	As reações químicas e caminhos envolvendo os ácidos biliares.
Ligação a íon metálico	Interagindo de forma seletiva e não-covalentemente com qualquer íon metálico.
Função molecular	Atividades elementares, tais como catálise ou ligação, descrevendo as ações de um produto gene ao nível molecular.
Processo biológico	Qualquer processo especificamente pertinentes para o funcionamento das unidades de vida integrado: células, tecidos, órgãos e organismos.
Componente celular	A parte de uma célula ou seu ambiente extracelular em que um produto gene está localizado.
Atividade de inibidor de peptidase	Pára, evita ou reduz a atividade de uma peptidase, qualquer enzima que catalisa a hidrólise ligações peptídicas.
Atividade da quinase do piruvato	Catálise da reação: ATP piruvato em ADP fosfoenolpiruvato.
Atividade molécula estrutural	Têm por ação contribuir para a integridade estrutural de uma molécula ou complexo, bem como a montagem do mesmo dentro ou fora de uma célula.
Atividade de transportador	Permite o movimento dirigido de substâncias (como macromoléculas, pequenas moléculas, íons) para fora ou dentro de uma célula, ou entre as células.
Resposta à hipóxia	Qualquer processo que resulta em uma mudança de estado ou atividade de uma célula ou um organismo como resultado de um estímulo, indicando baixa tensão de oxigênio.
Resposta imune	Qualquer processo do sistema imunológico que funciona na resposta calibrada de um organismo para uma ameaça potencial interno ou invasivo.

Na figura 13 estão expressas as funções e atividades mais representativas encontradas durante a classificação das ESTs de cada biblioteca. A biblioteca EGG (ovos), apresentou a maior quantidade de ESTs codificando proteínas de ligação, as quais têm por função interagir de forma seletiva e não-covalentemente com qualquer complexo protéico ou proteína. Também apresentou maior número de ESTs para atividade molecular estrutural, que têm por ação contribuir para a integridade estrutural de uma molécula ou complexo, bem como a montagem do mesmo dentro ou fora de uma célula.

Para a biblioteca NP (estágios larvais) foram identificadas proteínas com atividades de transportador que permite o movimento dirigido de substâncias (como macromoléculas, pequenas moléculas, íons) para fora ou dentro de uma célula, ou entre as células. Juntamente com a biblioteca MC (músculos), a NP tem genes codificantes para proteína de ligação a actina, que interage de forma seletiva e não-covalentemente com formas monomérica ou multimérica de actina, incluindo filamentos de actina; e proteínas de ligação a nucleotídeo, que interage de forma seletiva e não-covalentemente com um nucleotídeo, ou com composto qualquer consistindo de um nucleosídeo que é esterificado com (orto) fosfato ou um oligofosfato em qualquer grupo hidroxila na ribose ou desoxirribose.

Na biblioteca PO (Pedúnculo ocular), foi observada uma maior codificação de proteínas com funções moleculares, que realizam atividades elementares, tais como catálise ou ligação, descrevendo as ações de um produto gene ao nível molecular; e processo biológico, que são processos especificamente pertinentes para o funcionamento das unidades de vida integradas: células, tecidos, órgãos e organismos.

Genes que codificam proteínas envolvidas em processos metabólicos ácidos, que são reações químicas e caminhos envolvendo os ácidos biliares ou qualquer um dos grupos de esteróides ácidos carboxílicos para processos digestivos, e íon metálico de ligação, que interagem de forma seletiva e não-covalentemente com qualquer íon metálico, foram encontradas somente na biblioteca HP (hepatopâncreas), originária do tecido hepatopâncreas, que tem por função a digestão, absorção e armazenagem de nutrientes.

A biblioteca MC apresentou genes codificadores de proteínas ligadoras a íons cálcio, que interage de forma seletiva e não-covalentemente com íons cálcio ( $\text{Ca}^{2+}$ ); componente celular, parte de uma célula ou seu ambiente extracelular em que um produto do gene está localizado; atividade de inibidor de peptidase, que evita ou reduz a atividade de uma peptidase ou qualquer enzima que catalisa a hidrólise de ligações peptídicas; e atividade da quinase do piruvato, que catalisa a reação de ATP piruvato à ADP fosfoenolpiruvato.

Em 2009, Robalino *et al.* propõem que proteínas de ligação (*Binding protein*) estariam relacionadas com a resposta imune contra infecção do vírus WSSV. Esta classe de proteínas foi encontrada em vários tecidos, estando em menor proporção no pedúnculo ocular.

Na biblioteca MC também foi encontrado gene que obteve classificação para “Resposta imune”, que funciona na resposta calibrada de um organismo para uma ameaça potencial interno ou invasivo.

Na biblioteca HP foi encontrada uma proteína relacionada à “Resposta à hipoxia”, que resulta em uma mudança de estado ou atividade de uma célula ou um organismo (em termos de secreção, movimento, produção de enzimas, expressão gênica, etc), como resultado de um estímulo, indicando baixa tensão de oxigênio. Hipóxia, definida como um declínio nos níveis de O<sub>2</sub> abaixo dos níveis de normóxica 20,8-20,95%, resulta em adaptação metabólica, tanto a nível celular e do organismo.

### 3.5 Caracterização das ESTs

#### 3.5.1 Proteínas codificadas relacionadas com a defesa inata

Robalino *et al.* (2009) relataram a identificação de algumas proteínas codificadas, através do sequenciamento de cDNA de tecidos das brânquias e hematócitos do *L. vannamei*, que possuíam funções conhecidas ou presumidamente identificados como supostamente imune estimulada por meio de hibridação subtrativa de supressão (HSS). O HSS é um dos métodos mais antigos para análise de expressão gênica que permite comparar duas populações de mRNAs (RNA mensageiro), mas que apresenta pouca sensibilidade e precisa de grandes quantidades de iniciais de RNA (DIATCHENKO *et al.*, 1996). Dentre os genes codificados estão as Histona H2A variação Z, Histona H4, Tachylectina-5A hipertérmica, Fator anti-LPS, Domínio imunoglobulínico, Inibidor de protease serínica, Quinase I-kappa-B, Crescimento Derivado de Plaquetas fator-like, Catepsina A, B e L, Lisozima, Lectina tipo C, Quitinases, Lacase, Proteínas ricas em repetições de leucina. Das ESTs seqüenciadas do banco de cDNA cedidas pela Universidade Federal de São Carlos (São Carlos, SP) houve a codificação de proteínas relacionados com Histona H2A-VI, Catepsina L e Lectina tipo C nas bibliotecas NP, EGG e HP, respectivamente (Tabela 6). O banco de cDNA foi feito com amostras de *L. vannamei* de vários centros de cultivos não havendo um padrão de estresse e infecção, e assim, os resultados sugerem que a codificação dessas proteínas se dá pela defesa inata do organismo perante microorganismos patogênicos presentes no ambiente de cultivo.

**Tabela 7** -. Algumas proteínas codificadas com conhecida ou presumida função imunológicas no camarão marinho *L. vannamei* segundo Robalino *et al.* em 2009

Identificação	Biblioteca	Indutor	Blastx	Tamanho	e - value
Histona H2A-VI	NP	Ação microbiana	<a href="#">emb CBM82441.1 </a> <i>Rhabdopleura compacta</i>	513	9e-39
Catepsina L	EKG	Infecção pelo WSSV	<a href="#">gb ABQ10739.1 </a> <i>Penaeus monodon</i>	367	8e-45
Lectina tipo C	HP	Infecção pelo WSSV e/ou Resistência	<a href="#">gb ABU62825.1 </a> <i>Litopenaeus vannamei</i>	746	2e-12

Em estudos feitos por Robalino *et al.* (2009) as histonas foram identificadas entre as proteínas cuja abundância aumenta nas brânquias com a infecção do WSSV, obtendo um interesse particular. Sinais de proteína de histonas H2A, H2B, e H4 foram observados unicamente associado com as amostras infectadas. Para Robalino *et al.* (2009) as histonas, em *L. vannamei*, têm função em resposta à estimulação de estresse, imune ou infecção, além de seus papéis em função da cromatina. Sendo que, ainda assim, precisa ser elucidado se histonas possuem atividade antiviral em si ou se elas regulam os processos que são relevantes para a montagem de uma resposta antiviral no camarão.

Robalino *et al.* (2007) sugeriram que várias proteases lisossomais (por exemplo, catepsinas) são de especial interesse, porque elas estão presentes em bibliotecas para transcrições induzidas pela infecção por WSSV e por dsRNA, um indutor de imunidade antiviral em camarão. No mesmo ano Zhao *et al.*, relatam que a Catepsina L é uma protease cisteínica onipresente, existente em todos os tecidos de mamíferos e desempenha um papel importante na degradação de proteínas exógenas e endógenas no lisossomo. Segundo Robalino *et al.*, 2007, em termos de resposta imune, a melhor descrição para funções de catepsinas são relacionados ao processamento de peptídeos durante a apresentação de antígenos em vertebrados. Zhao *et al.*, 2007, relataram que um aumento na Catepsina L é crucial para o processo proteolítico de antígenos sujeitos a endocitose e para a apresentação dos antígenos ao sistema imunológico dos mamíferos. Em invertebrados, proteases cisteína representam uma componente importante do sistema proteolítico lisossomal, e são responsáveis pela degradação da proteína intracelular. Em seus estudos, a Catepsina L foi altamente expressa durante a infecção do WSSV,

sugerindo que ela está envolvida na resposta de defesa contra o vírus. Contudo, para Robalino *et al.* em 2007, ainda são necessários estudos futuros e métodos de alta capacidade de perfil de expressão para confirmar regulação diferencial da maioria desses genes.

Para Zhao *et al.* (2007) o reconhecimento de microorganismos invasores é a primeira reação imune, que é mediada pelo hemócitos e proteínas plasmáticas. Invertebrados têm desenvolvido várias moléculas de superfície em reconhecer os determinantes potenciais agentes patogénicos. As moléculas podem interagir com o reconhecimento e ativar hemócitos, que desempenham um papel importante e central na defesa do hospedeiro. Devido ao reconhecimento específico de determinantes de açúcar na parede celular ou a cápsula de patógenos, as lectinas têm sido consideradas efetores da imunidade inata, o que pode induzir a aglutinação bacteriana, ou agir como opsoninas para aumentar a fagocitose de microorganismos por hemócitos. Assim a expressão das lectinas que podem atuar como receptores no reconhecimento de padrões pode ser desencadeada pelo contato com o vírus, que por sua vez ativa cascatas enzimáticas e induz opsonização e fagocitose contra o patógeno.

Robalino *et al.* (2009) relataram que alguns genes codificaram peptídeos que possuíam funções conhecidas, ou presumida, funções imunológicas perante a infecção do vírus WSSV, tais como: Beta-Actina, H<sup>+</sup> ATPase, Proteína sarcoplasmático de ligação a cálcio, Quinase arginina, ácido retinóico celular/retinol, Proteínas de ligação, Crustacianina, Cofilina/ADF, Beta tubulina. Dentre as codificadas pelas bibliotecas seqüenciadas então: Proteína sarcoplasmática de ligação a cálcio, Actina A1, Crustacianina subunidade A2, Beta-1 tubulina, Alfa-2 tubulina (Tabela 7).



**Tabela 8** - Alguns peptídeos codificados com conhecida ou presumida função imunológicas contra a infecção WSSV no camarão marinho *L. vannamei* segundo Robalino *et al.* em 2009

<b>Identificação</b>	<b>Biblioteca</b>	<b>Blastx</b>	<b>Tamanho</b>	<b>e - value</b>
Proteína sarco-plasmática de ligação a cálcio	MC	<u>sp P02635.1 SCPB_PENSP</u> <i>Penaeus sp.</i>	892	1e-140
Actina A1	NP	<u>gb AAF81598.1 AF277456_1</u> <i>Papilio glaucus</i>	371	3e-25
	MC	<u>gb AAC78682.1 </u> <i>Penaeus monodon</i>	1217	4e-145
Crustacianina subunidade A2	EGG	<u>sp P80007.1 CRA2_HOMGA</u>	399	1e-38
	NP	<i>Homarus gammaru</i>	832	2e-49
Beta-1 tubulina	PO		449	5e-07
	EGG	<u>sp Q25009.1 TBB1_HOMAM</u> <i>Homarus americanus</i>	532	8e-85
Alfa-2 tubulina	MC	<u>gb AAC47523.1 </u> <i>Gecarcinus lateralis</i>	378	2e-10

Outras proteínas foram encontradas que codificam funções imunes, tais como Ferritina, Proteína QM, Selenoproteína, Hemocianina e Tiorredoxina (Tabela 8).

**Tabela 9** – Outras proteínas codificadas de função imunológica.

<b>Identificação</b>	<b>Biblioteca</b>	<b>Blastx</b>	<b>Tamanho</b>	<b>e - value</b>
Ferritina	EGG	<u>gb AAX55641.1 </u> <i>Litopenaeus vannamei</i>	592	4e-27
Proteína QM	EGG	<u>gb ACV72062.1 </u>	483	9e-77
	PO	<i>Penaeus monodon</i>	616	9e-52
Selenoproteína	EGG	<u>ref NP_840072.3 </u> <i>Danio rerio</i>	530	2e-27
	HP	<u>emb CAA57880.1 </u> <i>Litopenaeus vannamei</i>	442	2e-53
Tiorredoxina	NP	<u>Gb AEI25985.1 </u> <i>Penaeus monodon</i>	562	9e-56

Ruan *et al.* (2010) diz que a ferritina foi identificada em uma ampla gama de organismos procariontes e eucariontes. Em sistemas de vertebrados, ferritina é conhecida por

desempenhar um papel importante na desintoxicação e de resistência contra o estresse oxidativo e também está envolvida em processos imune. Em invertebrados, Karlsson *et al.* (2004) mostrou que a ferritina plasmática em *Drosophila* é regulada mediante a coagulação do sangue, além de ser considerado um elemento importante no sistema imunológico única de crustáceos.

Segundo o Xu *et al.* (2008) a proteína QM foi proposta para ser um supressor de tumor, em 1991, e assim atraído estudos intensivos em plantas, animais e fungos. Contudo, até a data a função da proteína QM permanece desconhecida. O trabalho sugeriu que a proteína QM estava envolvida na imunidade do camarão, pois a proteína QM interagiu com a hemocianina, indicando que a proteína poderá participar na oxidase *prophenol* (proPO) sistema de ativação da imunidade de camarão, podendo regular a atividade de oxidase de fenol, uma enzima-chave no sistema de ativação de imunidade proPO invertebrados.

Para Hwang *et al.* (2008) selenoproteína está associada com uma variedade de doenças graves, incluindo doenças infecciosas, doenças neurodegenerativas, câncer e doenças cardiovasculares. Pode funcionar como uma oxidoredutase tiol dissulfeto, que participa na formação da ligação dissulfeto e servir para a manutenção de um elevado nível de estado antioxidante. Segundo Garcia-Triana *et al.* (2010) a selenoproteína M é uma selenocisteína contendo proteínas com atividade redox envolvidas na resposta antioxidante. No camarão branco *L. vannamei*, a expressão de selenoproteína M nas brânquias é induzida transitoriamente durante a infecção viral pela Síndrome do Vírus da Mancha Branca (WSSV). Tais autores ainda relataram que a expressão de selenoproteína M foi detectada em camarão *L. vannamei* saudável nas brânquias, hepatopâncreas, músculo e pleópodos, com mais abundância no hepatopâncreas e brânquias. Peroxidases não-heme são os novos jogadores no sistema de estresse oxidativo, que precisam ser abordadas em detalhe, bem como o selênio como um micronutriente essencial para o antioxidante e sistema imunológico inato em crustáceos. O primeiro relato da expressão de selenoproteína M em um tecido específico em crustáceos foi em 2010 por Garcia-Triana *et al.*

A hemocianina é o principal componente protéico da hemolinfa, que tem como função primária transportar oxigênio, mas também está envolvida em outros processos fisiológicos como a osmorregulação e atividades enzimáticas (ZHANG *et al.*, 2004). Lei *et al.* (2008) afirmam que a hemocianina, além de seu papel como transportadora de oxigênio, parece também envolvida na defesa de invasores, incluindo bactérias e fungos. No entanto, estudos envolvidos em sua atividade antiviral continuam limitados. Já foi relatado que o complexo da

proteína nativa de hemocianina poderia atrasar a infecção pelo vírus WSSV *in vivo* em um nível relativamente elevado de concentração, que é semelhante à concentração de hemocianina em camarões resistentes à WSSV.

Tiorredoxina (TRX) é a principal componente da maquinaria redox na regulação da homeostase na célula e é necessária para a função ribonucleotídeo redutase, entre outras. Em invertebrados, o equilíbrio redox é comprometido durante doenças e mudanças no estado fisiológico e é um dos componentes da resposta imune inata. Em um estudo sobre TRX em *L. vannamei* (LvTRX) esta proteína foi capaz de reduzir dissulfetos de insulina e foi o melhor antioxidante em comparação com glutathiona reduzida e ácido ascórbico, por meio do ensaio da *Trolox Equivalent Antioxidant Capacity* (TEAC), Capacidade Antioxidante Equivalente Trolox. LvTRX se destaca como um importante antioxidante que deve ser considerado em futuros estudos fisiológicos e desafios imune (AISPURO-HERNANDEZ *et al.*, 2008).

### 3.5.2 Proteínas codificadas relacionadas com o estresse

Algumas proteínas codificadas com funções relacionadas a estresse foram encontradas, tais como proteínas de choque térmico 90, Beta-adaptina e Metalotioneína (Tabela 9).

**Tabela 10** – Proteínas codificadas com função relacionada em resposta ao estresse

Identificação	Biblioteca	Blastx	Tamanho	e - value
Proteínas de choque térmico 90	EGG	<a href="#">gb ADU03767.1 </a> <i>Litopenaeus vannamei</i>	464	3e-15
Beta-adaptina	PO	<a href="#">gb AAQ86830.1 </a> <i>Ixodes scapularis</i>	452	4e-52
Metalotioneína	HP	<a href="#">gb ADQ28316.1 </a> <i>Penaeus monodon</i>	370	4e-07

Proteínas de choque térmico (HSP) compõem uma família de proteínas expressas em resposta a uma ampla gama de fatores de estresse biótico e abiótico, sendo assim conhecidas também como proteínas de estresse. Sua expressão em peixes tem sido descrita em linhas celulares, culturas primárias de várias células e nos tecidos de organismos inteiros, e estudos mostram que a expressão de HSPs é afetada em uma ampla variedade de células e tecido de peixe, em resposta tanto a estresses biológicos, tais como agentes infecciosos, bem como a estresses abióticos como o calor e choque térmico, e contaminantes ambientais (IWAMA *et al*, em 1998 ).

Estudos recentes têm mostrado uma relação entre a expressão de HSP e a resposta ao estresse generalizado em peixes, mas são necessárias mais pesquisas para esclarecer as relações complexas entre os hormônios do estresse e da resposta HSP celular. Em geral, a resposta HSP parece estar relacionada com a detecção do estresse e os efeitos celulares subsequentes que podem auxiliar na adaptação das células ao estresse (IWAMA *et al*, 1998 ).

Receptores scavenger (SR) internalizam partículas quimicamente modificadas de lipoproteína de baixa densidade (LDL-ac) e outros ligantes através do processo de endocitose

mediada pelo receptor. Durante uma investigação com b-amilóide como um ligante natural para o SR, sob uma condição de estresse oxidativo, houve experimentos de expressão diferencial de b-adaptina, a-adaptina, SR-AI, e SR-BI em células RAW (macrófagos) e em células EOC (microglia). Os resultados mostram que, de acordo com a dosagem, b-amilóide produz um estado de estresse oxidativo, que afeta principalmente a disponibilidade de b-adaptina. Sob essas condições, ensaios de RT-PCR (reação em cadeia da polimerase em tempo real) mostraram que o mRNA da b-adaptin é normalmente sintetizado, razão pela qual a tradução proteica ou estrutura da proteína de b-adaptin pode ser alterada (MANZANO-LEÓN *et al.*, 2006).

Metalotioneína (MT) foi encontrada no hepatopâncreas de camarão. Esta proteína de baixo peso molecular (6-7 kd) rica em cisteína, cuja síntese é induzida por uma ampla gama de compostos, incluindo metais pesados (Cd, Zn, Cu e), agentes alquilantes, glicocorticóides, citocinas, e pelo estresse oxidativo. Estudos *in vitro* mostraram que MT pode participar de desintoxicação metal, limpeza de radicais livres, resistência aos agentes anticâncer e alquilantes, bem como radiação ionizante. Além de seu papel central na regulação de metal essencial, nomeadamente para Zn e Cu, funções fisiológicas da MT ainda permanecem obscuros (MOFFATT *et al.*, 1995).

### 3.5.3 Codificação de outras proteínas

A tropomiosina e a cadeia leve da miosina foram encontradas em músculos e náuplios do *L. vannamei*. Até recentemente, a tropomiosina, uma proteína muscular, foi o principal alérgeno identificado em diferentes espécies de camarão. No entanto, temos recentemente a descrição da cadeia leve de miosina (MLC), como uma nova proteína alergênica de camarão (AYUSO, *et al.*, 2009).

A oxidase do citocromo c foi encontrada em todos os tecidos. Esta é uma proteína integrante da membrana mitocondrial interna de eucariotos e membrana plasmática de alguns procariontes. Ela desempenha um papel fundamental na vida aeróbia, com 90% de todo o oxigênio respirado reduzido a água por esta enzima. Em eucariotos, a reação catalisada pelo citocromo c oxidase é o estágio final da cadeia mitocondrial de transporte de elétrons. A função única e complexa da enzima é para acoplar a oxidação de redução do citocromo c, um doador de elétrons único, para a redução do oxigênio molecular à água, uma reação de quatro elétrons (BLENKINSOP; AITKEN; WILSON, 1996).

Uma proteína ribossomal é qualquer uma das proteínas que, em conjunto com rRNA, compõem a subunidades ribossomal envolvidas no processo celular de tradução. Estas proteínas foram amplamente encontradas em ovos, náuplios, músculos e no pedúnculo ocular. A montagem das subunidades ribossomais é um processo multi-estágio complicada que envolve recursos significativos celulares (ILIN; MALYGIN; KARPOVA, 2011). Um dos objetivos principais da pesquisa sobre o ribossomo é entender o papel funcional de proteínas constituintes no controle e regulação da síntese protéica celular. Proteínas ribossomais muitas vezes exercem funções extraribossomal, sugerindo papéis reguladores e sinalização para eles em várias funções celulares (YADAIHAH *et al.*, 2012).

NADH desidrogenase desempenha o seu papel mais importante como uma desidrogenase primária, relacionada com o metabolismo central, na cadeia respiratória de todos os organismos tendo um aeróbico ou anaeróbico sistema de transporte de elétrons. Genes codificadores de NADH desidrogenase foram encontrados em ovos e músculos. As mitocôndrias de mamíferos têm somente uma NADH desidrogenase, que tem uma capacidade de bombeamento de prótons. No entanto, as membranas citoplasmática de quase todos os procariontes e também as membranas mitocondriais de alguns eucariontes, incluindo fungos e plantas, também

têm um NADH desidrogenase alternativa, que não tem qualquer capacidade de geração de energia (MATSUSHITA *et al.*, 2001).

Gliceraldeído-3-fosfato desidrogenase catalisa a formação de 1,3-difosfoglicerato de gliceraldeído 3-fosfato na via glicolítica. Estas enzimas foram encontradas no músculo que necessita de muita energia em processos de contração para locomoção. Gliceraldeído-3-fosfato desidrogenase (GAPDH), foi considerada uma proteína glicolítica clássica por seu papel central na produção de energia. Também foi usado como uma proteína modelo de análise da estrutura da proteína e os mecanismos enzimáticos. O gene GAPDH foi utilizado como protótipo para estudos genéticos de organização, expressão e regulação (SIROVER, 1999).

A expressão de poliubiquitina, bem como os níveis de conjugados de ubiquitina, uma enzima ubiquitina-conjugando envolvidos na via proteolítica ATP / ubiquitina-dependente, e proteassoma, foram examinadas para determinar o papel do ATP / ubiquitina dependente da proteólise na degradação de proteínas melhorado miofibrilar durante a atrofia muscular. A poliubiquitina é detectada em todas as fases do ciclo de intermuda (SHEAN; MYKLES, 1995). Tais autores afirmam que os resultados sugerem que a via ATP / ubiquitina-dependente proteolítica contribui para as mudanças no metabolismo das proteínas que ocorrem durante a muda induzida por atrofia muscular. Os autores também relatam que o aumento de mRNA de poliubiquitina é consistente com a hipótese de que a expressão aumentada fornece monômeros adicionais necessários para a degradação acelerada através do caminho ATP / ubiquitina-dependente. Esta proteína foi encontrada na biblioteca NP, que constitui tecidos de vários estágios larvais de *L. Vannamei*.

#### 4 CONCLUSÃO

Dos 4128 clones submetidos ao sequenciamento foram obtidos um total de 3643 ESTs, apresentado uma eficiência de sequenciamento de 88,25%. Após a sobreposição das ESTs pelo programa *Sequence Assembly* foram obtidas 245 *contigs* (Figura 10) e 609 *singletons* válidas para a caracterização. Dos unigenes validados, foram identificadas, em diferentes tecidos, sequências que apresentaram similaridade com proteínas com funções já conhecidas, tais como proteínas relacionadas com o sistema imunológico, peptídeos antimicrobianos, proteínas relacionadas a estresse, bem como proteínas envolvidas em diferentes processos celulares vitais.

Os resultados encontrados nesse trabalho podem servir de base para a geração de hipóteses para orientar futuras pesquisas em crustáceo sobre interações hospedeiro-vírus, além de estudos sobre a interação do camarão marinho *L. vannamei* como o meio no qual vive.



## REFERÊNCIAS

ADAMS, M., *et al.*. Complementary DNA sequencing: expressed sequence tags and human genome project. **Science, New Series**, v. 252, n. 5013, p. 1651–1656, 1991.

AISPURO-HERNANDEZ, *et al.* Shrimp thioredoxin is a potent antioxidant protein. **Comparative Biochemistry and Physiology**, v. 148, p. 94–99, 2008.

ALTSCHUL, S.F.; GISH, W.; MILLER, W.; MYERS, E.W.; LIPMAN, D.J. Basic local alignment search tool. **Journal of Molecular Biology**. 215,403-410, 1990.

ANDRADE, A. A.; SANDER, C. Bioinformatics: from genome data to biological knowledge. **Biotechnology**, v. 8, p. 675-683, 1997.

ANDREW, D. R. A new view of insectecrustacean relationships II. Inferences from expressed sequence tags and comparisons with neural cladistics. **Arthropod Structure & Development**, v. 40, p. 289-302, 2011.

AYUSO, R., *et al.* Sarcoplasmic calcium-binding protein is an EF-hand-type protein identified as a new shrimp allergen. **Journal of Allergy and Clinical Immunology**, v. 124, p. 114-120, 2009.

BLENKINSOP, C.; AITKEN, A.E.; WILSON, M.T. Physical and functional characterisation of monomeric and dimeric eukaryotic cytochrome c oxidases. **Biochemistry and Molecular Biology**, v. 115, p. 421-428, 1996.

BROCK, J. A.; MAIN, K. L. A. Guide to the Common Problems and Diseases of cultured *Penaeus vannamei*. **The Oceanic Institute**, p. 242, 1994.

CAMARGO, S. G. O.; POUHEY, J. L. O. F. Aqüicultura - Um mercado em expansão. **Revista Brasileira de Agrociência**, Pelotas, v. 11, n. 4, p. 393-396, 2005.

CAVALLI, L. V.; MARINS, L. F.; NETTO, S.; ABREU, P. C. Evaluation of white spot syndrome virus (wssv) in wild shrimp after a major outbreak in shrimp farms at laguna, southern brazil. **Atlântica**, v. 30, n. 1, p. 45-52, 2008.

COELHO, M. A. S. Análise de custo/volume/lucro e investimentos em carcinicultura de pequeno porte. **Custos e Agronegócio**, Tamarineira, Recife-PE, v.1, n. 1, p. 63-64, 2005.

CHRISTIE, A.E., *et al.* Identification of putative crustacean neuropeptides using in silico analyses of publicly accessible expressed sequence tags. **General and Comparative Endocrinology**, v. 156, p. 246–264, 2008.

Departamento de Pesca y Acuicultura, **FAO**, 2011. Disponível em: <[http://www.fao.org/fishery/culturedspecies/Litopenaeus\\_vannamei/es](http://www.fao.org/fishery/culturedspecies/Litopenaeus_vannamei/es)>. Acessado em 7 de outubro de 2011 às 10:21h

Depósito de documentos: Centro de desove, capacidad 160 millones de PL/ANO. Ante proyecto (1988). **FAO**, 2011. Disponível em: <<http://www.fao.org/docrep/field/003/AC410S/AC410S02.htm>>. Acessado em 7 de outubro de 2011 às 11:18h

DIATCHENKO, L.; LAU, *et al.* Suppression subtractive hybridization: A method for generating differentially regulated or tissue-specific cDNA probes and libraries. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 93, p. 6025-6030, 1996.

**FAO. The State of World Fisheries and Aquaculture.** Rome, Italy, p. 18, 2010

FRANCISCO, Ana Karina de. **Relações genéticas entre plantís de reprodutores do camarão marinho *Litopenaeus vannamei*, através do sequenciamento dos genes RNAr 16S e COI do DNAm.** 2003. 73f. Dissertação (Mestrado). Programa de pós-graduação em genética e evolução, Universidade Federal de São Carlos, São Paulo, 2003.

FREITAS, Patrícia Domingues de. **Estudos de diversidade genética em estoques reprodutivos de camarões *Litopenaeus vannamei* cultivados no Brasil.** 2003. 120f. Tese (Doutorado). Centro de ciências biológicas e da saúde, Universidade Federal de São Carlos, São Paulo, 2003.

FREITAS, P., *et al.* Development of EST-SSR Markers by Data Mining in Three Species of Shrimp: *Litopenaeus vannamei*, *Litopenaeus stylirostris*, and *Trachypenaeus birdy*. **Marine Biotechnology**, v. 7, p. 554–569, 2005.

GANEM, R.S. **Conservação da Biodiversidade: Legislação e Políticas Públicas.** Brasília: Câmara dos Deputados, Edições Câmara, 2011.

GARCÍA-TRIANA, A., *et al.* Expression and silencing of Selenoprotein M (SelM) from the white shrimp *Litopenaeus vannamei*: Effect on peroxidase activity and hydrogen peroxide concentration in gills and hepatopancreas. **Comparative Biochemistry and Physiology, Part A** 155, p. 200–204, 2010.

GROSS, P.S., *et al.* Immune gene discovery by expressed sequence tag analysis of hemocytes and hepatopancreas in the Pacific White Shrimp, *Litopenaeus vannamei*, and the Atlantic White Shrimp, *L. setiferus*. **Developmental and Comparative Immunology**, v. 25, p. 565-577, 2001.

HICKMAN JUNIOR; C. P.; ROBERTS, L. S.; LARSON, A. **Princípios integrados de zoolo-  
logia**. 11ª Ed. Rio de Janeiro, RJ. Editora Guanabara Koogan S.A. p 368, 2004.

HUANG, X.; MADAN, A. CAP3: A DNA sequence assembly program. **Genome Re-  
search.**, n. 9, p. 868-877, 1999.

HWANG, D. Y., *et al.* Overexpression of human selenoprotein M differentially regulates the concentrations of antioxidants and H<sub>2</sub>O<sub>2</sub>, the activity of antioxidant enzymes, and the composition of white blood cells in a transgenic rat. **International Journal of Molecular Medi-  
cine**, v. 21 n. 2, p. 169-179, 2008.

ILIN, A. A.; MALYGIN, A. A.; KARPOVA, G. G. Ribosomal protein S18e as a putative molecular staple for the 18S rRNA 3'-major domain core. **Biochimica et Biophysica Acta**, v. 1814, p. 505-512, 2011

IWAMA, G.K.; THOMAS, P. T.; FORSYTH, R. B.; VIJAYAN, M. V. Heat shock protein expression in fish. **Reviews in Fish Biology and Fisheries**, v. 8, p. 35-56, 1998.

KARLSSON C., *et al.* Proteomic analysis of the Drosophila larval hemolymph clot. **The Journal Biological Chemistry**, v. 279, n. 50, p. 52033 – 52041, 2004

LEI, K., *et al.* Difference between hemocyanin subunits from shrimp *Penaeus japonicus* in anti-WSSV defense. **Developmental and Comparative Immunology**, v. 32, p. 808–813, 2008.

MANEERUTTANARUNGROJ, C. *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**, v. 37, p. 363–366, 2006.

MANZANO-LEÓN, N.; DELGADO-COELLO, B.; GUADERRAMA-DÍAZ, M.; MAS-OLIVA, J. b-adaptin: Key molecule for microglial scavenger receptor function under oxidative estresse. **Biochemical and Biophysical Research Communications**, v. 351, p. 588-594, 2006.

MARTINS, Pedro Carlos Cunha. **Influência das condições ambientais e técnicas de produção sobre a incidência de enfermidades no cultivo de camarão marinho, *Litopenaeus vannamei*, no estado do Ceará.** 2003. 117f. Tese (Doutorado). Centro de ciências biológicas e da saúde, Universidade Federal de São Carlos, São Paulo, 2003.

MATSUSHITA, K., *et al.* NADH dehydrogenase of *Corynebacterium glutamicum*. Purification of an NADH dehydrogenase II homolog able to oxidize NADPH. **FEMS Microbiology Letters**, v. 204, p. 271-276, 2001

MCCULLOUGH, B.J., *et al.* 3P- syndrome defines a hearing loss locus in 3p25. 3. **Hearing research**, v. 224, p. 51-60, 2007.

MCGRATH, A. Genome Sequencing and Assembly. In: ELSEVIER, B. V. **Perspectives in Bioanalysis: New High Throughput Technologies for DNA Sequencing and Genomics.** Beijing, China: Elsevier, v. 2, cap. 11, p 327-355, 2007.

OFFATT, P. PLAA; G. L.; DENIZEAU, F. Induction of metallothionein gene expression by epidermal growth factor and its inhibition by transforming growth factor-p and dexamethasone in rat hepatocytes. **Hepatology**, v. 21, n. 4, 1995.

ORMOND, J. G. P. *et al.* A carcinicultura brasileira. **BNDES Setorial**, Rio de Janeiro, n. 19, p. 91-118, 2004.

POLI, C. R., *et al.* Aquicultura: experiências brasileiras. Florionópolis: **Multisafera**, p199-200, 2004.

PROSDOCIMI, F. *et al.* Bioinformática: Manual do Usuário. **Biotecnologia Ciência & Desenvolvimento**, n. 29, p. 18-31, 2003.

Projeto Genoma EST do Camarão *Litopenaeus vannamei* (ShEST). Disponível em: <<http://www.shrimp.ufscar.br>>. Acessado em 23 de julho de 2011 às 12:30h

ROBALINO, J., *et al.* Insights into the immune transcriptome of the shrimp *Litopenaeus vannamei*: tissue-specific expression profiles and transcriptomic responses to immune challenge. **Physiol. Genomics** v. 29, p. 44-45, 2007.

ROBALINO, J., *et al.* Contributions of functional genomics and proteomics to the study of immune responses in the Pacific white leg shrimp *Litopenaeus vannamei*. **Veterinary Immunology and Immunopathology**, v. 128, p. 110-118, 2009.

RUAN Y.-H., et al. Ferritin administration effectively enhances immunity, physiological responses, and survival of Pacific white shrimp (*Litopenaeus vannamei*) challenged with white spot syndrome virus. **Fish & Shellfish Immunology**, v. 28, p. 542 – 548, 2010.

SAMBROOK, J.; FRITSCH, E. F.; MANIATIS, T. **Molecular Cloning: A Laboratory Manual**. 2nd ed. New York: Cold Spring Harbor Laboratory Press, 1989.

SANTOS, C.A; FREITAS, P.D. Anotação genômica e caracterização de locos microsatélites em sequências expressas do genoma do camarão marinho *Litopenaeus vannamei*. **Departamento de Genética e Evolução**, Universidade Federal de São Carlos, p. 6, 2008.

SANTOS, F.C.; ORTEGA, J.M. Bioinformática aplicada à genômica. In: **Melhoramento Genômico**, Minas Gerais: UFV, p. 5, 2003.

SCHEIBYE-ALSING, K. *et al.* Sequence assembly, Review. **Computational Biology and Chemistry**. v. 33 , p. 121–136, 2009.

SEBRAE. **Aquicultura e pesca: camarões**. Estudos de mercado sebrae/espm (Sumário), p. 20, 2008.

SHEAN, B.S.; MYKLES, D. L. Polyubiquitin in crustacean striated muscle: increased expression and conjugation during molt-induced claw muscle atrophy. **Biochimica et Biophysica Acta**, v. 1264, p. 312-322, 1995.

SILVA, Marcelo Margalhães. **Caracterização da variabilidade genética do camarão *Litopenaeus vannamei* BOONE, 1931 em fazendas de produção da região de canavieiras (BA)**. 2007. 74f. Dissertação (Mestrado). Universidade Estadual de Santa Cruz, Bahia, 2007.

SIROVER, M. A. New insights into an old protein: the functional diversity of mammalian glyceraldehyde-3-phosphate dehydrogenase. **Biochimica et Biophysica Acta**, v. 1432, p. 159-184, 1999.

Wikipédia, **Penaeidae**, 2011. Disponível em:< <http://pt.wikipedia.org/wiki/Penaeidae>>. Acessado em 7 de outubro de 2011 às 10:28h

WHITFIELD, E. J.; PRUESS, M.; APWEILER, R. Bioinformatics database infrastructure for biotechnology research. **Journal of Biotechnology**, v. 124, p. 629-639, 2006.

XU, J.; WU, S.; ZHANG, X. Novel Function of QM Protein of Shrimp (*Penaeus Japonicus*) in Regulation of Phenol Oxidase Activity by Interaction with Hemocyanin. **Cell Physiol Biochem**, v. 21, p. 473-480, 2008.

YADAI AH, M, *et al.* Cloning, Escherichia coli expression, purification, characterization, and enzyme assay of the ribosomal protein S4 from wheat seedlings (*Triticum vulgare*). **Protein Expression and Purification**, v. 81, p. 55–62, 2012.

ZHANG, X.; HUANG, C.; QIN, Q. Antiviral properties of hemocyanin isolated from shrimp *Penaeus monodon*. **Antiviral Research**, v. 61, p. 93-99, 2004.

ZHAO, Z.Y., *et al.* Profiling of differentially expressed genes in hepatopancreas of white spot syndrome virus-resistant shrimp (*Litopenaeus vannamei*) by suppression subtractive hybridization. **Fish Shellfish Immuno**, v. 22, p. 520–534, 2007.

## ANEXOS

ANEXO 1 – Resultado de similaridade no *BlastX*, referente às *contigs* da biblioteca EGG, levando em consideração maior número de identidade (mínimo 50%) e *e-value* superior a  $10^{-5}$ .

Biblioteca EGG (n° de ESTs)	Resultado de similaridade Blastx	Organismo	Identidade	e-Value
Contig3 (7)	<u>Gb ACY66538.1 </u> Putative 60S ribosomal protein RPL34	<i>Scylla</i> <i>paramamosain</i>	38/53 (72%)	6e-16
Contig5 (3)	<u>Ref NP_001124585.1 </u> Coatomer subunit epsilon	<i>Pongo abelii</i>	46/77 (60%)	2e-31
Contig (21)	<u>Ref YP_001382109.1 </u> cytochrome c oxidase subunit II	<i>Fenneropenaeus</i> <i>chinensis</i>	44/51 (86%)	1e-31
Contig9 (2)	<u>Sp P80007.1 CRA2_HOMG</u> <u>A</u> Crustacyanin-A2 subunit	<i>Homarus</i> <i>gammaru</i>	58/76 (76%)	1e-38
Contig12 (5)	<u>Gb ACD65185.1 </u> Putative 60S ribosomal protein RPL39	<i>Phoronis muelleri</i>	37/44 (84%)	6e-20
Contig13 (4)	<u>Gb AEB54627.1 </u> Trap	<i>Procambarus</i> <i>clarkii</i>	88/101 (87%)	7e-47
Contig17 (3)	<u>Gb ADO32980.1 </u> Ubiquitin b	<i>Eriocheir sinensis</i>	122/127 (96%)	2e-85
Contig19 (4)	<u>Ref XP_970247.1 </u> PREDICTED: similar to ribo- somal protein L17e	<i>Tribolium</i> <i>castaneum</i>	60/82 (73%)	5e-36
Contig21 (4)	<u>Ref XP_974524.1 </u> PREDICTED: similar to S- phase kinase-associated pro- tein	<i>Tribolium</i> <i>castaneum</i>	92/113 (81%)	5e-36
Contig22 (5)	<u>Gb ADD14882.1 </u> Cytochrome c oxidase sub- unit I	<i>Pagurus</i> <i>granosimanus</i>	132/154 (86%)	1e-54
Contig25 (2)	<u>Gb AAU11097.1 </u> Ribosomal protein L7	<i>Doryteuthis</i> <i>pealeii</i>	45/65 (69%)	1e-09
Contig26 (7)	<u>Gb ADY48680.1 </u> 60S ribosomal protein L27a	<i>Ascaris suum</i>	122/147 (83%)	3e-69
Contig27 (2)	<u>Gb EFN72943.1 </u> 60S ribosomal protein L13a	<i>Camponotus</i> <i>floridanus</i>	68/129 (53%)	2e-31
Contig28 (5)	<u>Gb ABG65676.1 </u> Cytochrome b	<i>Fenneropenaeus</i> <i>chinensis</i>	39/41 (95%)	4e-16
Contig35 (21)	<u>Ref YP_001315037.1 </u> Cytochrome c oxidase sub- unit III	<i>Litopenaeus</i> <i>vannamei</i>	164/183 (90%)	3e-62
Contig36 (4)	<u>Ref XP_002735558.1 </u> PRE- DICTED: putative 40S ribo- somal protein S15-like	<i>Saccoglossus</i> <i>kowalevskii</i>	109/130 (84%)	2e-29
Contig39 (8)	<u>Gb ABM55579.1 </u> Ribosomal protein L28-like	<i>Maconellicoccus</i> <i>hirsutus</i>	56/113 (50%)	5e-32

Contig40 (2)	<u>Gb ABQ10739.1 </u> Cathepsin L	<i>Penaeus monodon</i>	69/76 (91%)	8e-45
Contig43 (2)	<u>Gb AAX55641.1 </u> Ferritin	<i>Litopenaeus vannamei</i>	55/62 (89%)	4e-27
Contig45 (6)	<u>Ref YP_002587073.1 </u> NADH dehydrogenase subunit 2	<i>Litopenaeus stylirostris</i>	152/163 (93%)	2e-43
Contig46 (3)	<u>Ref XP_002739754.1 </u> PREDICTED: 40S ribosomal protein S12-like isoform 1	<i>Saccoglossus kowalevskii</i>	86/115 (75%)	2e-50
Contig48 (4)	<u>Emb CAA47680.1 </u> Cytochrome oxidase subunit I	<i>Artemia parthenogenetica</i>	33/41 (80%)	1e-13
Contig49 (2)	<u>Gb AEB54646.1 </u> Ribosomal protein S11	<i>Procambarus clarkii</i>	115/144 (80%)	9e-71
Contig50 (3)	<u>Gb ADT80672.1 </u> Ribosomal protein S8	<i>Euphydrias aurinia</i>	121/145 (83%)	3e-73
Contig56 (6)	<u>Ref NP_001037243.1 </u> Ribosomal protein L35A	<i>B ombyx mori</i>	81/124 (65%)	6e-48
Contig59 (2)	<u>Ref XP_973003.1 </u> PRE-DICTED: similar to 39S ribosomal protein L13	<i>Tribolium castaneum</i>	30/52 (58%)	1e-14
Contig60 (2)	<u>Gb ABC48600.1 </u> Ribosomal protein L8	<i>Litopenaeus vannamei</i>	203/206 (99%)	2e-146
Contig64 (2)	<u>Gb AEB54641.1 </u> Ribosomal protein L32	<i>Procambarus clarkii</i>	76/87 (87%)	2e-51
Contig65 (3)	<u>Gb ACV72062.1 </u> QM protein	<i>Penaeus monodon</i>	120/123 (98%)	9e-77
Contig68 (4)	<u>Ref YP_002587083.1 </u> NADH dehydrogenase subunit 6	<i>Litopenaeus stylirostris</i>	55/67 (82%)	4e-26
Contig69 (3)	<u>Gb ABF58009.1 </u> ATP synthase F0 subunit 6	<i>Litopenaeus vannamei</i>	99/120 (83%)	8e-42
Contig71 (4)	<u>Gb ACO11016.1 </u> Trans-membrane emp24 domain-containing protein 9 precursor	<i>Caligus rogercresseyi</i>	128/158 (81%)	2e-76
Contig72 (8)	<u>Gb ACR56783.1 </u> Small ubiquitin-like modifier-1	<i>Litopenaeus vannamei</i>	92/93 (99%)	3e-59
Contig73 (2)	<u>Sp Q4GXU6.1 RS4_CARGR</u> 40S ribosomal protein S4	<i>Carabus granulatus</i>	62/120 (52%),	2e-22
Contig74 (10)	<u>Gb ABX44843.1 </u> Putative 60S ribosomal protein RPL38	<i>Flustra foliacea</i>	61/70 (87%)	2e-35



ANEXO 2 – Resultado de similaridade no *BlastX*, referente aos *singletons* da biblioteca EGG, levando em consideração maior número de identidade (mínimo 50%) e *e-value* superior a  $10^{-5}$ .

Biblioteca EGG	Resultado de similaridade Blastx	Organismo	Identidade	e-Value
A04 EGG73	<a href="#">Ref XP_002408950.1 </a> Secreted protein	<i>Ixodes</i> <i>scapularis</i>	28/55 (51%)	4e-13
B01 EGG73	<a href="#">Sp P80007.1 CRA2_HOMGA</a> Crustacyanin-A2 subunit	<i>Homarus</i> <i>gammarus</i>	121/167 (72%)	5e-62
B02 EGG73	<a href="#">Ref XP_001607937.1 </a> PREDICTED: glycerol-3-phosphate dehydrogenase [NAD+], cytoplasmic	<i>Nasonia</i> <i>vitripennis</i>	98/145 (68%)	5e-44
B05 EGG73	<a href="#">Gb AA96646.1 </a> Ras-related nuclear protein isoform	<i>Marsupenaeus</i> <i>japonicus</i>	123/161 (76%)	2e-80
B07 EGG73	<a href="#">Ref XP_002427259.1 </a> 60S ribosomal protein L24, putative	<i>Pediculus</i> <i>humanus</i> <i>corporis</i>	69/106 (65%)	6e-45
C05 EGG73	<a href="#">Gb ACR54112.1 </a> Ribosomal protein P1	<i>Palaemonetes</i> <i>varians</i>	38/51 (75%)	1e-16
C09 S EGG073	<a href="#">Sp Q25009.1 TBB1_HOMAM</a> Tubulin beta-1	<i>Homarus</i> <i>americanus</i>	140/149 (94%)	8e-85
B11 EGG73	<a href="#">Ref XP_972981.1 </a> PREDICTED: similar to Mps one binder kinase activator-like4	<i>Tribolium</i> <i>castaneum</i>	159/189 (84%)	2e-114
E02 EGG73	<a href="#">Gb EFN63123.1 </a> Myotrophin	<i>Camponotus</i> <i>floridanus</i>	80/116 (69%)	5e-30
E04 EGG73	<a href="#">Gb ABC25067.1 </a> Fibrillar protein	<i>Glossina</i> <i>morsitans</i>	119/145 (82%)	2e-83
D12 EGG73	<a href="#">Gb ACO12227.1 </a> 60S ribosomal protein l27a	<i>Lepeophtheirus</i> <i>salmonis</i>	71/94 (76%)	1e-23
F11 EGG073	<a href="#">Ref XP_003424561.1 </a> PREDICTED: guanine nucleotide-binding protein subunit gamma-1-like	<i>Nasonia</i> <i>vitripennis</i>	44/65 (68%)	2e-22
G07 EGG073	<a href="#">Gb ADU03767.1 </a> Heat shock protein 90	<i>Litopenaeus</i> <i>vannamei</i>	53/54 (98%)	3e-15
G12 EGG	<a href="#">Gb ABF58018.1 </a> NADH dehydrogenase subunit 2	<i>Litopenaeus</i> <i>vannamei</i>	65/74 (88%)	2e-30
H02 EGG73	<a href="#">Ref XP_003394747.1 </a> PRE-DICTED: dynein light chain roadblock-type 2-like	<i>Bombus</i> <i>terrestris</i>	80/94 (85%)	4e-51
H11 EGG73	<a href="#">Gb ADM64579.1 </a> Ribosomal protein S24	<i>Hypophthalmichthys</i> <i>nobilis</i>	92/114 (81%)	2e-60
A12 EGG74	<a href="#">Gb ACI30106.1 </a> NADH ubiquinone dehydrogenase	<i>Anopheles</i> <i>darlingi</i>	57/83 (69%)	5e-35
B12 EGG74	<a href="#">Gb AEB54647.1 </a> Ribosomal protein S18	<i>Procambarus</i> <i>clarkii</i>	139/150 (93%)	4e-69
C01	<a href="#">Ref XP_003221225.1 </a>	<i>Anolis</i>	113/153	2e-79

EGG74	PREDICTED: 60S ribosomal protein l18a-like	<i>carolinensis</i>	(74%)	
C02 EGG74	<a href="#">Ref NP_840072.3 </a> Selenoprotein W	<i>Danio rerio</i>	48/83 (58%)	2e-27
C03 EGG74	<a href="#">Ref XP_971769.1 </a> PREDICTED: similar to GA21547-PA	<i>Tribolium castaneum</i>	23/46 (50%)	6e-07
C06 EGG74	<a href="#">Gb ACY71322.1 </a> Non-histone chromosome protein 2	<i>Chrysomela tremulae</i>	65/101 (64%)	2e-24
C08 EGG74	<a href="#">Ref YP_001315041.1 </a> NADH dehydrogenase subunit 4L	<i>Litopenaeus vannamei</i>	88/99 (89%)	3e-36
C12 EGG74	<a href="#">Gb ADD20406.1 </a> Cytochrome c oxidase subunit vib COX12	<i>Glossina morsitans</i>	36/52 (69%)	1e-21
D01 EGG74	<a href="#">Gb EGI66132.1 </a> Transcription initiation factor TFIID subunit 9	<i>Acromyrmex echinator</i>	86/121 (71%)	4e-39
D03 EGG74	<a href="#">Gb ACI30106.1 </a> NADH ubiquinone dehydrogenase	<i>Anopheles darlingi</i>	36/47 (77%)	5e-18
E05 EGG74	<a href="#">Ref XP_001607077.1 </a> PREDICTED: eukaryotic initiation factor 4A-III-	<i>Nasonia vitripennis</i>	136/161 (84%)	5e-79
F01 EGG74	<a href="#">Ref XP_001652964.1 </a> Ubiquinol-cytochrome c reductase complex 14 kd protein	<i>Aedes aegypti</i>	55/90 (61%)	1e-34
F04 EGG74	<a href="#">Gb AAZ08240.1 </a> Elongation factor-1 alpha	<i>Penaeus monodon</i>	145/165 (88%)	3e-80
G04 EGG74	<a href="#">Gb ACY66573.1 </a> Ribosomal protein S4	<i>Scylla paramamosain</i>	68/78 (87%)	9e-42
A02 EGG75	<a href="#">Gb AAX55641.1 </a> Ferritin	<i>Litopenaeus vannamei</i>	110/116 (95%),	5e-63
C05 EGG75	<a href="#">Gb ACR56783.1 </a> Small ubiquitin-like modifier-1	<i>Litopenaeus vannamei</i>	82/82 (100%)	8e-49
F02 EGG75	<a href="#">Ref YP_001315034.1 </a> Cytochrome c oxidase subunit II	<i>Litopenaeus vannamei</i>	160/190 (84%)	2e-58
F06 EGG75	<a href="#">Gb ACO11016.1 </a> Transmembrane emp24 domain-containing protein 9 precursor	<i>Caligus rogercresseyi</i>	90/111 (81%)	4e-63
G01 EGG75	<a href="#">Ref XP_968508.1 </a> PREDICTED: similar to GA11400-PA	<i>Tribolium castaneum</i>	75/102 (74%)	7e-54

ANEXO 3 – Resultado de similaridade no *BlastX*, referente às *contigs* biblioteca NP, levando em consideração maior número de identidade (mínimo 50%) e *e-value* superior a  $10^{-5}$ .

Biblioteca NP (n° de ESTs)	Resultado de similaridade Blastx	Organismo	Identidade	e-Value
Contigs1 (3)	Gb AEI25985.1  Thioredoxin	<i>Penaeus monodon</i>	87/89 (98%)	9e-56
Contigs3 (3)	Gb ADY16617.1  Ribosomal protein	<i>Fenneropenaeus merguiensis</i>	125/126 (99%)	5e-82
Contigs4 (3)	Ref XP_002434050.1  60S ribosomal protein L5	<i>Ixodes scapularis</i>	114/143 (80%)	5e-68
Contigs6 (2)	Gb AEB54644.1  Ribosomal protein L44	<i>Procambarus clarkii</i>	57/61 (93%)	3e-27
Contigs7 (6)	Gb ADC55381.4  Tropomyosin	<i>Fenneropenaeus merguiensis</i>	59/59 (100%)	5e-19
Contigs9 (2)	Gb AAT92102.1  Isopentenyl- diphosphate delta-isomerase	<i>Ixodes pacificus</i>	34/42 (81%)	1e-26
Contigs10 (2)	Gb ADW10720.1  Beta-1,3-glucan-bindingprotein	<i>Fenneropenaeus chinensis</i>	76/87 (87%)	8e-57
Contigs11 (4)	Gb ACO12227.1  60S ribosomal protein l27a	<i>Lepeophtheirus salmonis</i>	123/149 (83%)	3e-51
Contigs15 (3)	Ref XP_001064912.2  PREDICTED: ribosomal pro- tein S10-like	<i>Rattus norvegicus</i>	76/122 (62%)	1e-42
Contigs19 (3)	Gb AEB54641.1  Ribosomal protein L32	<i>Procambarus clarkii</i>	91/108 (84%)	4e-52
Contigs20 (3)	Ref YP_001315034.1  Cytochrome c oxidase subunitII	<i>Litopenaeus vannamei</i>	130/153 (85%)	5e-56
Contigs21 (8)	Ref NP_001037243.1  Ribosomal protein L35A	<i>Bombyx mori</i>	81/124 (65%)	2e-49
Contigs23 (3)	Gb ABD98778.1  Putative mitochondrial ATP synthase epsilon chain	<i>Graphocephala atropunctata</i>	34/55 (62%)	3e-15
Contigs25 (4)	Gb ACY66482.1  NADH: ubi- quinone dehydrogenase	<i>Scylla paramamosain</i>	49/76 (64%)	2e-28
Contigs27 (4)	Ref YP_002922052.1  Cytochrome c oxidase sub III	<i>Farfantepenaeus californiensis</i>	114/131 (87%)	7e-51
Contigs28 (5)	Gb ADW95789.1  Ribosomal protein S30-like protein	<i>Pectinaria gouldii</i>	74/127 (58%)	3e-32
Contigs29 (2)	Gb ADI56518.1  GABA(A) receptor-associated protein	<i>Haliotis diversicolor</i>	113/116 (97%)	5e-80
Contigs33 (7)	Gb ABZ04247.1  Ribosomal protein rpl38	<i>Lineus viridis</i>	63/70 (90%)	8e-36
Contigs35 (8)	Prfl 1212243D Ubiquitin S2	<i>Drosophila melanogaster</i>	72/76 (95%)	4e-44
Contigs38 (3)	Ref XP_862110.1  PREDICTED: similar to Pro- teasome subunit alpha type6	<i>Proteasome iota chain</i>	72/94 (77%)	2e-44

Contigs39 (4)	<a href="#">Gb AAM18049.1 </a> Ribosomal protein S24	<i>Marsupenaeus japonicus</i>	112/117 (96%)	3e-58
Contigs40 (2)	<a href="#">Gb ABR27899.1 </a> 60S ribosomal protein L37	<i>Triatoma infestans</i>	72/84 (86%)	3e-35
Contigs41 (14)	<a href="#">P P80029.1 CRC1_HOMGA</a> Crustacyanin-C1 subunit	<i>Homarus gammarus</i>	96/153 (63%)	1e-73
Contigs43 (2)	<a href="#">Gb ABO26646.1 </a> Mitochondrial ATP synthase delta chain	<i>Haliotis discus</i>	43/76 (57%)	4e-15
Contigs44 (7)	<a href="#">Gb AEB54642.1 </a> Ribosomal protein L35	<i>Procambarus clarkii</i>	75/94 (80%)	1e-44
Contigs47 (4)	<a href="#">Gb ADN23573.1 </a> Ribosomal protein S25	<i>Hyalomma marginatum</i>	56/83 (67%)	6e-21
Contigs48 (7)	<a href="#">Ref XP_003272716.1 </a> PREDICTED: 60S ribosomal protein L31-like	<i>Nomascus leucogenys</i>	20/32 (63%)	4e-07
Contigs49 (3)	<a href="#">Gb AEF13174.1 </a> Polyubiquitin	<i>Delphacodes kuscheli</i>	35/35 (100%)	1e-14
Contigs54 (2)	<a href="#">Gb ABI93176.1 </a> Oncoprotein nm23	<i>Litopenaeus vannamei</i>	146/151 (97%)	6e-104
Contigs55 (2)	<a href="#">Gb AEB54647.1 </a> Ribosomal protein S18	<i>Procambarus clarkii</i>	36/37 (97%)	3e-28
Contigs56 (2)	<a href="#">Gb ABM55579.1 </a> Ribosomal protein L28-like protein	<i>Maconellicoccus hirsutus</i>	37/71 (52%)	1e-18
Contigs61 (2)	<a href="#">Gb ACD65185.1 </a> Putative 60S ribosomal protein RPL39	<i>Phoronis muelleri</i>	38/44 (86%)	5e-22
Contigs63 (10)	<a href="#">Ref XP_002735558.1 </a> PREDICTED: putative 40S ribosomal protein S15-like	<i>Saccoglossus kowalevskii</i>	113/130 (87%)	5e-49
Contigs67 (13)	<a href="#">Sp P80007.1 CRA2_HOMGA</a> Crustacyanin-A2 subunit	<i>Homarus gammarus</i>	92/126 (73%)	2e-49
Contigs68 (9)	<a href="#">Gb AAV74196.1 </a> Ribosomal protein L26	<i>Penaeus monodon</i>	126/136 (93%)	3e-82
Contigs71 (2)	<a href="#">Emb CBM82441.1 </a> Histone H2A-VI	<i>Rhabdopleura compacta</i>	76/88 (86%)	9e-39
Contigs73 (8)	<a href="#">Gb ADN23552.1 </a> 60S acidic ribosomal proteinp1	<i>Hyalomma marginatum</i>	45/64 (70%)	2e-13
Contigs74 (2)	<a href="#">Gb ACR54116.1 </a> Myosin 1 light chain	<i>Palaemonetes varians</i>	37/44 (84%)	1e-13

**ANEXO 4** – Resultado de similaridade no *BlastX*, referente aos *singletons* biblioteca NP, levando em consideração maior número de identidade (mínimo 50%) e *e-value* superior a  $10^{-5}$ .

Biblioteca NP	Resultado de similaridade Blastx	Organismo	Identidade	e-Value
B04 NP1063	<a href="#">Gb ADZ31249.1 </a> Cytochrome b	<i>Marsupenaeus japonicus</i>	64/75 (85%)	9e-30
B05 NP1063	<a href="#">Gb AEB54648.1 </a> Ribosomal protein S19	<i>Procambarus clarkii</i>	31/45 (69%)	1e-10
E01 NP1063	<a href="#">Gb ACU52718.1 </a> Ribosomal protein L10a	<i>Fenneropenaeus merguensis</i>	49/49 (100%)	4e-36
F01 NP1063	<a href="#">Gb ABZ04247.1 </a> Ribosomal protein rpl38	<i>Lineus viridis</i>	37/43 (86%)	6e-20
F03 NP1063	<a href="#">Gb ADW10720.1 </a> Beta-1,3-glucan-binding protein	<i>Fenneropenaeus chinensis</i>	64/115 (56%)	1e-21
A02 NP1064	<a href="#">Ref XP_001516286.1 </a> PRE-DICTED: 40S ribosomal protein S10-like, partial	<i>Ornithorhynchus anatinus</i>	68/90 (76%)	7e-44
B05 NP1064	<a href="#">Gb AAB36093.1 </a> Neb-cgp=camp generating peptide	<i>Sarcophaga bullata</i>	32/48 (67%)	1e-14
C02 NP1064	<a href="#">Gb ABD98778.1 </a> Putative mitochondrial ATP synthase epsilon chain	<i>Graphocephala atropunctata</i>	34/55 (62%)	1e-15
C09 NP1064	<a href="#">Gb AEB54641.1 </a> Ribosomal protein L32	<i>Procambarus clarkii</i>	94/114 (82%)	5e-54
H09 NP1064	<a href="#">Gb AAF81598.1 </a> <a href="#">AF277456_1</a> Actin A1	<i>Papilio glaucus</i>	98/109 (90%)	3e-25
C02 NP1065	<a href="#">Gb AAA37165.1 </a> Alpha-cardiac actin	<i>Mus musculus</i>	99/108 (92%)	1e-20
F01 NP1065	<a href="#">Gb ACD69503.1 </a> Ubiquinone biosynthesis protein	<i>Gryllus pennsylvanicus</i>	46/66 (70%)	1e-19
F02 NP1065	<a href="#">Gb ABG65676.1 </a> Cytochrome b	<i>Fenneropenaeus chinensis</i>	115/158 (73%)	3e-63
H06 NP1065	<a href="#">Gb ABG65676.1 </a> Cytochrome b	<i>Fenneropenaeus chinensis</i>	58/61 (95%)	2e-38
G09 NP1065	<a href="#">Ref XP_001811714.1 </a> PRE-DICTED: similar to DC2 protein	<i>Tribolium castaneum</i>	61/68 (90%)	2e-35
G07 NP106	<a href="#">Ref XP_001648152.1 </a> U1 small nuclear ribonucleoprotein	<i>Aedes aegypti</i>	74/96 (77%)	1e-45
E12 NP106	<a href="#">Gb EFN82601.1 </a> Nuclear protein Hcc-1	<i>Harpegnathos saltator</i>	31/45 (69%)	7e-10
E09 NP106	<a href="#">Gb ACO12449.1 </a> 60S ribosomal protein L21	<i>Lepeophtheirus salmonis</i>	64/81 (79%),	7e-40
D11 NP1065	<a href="#">Ref XP_001599728.1 </a> PRE-DICTED: UPF0139 membrane protein pmsmaa27 like	<i>Nasonia vitripennis</i>	71/99 (72%)	1e-25

ANEXO 5 – Resultado de similaridade no *BlastX*, referente à biblioteca PO, levando em consideração maior número de identidade (mínimo 50%) e *e-value* superior a  $10^{-5}$ .

Biblioteca PO (n° de ESTs)	Resultado de similaridade Blastx	Organismo	Identidade	e-Value
Contig2 (4) PO1106	Gb ACV72062.1  QM protein	<i>Penaeus monodon</i>	65/68 (96%)	2e-60
B01 PO1106	Gb AAQ86830.1  Beta-adaptin	<i>Ixodes scapularis</i>	83/133 (62%)	4e-52
B08 PO1106	Gb ABW90384.1  Putative ribosomal protein S25	<i>Sipunculus nudus</i>	29/32 (91%)	4e-08
C02 PO1106	Gb AEB54642.1  Ribosomal protein L35	<i>Procambarus clarkii</i>	74/95 (78%)	1e-42
D02 PO1106	Sp P05547.1 TNNI_PONLE Troponin I	<i>Pontastacus leptodactylus</i>	65/90 (72%)	2e-15
D05 PO1106	Ref XP_975256.1  PREDICTED: similar to translocon-associated pro- tein subunit alpha	<i>Tribolium castaneum</i>	48/77 (62%)	1e-28
D06 PO1106	Pdb 2W1O A Chain A, Nmr Structure Of Dimerization Domain Of Human Ribosomal Protein P2	<i>Homo sapiens</i>	22/42 (52%)	3e-06
E06 PO1106	Gb AAH82286.1  Rps16 protein	<i>Mus musculus</i>	85/106 (80%)	2e-54
E09 PO1106	Ref XP_972603.1  PREDICTED: similar to lethal 35Di	<i>Tribolium castaneum</i>	42/67 (63%)	2e-22
F08 PO1106	Sp P80007.1 CRA2_HOMG A Crustacyanin-A2 subunit	<i>Homarus gammarus</i>	39/55 (71%)	5e-07
G06 PO1106	Ref YP_001382109.1  cy- tochrome c oxidase subunit II	<i>Fenneropenaeus chinensis</i>	88/100 (88%)	2e-30
H02 PO1106	Ref YP_001315037.1  cy- tochrome c oxidase subunit III	<i>Litopenaeus vannamei</i>	128/143 (90%)	8e-77
H06_2 PO1106	Gb ACH81781.1  Fructose 1,6-bisphosphate aldolase	<i>Artemia franciscana</i>	126/156 (81%)	2e-87

ANEXO 6 – Resultado de similaridade no *BlastX*, referente às *contigs* biblioteca HP, levando em consideração maior número de identidade (mínimo 50%) e *e-value* superior a  $10^{-5}$ .

Biblioteca HP (n° de ESTs)	Resultado de similaridade Blastx	Organismo	Identidade	e-Value
Contig1 (2)	Gb ABU62825.1  C-type lectin	<i>Litopenaeus vannamei</i>	64/99 (65%)	2e-12
Contig2 (10)	Emb CAB85965.1  Hemocyanin	<i>Litopenaeus vannamei</i>	90/99 (91%)	3e-53
Contig4 (6)	Emb CAA57880.1  Hemocyanin	<i>Litopenaeus vannamei</i>	43/49 (88%)	4e-21
Contig7 (3)	Ref YP_001315034.1  Cytochrome c oxidase subunit II	<i>Litopenaeus vannamei</i>	99/114 (87%)	2e-29
Contig10 (4)	Emb CAB85965.1  hemocyanin	<i>Litopenaeus vannamei</i>	83/87 (95%)	5e-36
Contig11 (2)	Gb ABD65298.1  Destabilase I	<i>Litopenaeus vannamei</i>	45/56 (80%)	8e-25
Contig12 (2)	Gb ADQ28316.1  Metallothionein	<i>Penaeus monodon</i>	32/47 (68%)	4e-07
Contig15 (3)	Gb ABD65306.1  Fatty acid binding protein 10	<i>Litopenaeus vannamei</i>	78/78 (100%)	3e-31
Contig17 (10)	Emb CAA57880.1  Hemocyanin	<i>Litopenaeus vannamei</i>	71/72 (99%)	5e-41
Contig18 (3)	Emb CAA57880.1  Hemocyanin	<i>Litopenaeus vannamei</i>	85/87 (98%)	5e-49

ANEXO 7 – Resultado de similaridade no *BlastX*, referente às *contigs* biblioteca MC, levando em consideração maior número de identidade (mínimo 50%) e *e-value* superior a  $10^{-5}$ .

Biblioteca MC (n° de ESTs)	Resultado de similaridade Blastx	Organismo	Identidade	e-Value
Contigs4 (10)	<a href="#">Gb ACM89179.1 </a> Sarcoplasmic calcium-binding protein	<i>Litopenaeus vannamei</i>	193/193 (100%)	1e-140
Contigs6 (3)	<a href="#">Gb ADK25704.1 </a> Glyceraldehyde-3-phosphate dehydrogenase	<i>Cancer borealis</i>	136/161 (84%)	2e-88
Contigs7 (6)	<a href="#">Gb ACL79888.1 </a> Putative elastin a	<i>Rimicaris exoculata</i>	54/90 (60%)	2e-31
Contigs8 (2)	<a href="#">Ref ZP_03347722.1 </a> L-lactate permease	<i>Salmonella enterica</i>	93/99 (94%)	3e-51
Contigs9 (2)	<a href="#">Gb EFN81819.1 </a> Probable ATP-dependent RNA helicase pitchoune	<i>Harpegnathos saltator</i>	112/156 (72%)	2e-49
Contigs10 (67)	<a href="#">Gb AAC78682.1 </a> Actin 2	<i>Penaeus monodon</i>	365/377 (97%)	4e-145
Contigs11 (3)	<a href="#">Sp A1KYZ2.1 TPM_PENMO</a> Tropomyosin	<i>Penaeus monodon</i>	93/93 (100%)	7e-54
Contigs13 (5)	<a href="#">Ref YP_002587080.1 </a> NADH dehydrogenase subunit 5	<i>Litopenaeus stylirostris</i>	178/203 (88%)	7e-65
Contigs14 (4)	<a href="#">Sp P02635.1 SCPB_PENSP</a> Sarcoplasmic calcium-binding protein	<i>Penaeus sp.</i>	136/143 (95%)	1e-97
Contigs15 (4)	<a href="#">Gb AAC78682.1 </a> Actin 2	<i>Penaeus monodon</i>	141/159 (89%)	3e-66
Contigs16 (4)	<a href="#">Gb ADC55381.4 </a> Tropomyosin	<i>Fenneropenaeus merguensis</i>	79/81 (98%)	5e-28
Contigs18 (2)	<a href="#">Dbj BAG30738.1 </a> Troponin T	<i>Papilio xuthus</i>	95/146 (65%)	2e-38
Contigs22 (21)	<a href="#">Gb ADM34185.1 </a> Myosin light chain	<i>Penaeus monodon</i>	99/154 (64%)	1e-12
Contigs24 (2)	<a href="#">Gb AAY33769.1 </a> Anti-lipopolysaccharide factor	<i>Litopenaeus stylirostris</i>	87/90 (97%)	2e-45
Contigs25 (3)	<a href="#">Gb AEK11985.1 </a> Calmin-like protein	<i>Tigriopus californicus</i>	42/76 (55%)	2e-20
Contigs27 (3)	<a href="#">Ref ZP_08369981.1 </a> Glucokinase	<i>Escherichia coli</i> TA271	140/153 (92%)	2e-94
Contigs29 (2)	<a href="#">Gb ABR27687.1 </a> Cuticle proprotein procp5.2	<i>Callinectes sapidus</i>	95/187 (51%)	1e-37
Contigs34 (2)	<a href="#">Ref XP_001656319.1 </a> Glycine cleavage system h protein	<i>Aedes aegypti</i>	79/123 (64%)	2e-49
Contigs37 (6)	<a href="#">Ref YP_001315043.1 </a> Cytochrome b	<i>Litopenaeus vannamei</i>	264/286 (92%)	2e-141
Contigs38 (4)	<a href="#">Gb AAP21806.1 </a> Cytochrome oxidase subunit I	<i>Litopenaeus vannamei</i>	90/96 (94%)	2e-53
Contigs39 (3)	<a href="#">Emb CAH04316.1 </a>	<i>Dascillus</i>	115/154	6e-67



	S5e ribosomal protein	<i>cervinus</i>	(75%)	
Contigs40 (7)	<u>Gb AAT66425.2 </u> Actin T2	<i>Litopenaeus</i> <i>vannamei</i>	131/150 (87%)	8e-59
Contigs41 (8)	<u>Gb EFW49978.1 </u> Nitrogen regulation protein NR(II)	<i>Shigella</i> <i>dysenteriae</i>	117/119 (98%)	2e-75
Contigs44 (3)	<u>Gb AAC78141.1 </u> Phosphopyruvate hydratase	<i>Penaeus</i> <i>monodon</i>	155/162 (96%)	1e-99
Contigs48 (3)	<u>Ref YP_001315044.1 </u> NADH dehydrogenase subunit 1	<i>Litopenaeus</i> <i>vannamei</i>	96/122 (79%)	5e-72
Contigs50 (3)	<u>Ref YP_001315034.1 </u> Cytochrome c oxidase subunit II	<i>Litopenaeus</i> <i>vannamei</i>	155/173 (90%)	1e-69
Contigs51 (2)	<u>Gb AEB54639.1 </u> Ribosomal protein L13	<i>Procambarus</i> <i>clarkii</i>	88/126 (70%)	1e-57
Contigs52 (15)	<u>Gb ADD70028.1 </u> Myosin light chain	<i>Marsupenaeus</i> <i>japonicus</i>	153/153 (100%)	3e-63
Contigs53 (2)	<u>Gb ABY66597.1 </u> Pyruvate kinase 2	<i>Litopenaeus</i> <i>vannamei</i>	156/158 (99%)	1e-73
Contigs54 (7)	<u>Gb ACV40756.1 </u> Troponin I	<i>Fenneropenaeus</i> <i>chinensis</i>	87/89 (98%)	1e-38
Contigs55 (2)	<u>Gb ABA29619.1 </u> Cytochrome oxidase subunit I	<i>Arctia caja</i>	43/52 (83%)	9e-24
Contigs57 (2)	<u>Gb ADO32980.1 </u> Ubiquitin b	<i>Eriocheir</i> <i>sinensis</i>	105/110 (95%)	7e-35
Contigs58 (2)	<u>Prf 2106164A</u> Cuticle protein		41/70 (59%)	1e-14
Contigs59 (7)	<u>B AAV41424.1 </u> Cytochrome oxidase subunit I	<i>Litopenaeus</i> <i>vannamei</i>	160/183 (87%)	6e-72
Contigs60 (6)	<u>Gb ABF58009.1 </u> ATP synthase F0 subunit 6	<i>Litopenaeus</i> <i>vannamei</i>	83/98 (85%)	7e-25
Contigs61 (2)	<u>Gb ACS36545.1 </u> Troponin C isoform 5	<i>Homarus</i> <i>americanus</i>	84/122 (69%)	5e-51
Contigs62 (12)	<u>Ref YP_003498402.1 </u> Rare lipoprotein A precursor	<i>Escherichia coli</i> <i>O55:H7</i>	174/192 (91%)	5e-114
Contigs63 (6)	<u>Ref YP_001315037.1 </u> Cytochrome c oxidase sub III	<i>Litopenaeus</i> <i>vannamei</i>	91/136 (67%)	3e-32
Contigs64 (6)	<u>Ref ZP_06654360.1 </u> Lipid A biosynthesis lauroyl acyltransferase	<i>Escherichia coli</i> <i>B354</i>	154/154 (100%)	6e-112
Contigs65 (9)	<u>Ref ZP_03385825.1 </u> Ubiquinone/menaquinone biosynthesis methyltransferase	<i>Salmonella</i> <i>enterica</i>	69/71 (97%)	1e-39

**ANEXO 8** – Resultado de similaridade no *BlastX*, referente aos *singlets* biblioteca MC, levando em consideração maior número de identidade (mínimo 50%) e *e-value* superior a  $10^{-5}$ .

Biblioteca MC	Resultado de similaridade Blastx	Organismo	Identidade	e-Value
A02 MC2077	<a href="#">Gb ADK66280.1 </a> Intracellular fatty acid-binding protein	<i>Litopenaeus vannamei</i>	115/116 (99%)	8e-46
A12 MC2077	<a href="#">Dbj BAJ53744.1 </a> Nucleotide excision repair protein	<i>Marsupenaeus japonicus</i>	65/80 (81%),	2e-15
C01 MC2077	<a href="#">Gb AAB82291.1 </a> Sarco/endoplasmic reticulum Ca <sup>2+</sup> -atpase	<i>Procambarus clarkii</i>	49/60 (82%)	2e-23
D07 MC2077	<a href="#">Gb ADM16544.2 </a> Carbonic anhydrase I	<i>Litopenaeus vannamei</i>	45/74 (61%)	7e-17
E05 MC2077	<a href="#">Ref YP_373554.1 </a> HAD family hydrolase	<i>Burkholderia sp</i>	120/125 (96%)	6e-61
G09 MC2077	<a href="#">Dbj BAK61430.1 </a> Myosin heavy chain type b	<i>Marsupenaeus japonicus</i>	68/91 (75%)	4e-23
A10 MC2078	<a href="#">Ref NP_001188130.1 </a> Mitochondrial malate dehydrogenase	<i>Ictalurus punctatus</i>	119/146 (82%)	5e-75
B01 MC2078	<a href="#">Gb ADY16617.1 </a> Ribosomal protein S3a	<i>Fenneropenaeus merguensis</i>	168/172 (98%)	7e-96
C04 MC2078	<a href="#">Dbj BAK61430.1 </a> Myosin heavy chain type b	<i>Marsupenaeus japonicus</i>	109/121 (90%)	5e-43
E08 MC2078	<a href="#">Ref YP_371765.1 </a> 2-oxoacid dehydrogenase subunit E1	<i>Burkholderia sp.</i>	24/40 (60%)	2e-05
G01 MC2078	<a href="#">Gb AEB54651.1 </a> Ribosomal protein S9	<i>Procambarus clarkii</i>	159/172 (92%)	2e-107
G08 MC2078	<a href="#">Gb ACO54864.1 </a> Mitochondrial aconitase 2	<i>Gillichthys mirabilis</i>	59/78 (76%)	4e-34
H07 MC2078	<a href="#">Gb ACY66490.1 </a> 40S ribosomal protein S12E	<i>Scylla paramamosain</i>	25/31 (81%)	2e-06
A01 MC2079	<a href="#">Ref ZP_08345958.1 </a> Protein phno	<i>Escherichia coli H736</i>	115/119 (97%)	3e-64
A07 MC2079b	<a href="#">Ref XP_001851278.1 </a> Cornichon protein	<i>Culex quinquefasciatus</i>	102/118 (86%)	2e-55
A09 MC2079b	<a href="#">Gb AAZ20120.1 </a> Myosin heavy chain	<i>Farfantepenaeus paulensis</i>	34/45 (76%)	4e-12
B04 MC2079	<a href="#">Ref XP_002404909.1 </a> Prefoldin, putative	<i>Ixodes scapularis</i>	41/59 (69%)	7e-08
C03 MC2079	<a href="#">Gb AAN03488.1 </a> Transferrin	<i>Mastotermes darwiniensis</i>	41/115 (36%)	2e-12
C04 MC2079	<a href="#">Ref XP_001607950.1 </a> PREDICTED: charged multivesicular body protein 1b-like	<i>Nasonia vitripennis</i>	86/104 (83%)	2e-35

F04 MC2079	<a href="#">Gb ACO09577.1 </a> Serine/threonine-protein phosphatase 2A catalytic subunit beta isoform	<i>Osmerus mordax</i>	75/91 (82%)	2e-51
H06 MC2079	<a href="#">Dbj BAK61430.1 </a> Myosin heavy chain type b	<i>Marsupenaeus japonicus</i>	115/122 (94%)	8e-68
A01 MC2080	<a href="#">Pdb 3AV2 A</a> Chain A, The Human Nucleosome Structure Containing The Histone	<i>Homo sapiens</i>	113/116 (97%)	2e-72
A05 MC2080	<a href="#">Gb EGB51590.1 </a> Hsp90-like protein	<i>Escherichia coli H263</i>	109/114 (96%)	2e-51
B06 MC2080	<a href="#">Gb EGI65334.1 </a> 60S ribosomal protein L37a	<i>Acromyrmex echinator</i>	67/93 (72%),	1e-22
B07 MC2080	<a href="#">Gb AAC47523.1 </a> Alpha-2-tubulin	<i>Gecarcinus lateralis</i>	71/74 (96%)	2e-10
F12 MC2080	<a href="#">Gb ACH81781.1 </a> Fructose 1,6-bisphosphate aldolase	<i>Artemia franciscana</i>	110/150 (73%)	9e-73
H01 MC2080	<a href="#">Gb EGP24490.1 </a> Regulatory protein ada	<i>Escherichia coli PCN033</i>	16/23 (70%)	1e-05
H02 MC2080	<a href="#">Gb EGR60947.1 </a> Putative uroporphyrinogen III C-methyltransferase	<i>Escherichia coli O104:H4 str. 01-09591</i>	118/126 (94%)	7e-27
H11 MC2080	<a href="#">Emb CAI40951.1 </a> Eukaryotic translation elongation factor 1 alpha 1	<i>Homo sapiens</i>	131/145 (90%)	2e-92
H12 MC2080	<a href="#">Gb ABU41055.1 </a> Troponin T-5	<i>Lepeophtheirus salmonis</i>	42/62 (68%),	2e-26
A05 MC2081	<a href="#">Gb ACM89179.1 </a> Sarcoplasmic calcium-binding protein	<i>Litopenaeus vannamei</i>	67/124 (54%)	9e-25
D01 MC2081	<a href="#">Gb AEK78307.1 </a> Adenine nucleotide translocase 2	<i>Litopenaeus vannamei</i>	113/114 (99%)	3e-33
F04 MC2081	<a href="#">Gb ABI93176.1 </a> Oncoprotein nm23	<i>Litopenaeus vannamei</i>	121/124 (98%)	2e-83
F11 MC2081	<a href="#">Gb AEB54792.1 </a> Single VWC domain protein 2	<i>Litopenaeus vannamei</i>	99/104 (95%)	7e-65
G09 MC2081	<a href="#">Ref XP_002407905.1 </a> Peptide methionine sulfoxide reductase, putative	<i>Ixodes scapularis</i>	48/80 (60%)	5e-29
H12 MC2081	<a href="#">Ref XP_003423940.1 </a> PREDICTED: CCHC-type zinc finger protein CG3800-like	<i>Nasonia vitripennis</i>	18/35 (51%)	9e-07
A06 MC2082	<a href="#">Dbj BAK61429.1 </a> Myosin heavy chain type a	<i>Marsupenaeus japonicus</i>	131/143 (92%)	6e-50
D02 MC2082	<a href="#">Ref YP_001315040.1 </a> NADH dehydrogenase subunit 4	<i>Litopenaeus vannamei</i>	60/68 (88%)	3e-31
D04 S MC2082	<a href="#">Dbj BAB78485.1 </a> Eukaryotic initiation factor eif-	<i>Marsupenaeus japonicus</i>	39/41 (95%)	4e-11

---

4A like protein				
E01 S	<u>Gb ABH10628.1 </u>	<i>Litopenaeus</i>	167/168	2e-104
MC2082	Laminin receptor	<i>vannamei</i>	(99%)	

---