



Universidade Federal do Rio de Janeiro
Centro de Ciências da Saúde
Instituto de Biofísica Carlos Chagas Filho

Juliana Alves Americo

Construção, sequenciamento e anotação funcional de bibliotecas de cDNA normalizadas para dois bivalves tropicais: a ostra *Crassostrea rhizophorae* e a vieira *Nodipecten nodosus*.

Dissertação submetida ao Instituto de Biofísica Carlos Chagas Filho da Universidade Federal do Rio de Janeiro para a obtenção do grau de mestre em Ciências Biológicas (Biofísica)

Rio de Janeiro
2010

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

Universidade Federal do Rio de Janeiro

Juliana Alves Americo

Construção, sequenciamento e anotação funcional de bibliotecas de cDNA normalizadas para dois bivalves tropicais: a ostra *Crassostrea rhizophorae* e a vieira *Nodipecten nodosus*.

Orientador: Prof^o Dr. Mauro de Freitas Rebelo

Laboratório Intermediário de Biologia Molecular Ambiental.

Programa de Biofísica Ambiental.

Instituto de Biofísica Carlos Chagas Filho.

Centro de Ciências da Saúde.

Americo, Juliana Alves.

Construção, sequenciamento e anotação funcional de bibliotecas de cDNA normalizadas para dois bivalves tropicais: a ostra *Crassostrea rhizophorae* e a vieira *Nodipecten nodosus* /Juliana Alves Americo. Rio de Janeiro, UFRJ/IBCCF, 2010.

vi, 98f: il., 31cm.

Orientador: Mauro de Freitas Rebelo.

Dissertação (mestrado) – UFRJ/ CCS/ IBCCF/ Programa de pós-graduação em Ciências Biológicas (modalidade biofísica), 19/03/2010.

Referências Bibliográficas: f75-81.

1. *Crassostrea rhizophorae* 2. *Nodipecten nodosus* 3. Bibliotecas de cDNA I. Rebelo, Mauro de Freitas. II. Universidade Federal do Rio de Janeiro, Centro de Ciências da Saúde, Instituto de Biofísica Carlos Chagas Filho, Programa de pós-graduação em Ciências Biológicas (modalidade biofísica). III. Construção, sequenciamento e anotação funcional de bibliotecas de cDNA normalizadas para dois bivalves tropicais: a ostra *Crassostrea rhizophorae* e a vieira *Nodipecten nodosus*.

**Aos meus pais e a minha irmã,
pelo apoio, força e exemplo.**

“Saber não basta, devemos aplicar,
Querer não basta, devemos fazer”

Goethe

“Mas é preciso ter manha
É preciso ter graça
É preciso ter sonho sempre
Quem traz na pele essa marca
Possui a estranha mania
De ter fé na vida”

Milton Nascimento

AGRADECIMENTOS

Ao professor Mauro Rebelo pela orientação e paciência, pelas muitas oportunidades de aprendizado e crescimento durante os últimos anos e por ter me confiado um projeto no qual eu realmente gosto de trabalhar.

Ao professor Aldo Viarengo por me receber em seu laboratório e ao professor Francesco Dondero por toda paciência, disponibilidade e apoio na realização deste trabalho.

A Caterina e Alessandro pela boa vontade e ajuda no laboratório, na Itália.

A Diego Coltella pela ajuda com a transformação das bibliotecas de cDNA.

Ao IED-BIG pela disponibilização das vieiras.

Ao CNPQ pela bolsa de mestrado.

Aos meus amigos de laboratório, Lia, Gisele, Milena e Thiago pela convivência, viagens, conversas e por toda a ajuda! Agradeço especialmente a Lia pelos “baldes” de hemolinfa e a Gisele, pelos “baldes” de RNA. Fazer o mestrado sem vocês certamente não teria sido tão divertido!

A Aline, Diana, Raquel e Vania pela constante amizade e por suportarem minha nerdice e piadas biológicas muitas vezes sem graça desde 2003/01! E a Diana e Raquel por estarem no mesmo barco que eu mais uma vez, no mestrado.

As amigas que ajudaram a tornar a saudade de casa menor, Floriane, Anastazija e Alexandra.

Aos meus pais por terem me ensinado desde cedo a valorizar a educação e o conhecimento, pelo incentivo, apoio, carinho e compreensão nos meus muitos momentos de estresse.

A minha irmã, pela amizade, pelo apoio constante, pela enorme paciência e, além de tudo isso, pelo suporte acadêmico-científico assistindo pacientemente todas as minhas apresentações! Muito obrigada, Tati!

A Deus pelas pessoas e oportunidades que coloca no meu caminho.

ÍNDICE

AGRADECIMENTOS.....	vi
ÍNDICE.....	vii
LISTA DE ABREVIATURAS.....	ix
RESUMO.....	x
ABSTRACT.....	xi
1. INTRODUÇÃO.....	12
1.1. A era das ômicas.....	13
1.2. Transcriptômica.....	15
1.2.1. A anotação funcional de genes.....	20
1.3. Relações entre genes e ambiente.....	22
1.3.1. Organismos modelo para o estudo das relações entre genes e ambiente.....	23
2. OBJETIVOS.....	27
2.1. Objetivos específicos.....	27
3. MATERIAIS E MÉTODOS.....	28
3.1. Extração e purificação de RNA.....	31
3.2. Síntese e amplificação de cDNA.....	32
3.2.1. Amplificação de cDNA por LD-PCR e regulação do tamanho médio dos produtos de PCR.....	35
3.3. Normalização do cDNA.....	39
3.4. Amplificação do cDNA normalizado.....	42
3.5. Digestão com enzima restrição (<i>Sfi</i> I) e fracionamento do cDNA normalizado.....	41
3.6. Reação de ligação do cDNA ao vetor pDNR-LIB.....	48
3.7. Transformação das bibliotecas de cDNA em <i>Escherichia coli</i> , titulação e avaliação inicial da proporção de clones recombinantes.....	50
3.8. Extração de DNA plasmidial e seqüenciamento.....	52
3.9. Processamento das sequências e anotação funcional de genes.....	53
4. RESULTADOS E DISCUSSÃO.....	56
4.1. Avaliação inicial de parâmetros de qualidade das bibliotecas de cDNA.....	56
4.2. Limpeza, agrupamento e montagem das sequências.....	57
4.3. Caracterização e anotação funcional das sequências.....	59
4.3.1. Identificação de sequências similares.....	59

4.3.2. Anotação funcional das sequências.....	62
4.3.3. Identificação de genes de <i>Nodipecten nodosus</i>	64
4.3.4. Identificação de genes de <i>Crassostrea rhizophorae</i>	68
5. CONCLUSÕES.....	73
6. PERSPECTIVAS.....	74
7. REFERÊNCIAS BIBLIOGRÁFICAS.....	75
APÊNDICE A – Anotações funcionais de genes de <i>N. nodosus</i>	82
APÊNDICE B – Genes de <i>N. nodosus</i> com <i>hits</i> de BLAST e sem anotação funcional.....	85
APÊNDICE C – Anotações funcionais de genes de <i>C. rhizophorae</i>	86
APÊNDICE D – Genes de <i>C. rhizophorae</i> com <i>hits</i> de BLAST e sem anotação funcional.....	95
APÊNDICE E – Sequências que não apresentaram <i>hits</i>	98

LISTA DE ABREVIATURAS

cDNA – DNA complementar

DNA – Ácido desoxiribonucléico

dNTP - desorribonucleotídeo trifosfatado

EDTA - ácido etilenodiamino tetra-acético

EC – código de enzima

EST- Sequência expressa

GO – ontologia gênica

LD-PCR – Reação em cadeia da polimerase de longa distância.

ORF – ‘quadro’ aberto de leitura

PCR – Reação em cadeia da polimerase

RNA – Ácido ribonucléico

TRG – Gene taxonomicamente restrito

RESUMO

Bivalves aquáticos são excelentes modelos para estudos de mecanismos e efeitos de substâncias tóxicas, já que através da filtração da água continuamente acumulam poluentes. Apesar da genômica e transcriptômica de bivalves estarem avançando, os esforços ainda estão concentrados em espécies de regiões temperadas. Para viabilizar a construção de microrranjos de DNA de bivalves tropicais, iniciamos o sequenciamento do transcriptoma de duas espécies nativas, a ostra *Crassostrea rhizophorae* e a vieira *Nodipecten nodosus*. Foram construídas nove bibliotecas de cDNA normalizadas para tecidos de relevância toxicológica, os hemócitos da ostra e a glândula digestiva da ostra e da vieira. As bibliotecas de cDNA apresentaram altas porcentagens de clones recombinantes (62,5-92,5%), elevado número de clones independentes (10^6 - 10^7 cfu) e insertos entre 250 e 2500 pb. Um total de 96 clones foi selecionado para sequenciamento a partir das bibliotecas que apresentaram os melhores resultados. Somente a biblioteca dos hemócitos da ostra apresentou alguma redundância (14,8%), enquanto as demais resultaram somente em sequências únicas. Entre as sequências obtidas, 22 da ostra *C. rhizophorae* e 8 de *N. nodosus* foram funcionalmente anotadas com termos de ontologia gênica e/ou domínios conservados. Os genes identificados estão envolvidos em uma diversidade de processos como controle da expressão gênica, divisão celular, reconhecimento de patógenos, adesão celular e nos metabolismos de proteínas, RNA e lipídeos. Por fim, 82 novas sequências de alta qualidade (54 de *C. rhizophorae* e 28 de *N. nodosus*) foram depositadas no banco de dados *genbank*. Os resultados indicam o potencial das bibliotecas de cDNA geradas neste trabalho para o isolamento e identificação de um número ainda maior de genes destas duas espécies, necessário para a construção de microrranjos de DNA.

ABSTRACT

Bivalves are excellent models for studying mechanisms and effects of toxic substances since they continuously accumulate environmental pollutants through water filtration. Although bivalve genomics and transcriptomics are in advanced stage, efforts are still concentrated in temperate species. To enable the construction of DNA microarrays for tropical bivalves, we started the transcriptome sequencing of two native species, the oyster *Crassostrea rhizophorae* and the scallop *Nodipecten nodosus*. Nine cDNA libraries were built from tissues of toxicological relevance, the oyster's haemocytes and the scallop and oyster's digestive gland. The cDNA libraries obtained showed a high percentage of recombinant clones (62,5-92,5%), high number of independent clones (106-107 cfu) and insert sizes between 250 and 2500 bp. 96 clones were selected for sequencing from the libraries that showed the best results. Only the oyster's haemocytes library showed some redundancy (14,8%), while the remaining libraries resulted only in unique sequences. Twenty two sequences of the oyster *C. rhizophorae* and 8 of the scallop *N. nodosus* were functionally annotated with gene ontology terms and/or conserved domains. The identified genes are involved in a variety of processes such as gene expression control, cell division, pathogen recognition, cell adhesion and metabolism of proteins, RNA and lipids. Finally, a total of 82 high quality sequences (54 of *C. rhizophorae* and 28 of *N. nodosus*) were obtained and submitted to the Genbank database. The results indicate the potential of the generated cDNA libraries for the isolation and identification of an even larger number of genes from these two species, necessary for the construction of DNA microarrays.

1. INTRODUÇÃO

A última década foi marcada por grandes avanços no desenvolvimento de tecnologias aplicadas a biologia molecular. O sequenciamento do primeiro genoma de um organismo eucariótico, a levedura *Saccharomyces cerevisiae* (Goffeau et al., 1995) e do genoma humano (Lander et al., 2001; Venter et al., 2001) deram início a “era dos genomas completos”. Desde a sua criação em 1982, o banco de sequências de ácidos nucleicos *genbank* dobra o número de sequências registradas a cada ano e em 2008 já contava com mais de 98 milhões (NCBI, 2010). Com o surgimento de uma nova geração de métodos de sequenciamento capaz de produzir até 3 Gb por corrida, a tendência é que esta taxa aumente ainda mais (Metzker, 2010). Atualmente, é possível não apenas sequenciar genomas e transcriptomas inteiros, mas quantificar a expressão gênica a nível transcricional e protéico de centenas e milhares de genes simultaneamente (Kandpal et al., 2009).

Apesar destes grandes avanços, a quantidade de sequências conhecidas para modelos não tradicionais de pesquisa (animais e plantas que não são comumente utilizados na pesquisa biomédica) ainda é um fator limitante para o estudo de mecanismos de toxicidade de poluentes ambientais em animais expostos no ambiente e para o desenvolvimento de novas abordagens ecotoxicológicas, baseadas nas relações entre genes e ambiente.

Moluscos bivalves - animais amplamente utilizados como espécies sentinela em programas de biomonitoramento ambiental - são reconhecidos como excelentes modelos para o estudo das relações entre genes e ambiente. Embora esforços sejam feitos para o sequenciamento do genoma e, principalmente, do transcriptoma de algumas espécies de bivalves nativas de regiões temperadas, o mesmo não ocorre para espécies de regiões tropicais, tornando o desenvolvimento de abordagens moleculares para espécies nativas do Brasil um desafio ainda maior. Para mudar este quadro, neste trabalho, demos início a um projeto de sequenciamento e anotação funcional do transcriptoma de duas espécies de bivalves tropicais de importância econômica

ecotoxicológica, a ostra de mangue *Crassostrea rhizophorae* e a vieira *Nodipecten nodosus*.

1.1. A era das “ômicas”

Desde a determinação da estrutura do DNA e do código genético por Watson e Crick (1953), os genes assumiram um papel central na pesquisa dos mais diversos processos biológicos. A descoberta de que o DNA contém toda a informação necessária para o desenvolvimento e manutenção da vida e a compreensão de como esta informação é codificada criou muitas expectativas quanto à identificação de genes associados a processos e fenótipos específicos, principalmente aqueles associados a doenças em humanos. Com o desenvolvimento do método de sequenciamento de DNA de Sanger, em 1977, e o seu aprimoramento e automação ao longo dos anos, uma grande corrida pelo sequenciamento de genes específicos e genomas inteiros foi disparada (Adams, 2008). O anúncio do rascunho do genoma humano em 2000 causou grande impacto e criou grandes expectativas no público, que esperava que com a sequência do genoma humano, a identificação das bases genéticas de doenças levasse a terapias gênicas em um curto espaço de tempo.

Dez anos depois, estas expectativas não se concretizaram totalmente. Em parte, por que apenas pouquíssimas doenças são causadas por alterações genéticas simples, em um único gene. A maior parte dos fenótipos, inclusive patológicos, possui causas multifatoriais e é resultado da interação entre genes e ambiente (Hernandez e Blazer, 2006). Além disso, o projeto genoma humano consistiu na produção de grupos de sequências que estatisticamente cobrem todo o genoma, mas que ainda contém discontinuidades e trechos não sequenciados. Com o recente sequenciamento do genoma de mais indivíduos, aos poucos estas discontinuidades estão sendo resolvidas, trechos antes não sequenciados, incluídos e um “pan genoma”, mais representativo da espécie humana, sendo montado (Li et al., 2010).

No entanto, a questão mais importante, é que apesar de o genoma humano ter sido quase totalmente sequenciado, ainda não está disponível uma relação completa e funcionalmente anotada de todos os genes humanos. A identificação e validação funcional de genes é justamente a etapa mais trabalhosa e desafiante da análise de um genoma, devido à grande quantidade de sequências geradas e ao fato de que a maior parte do DNA genômico não conter regiões codificantes. A identificação de regiões regulatórias, como promotores, atenuadores e acentuadores, tem sido um desafio a parte, uma vez que estas regiões não são necessariamente definidas por sequências conservadas e porque podem estar presentes não apenas em regiões próximas a seus genes alvo, como a milhares de pares de base de distância deles (Griffiths et al., 1999).

Um genoma totalmente sequenciado e funcionalmente anotado fornece um panorama do organismo, possibilitando identificar as vias metabólicas presentes e estudos de genômica comparativa. No entanto, embora seja um grande feito, a conclusão de um genoma é apenas um passo em direção ao completo entendimento das bases genéticas de processos biológicos. Isoladamente, a genômica não responde todas as perguntas.

Paralelamente ao desenvolvimento da genômica, surgiram outras disciplinas que buscam estudar, também em grande escala, aspectos dinâmicos da biologia molecular. Entre elas, estão a expressão gênica em nível transcricional (transcriptoma) e protéico (proteoma), a investigação do fluxo metabólico (metaboloma), a investigação de interações entre proteínas (interatoma), mapeamento do padrão de fosforilação de proteínas (fosfoproteoma) e a identificação do perfil lipídico (lipidoma) (Kandpal et al., 2009; Macek et al., 2009; Oresic, 2009). Da reunião de todas estas disciplinas, coletivamente chamadas “ômicas”, recentemente teve origem a “biologia de sistemas” que busca entender fenômenos biológicos, não através da análise de genes individuais, mas pela rede de interações entre todos os componentes celulares e bioquímicos em uma célula ou organismo. A biologia de sistemas busca integrar dados heterogêneos e tem o audacioso objetivo final de prever fenômenos biológicos, dados os componentes envolvidos (Liu, 2005).

1.2. Transcriptômica

O transcriptoma representa a porção do genoma expressa na forma de moléculas de RNA. No entanto, em comparação com as sequências genômicas, as suas diversidade e complexidade são maiores, uma vez que cada gene pode dar origem a mais de um transcrito devido ao processamento alternativo e a edição de RNA e ao uso alternativo de sítios de iniciação e terminação da transcrição (Frith et al., 2005). Recentemente, novas classes de RNA não codificantes foram identificadas em eucariotos, entre as quais, microRNAs e RNAs de interferência (Carthew e Sontheimer, 2009). Mais surpreendentemente, também foram identificados RNAs longos não codificantes, transcritos a partir de regiões do genoma antes tidas como transcricionalmente inativas, como regiões intergênicas e a partir da fita antisense de genes, mostrando que uma parte bem maior do genoma é transcrita (Ponting et al., 2009).

O transcriptoma é altamente dinâmico. Apesar de todas as células de um organismo possuírem, em princípio, o mesmo material genético, cada tipo celular expressa um conjunto de genes distinto. Além disso, em cada tecido ou tipo celular, os níveis de expressão gênica podem ainda variar em função de inúmeras variáveis, permitindo que a célula responda e se adapte a diferentes condições e sinais externos. Uma das dificuldades em se sequenciar todo o transcriptoma de um organismo é justamente o fato de que não há apenas um, mas vários transcriptomas (Alberts et al., 2004).

Quadro 1 – Os microarranjos de DNA e as bibliotecas de cDNA

Essas são as duas principais abordagens para o estudo de transcriptomas. O microarranjo de DNA (**fig. 1**) é uma das ferramentas usadas para investigar quantitativamente o perfil de expressão gênica, ou seja, o transcriptoma, de um tipo celular, tecido ou organismo em um determinado contexto (Schena et al., 1995; 1998). Este experimento consiste na construção de um *chip* de DNA, uma pequena lâmina de vidro, onde sondas para os diferentes genes de um organismo são ligadas, em posições conhecidas. É possível construir *chips* com milhares de sondas distintas e, portanto, até mesmo todos os genes de um organismo podem ter seus níveis de expressão medidos simultaneamente. Para avaliar comparativamente os níveis de expressão gênica entre duas amostras, o RNA é extraído e convertido em cDNA (DNA complementar). O cDNA de cada amostra é marcado com uma molécula fluorescente diferente e, então, ambos os cDNAs são submetidos a hibridização competitiva com as sondas presentes no *chip*. A quantificação das diferentes fluorescências emitidas em cada posição do chip, ou seja, para cada gene ali representado, indica em qual das duas amostras o gene está mais ou menos expresso ou se não há diferenças de expressão entre as duas amostras, no caso de ambas as fluorescências terem níveis semelhantes.

Os *chips* de DNA podem ser construídos com dois tipos de sonda, oligonucleotídeos sintéticos ou moléculas de cDNA. No primeiro caso, é necessário conhecer as sequências dos genes a serem analisados para, então, desenhar oligonucleotídeos que funcionem como sondas para as sequências expressas. No segundo, é necessário possuir uma coleção de cDNA como fonte para a produção de sondas por PCR. Bibliotecas de cDNA (**fig. 2**) têm sido construídas com este intuito, principalmente para organismos sem genoma e transcriptoma seqüenciados e, portanto, sem sequências disponíveis na literatura e nos bancos de dados (Barret e Kawasaki, 2003)

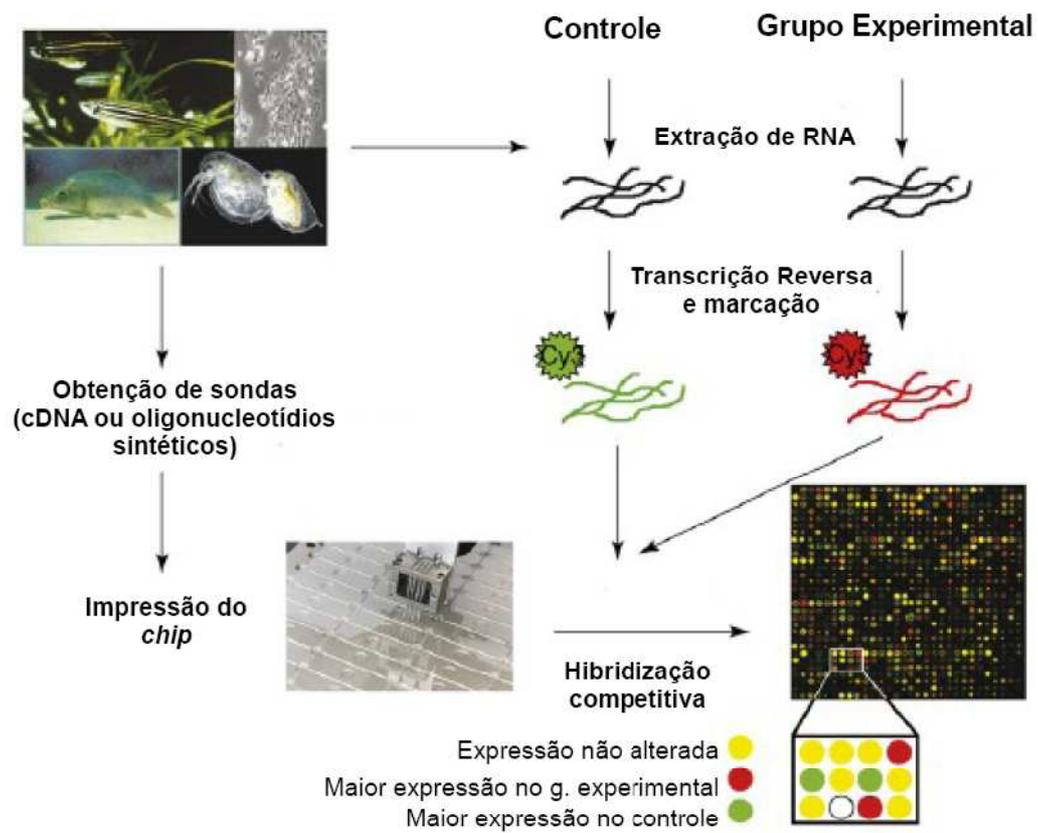


Figura 1 – Principais etapas da construção e utilização de um microrranjo de DNA para comparação quantitativa do perfil de expressão gênica entre duas amostras biológicas. Adaptado de Robbens et al., 2007.

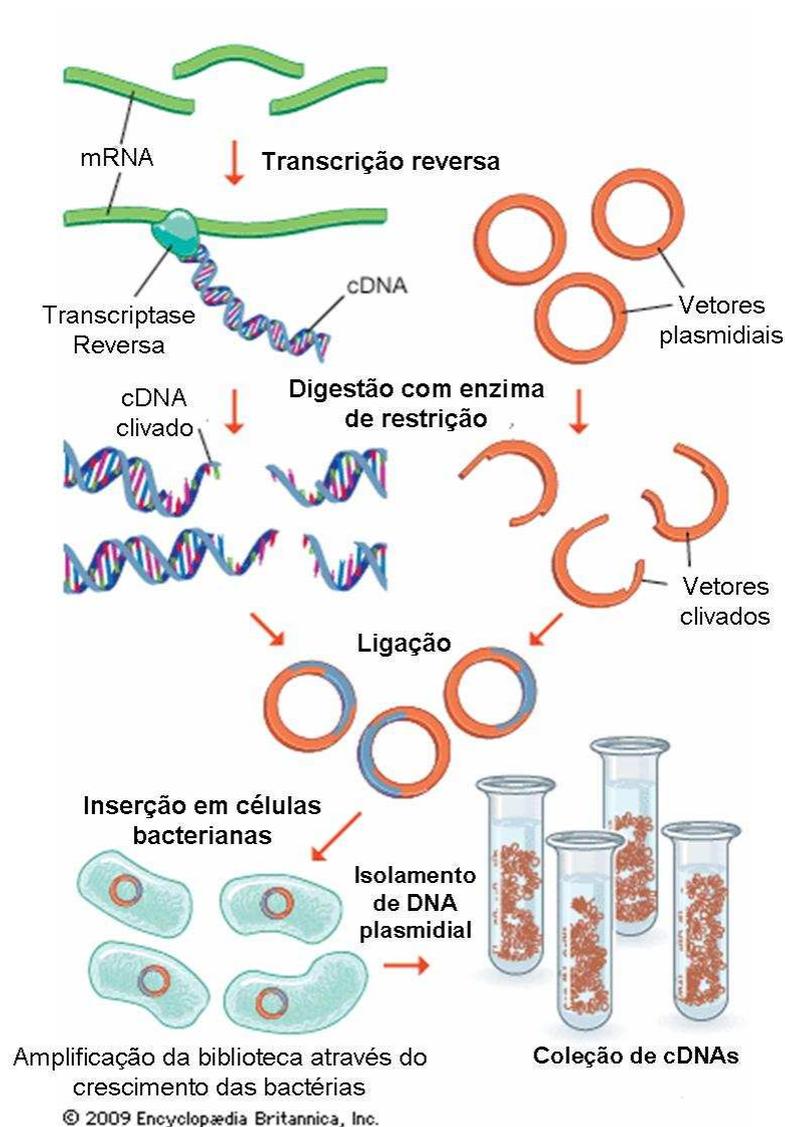


Figura 2: Principais etapas da construção de uma biblioteca de cDNA. Bibliotecas de cDNA são construídas a partir do RNA isolado de um organismo ou de um tipo celular ou tecido específicos. Através de uma reação de transcrição reversa, é sintetizado o cDNA que por fim é inserido em um vetor de clonagem, em geral, um plasmídeo. O DNA plasmidial contendo os insertos de cDNA são inseridos em células bacterianas, onde são replicados e de onde são posteriormente isolados. Em um vetor de clonagem, o inserto de cDNA pode ser facilmente sequenciado com um primer plasmídeo-específico ou ainda utilizado como molde para a produção de sondas por PCR (Barret e Kawasaki, 2003). Figura adaptada de <<http://www.britannica.com/EBchecked/topic/493667/recombinant-DNA-technology>>.

O sequenciamento de transcriptomas através de bibliotecas de cDNA vem sendo conduzido para suprir a falta de projetos genoma em modelos não tradicionais de pesquisa como algas (Ferreira et al., 2007), peixes (Brown et al., 2008), crustáceos (Li e Brouwer, 2009) e moluscos (Kang et al., 2006). Esse procedimento tem a vantagem de fornecer somente as sequências codificantes expressas, dispensando as trabalhosas etapas de identificação de exons e introns, necessária na análise de sequências de origem genômica. Em contrapartida, a redundância das sequências é um fator limitante para a identificação de novos genes através do sequenciamento de bibliotecas de cDNA.

Células eucarióticas (de levedura, camundongo e humanas) apresentam um padrão de expressão onde poucos genes são expressos em níveis muito altos de até 20000 transcritos por célula, enquanto a maioria dos genes é expressa em níveis tão baixos quanto o de 1 molécula por célula. (Bishop et al., 1974; Velculescu et al., 1999; Kuznetsov et al., 2002).

Este perfil de expressão gênica desigual será refletido nas concentrações com que os diferentes genes serão encontrados na biblioteca de cDNA, dificultando o isolamento de genes raros, justamente a maior parte dos genes expressos em uma célula.

Bibliotecas de cDNA com alto grau de redundância podem ser pouco produtivas e conseqüentemente caras para a identificação *de novo* de genes. O número de clones sequenciados para encontrar uma nova sequência pode ser extremamente alto. Por exemplo, para encontrar uma sequência cuja frequência seja 1:10.000 seria necessário sequenciar até 100.000 clones. Para superar este problema, diferentes estratégias de normalização de cDNA, que removem do cDNA o excesso de transcritos super expressos (ou seja, a fração redundante), vem sendo desenvolvidas. Assim, é possível que diferentes sequências estejam presentes em concentrações aproximadamente iguais na biblioteca de cDNA e tenham as mesmas probabilidades de serem isoladas e sequenciadas (Soares et al., 1994; Bonaldo et al., 1996).

1.2.1. A anotação funcional de genes

A bioinformática consiste na aplicação de técnicas de informática (derivadas da matemática, estatística e ciências computacionais) na organização, análise e compreensão de dados biológicos (Luscombe et al., 2001). O desenvolvimento destas ferramentas tem sido essencial para lidar com a grande quantidade de dados gerados na era genômica, em especial, na identificação das possíveis funções dos produtos codificados por genes, porque seria impossível testar todos experimentalmente.

A anotação funcional eletrônica se baseia na premissa de que duas moléculas com funções relacionadas, em geral, apresentam sequências similares e vice-versa. A bioinformática possibilita a identificação de similaridade entre sequências e calcula sua significância estatística, para que se possa avaliar se é possível inferir a função da sequência desconhecida pela sua similaridade com outra de funções conhecidas (Dardel e Képès, 2006).

Quadro 2 – Homologia x Similaridade

É muito comum o uso equivocado do termo homologia com o sentido de similaridade, por exemplo, quando se diz que “duas sequências apresentam 70% de homologia”. Homologia é um conceito biológico qualitativo que indica uma origem comum. Genes homólogos podem ter se originado de um gene presente no último ancestral comum entre as espécies comparadas (genes ortólogos) ou podem estar relacionados por eventos de duplicação (genes parálogos) (Koonin, 2005).

As ferramentas de bioinformática avaliam e quantificam a ocorrência de similaridade e não, de homologia, entre sequências. Genes homólogos, descendentes de um ancestral comum, de forma geral, apresentam sequências e funções similares. Sendo a similaridade um valor numérico, pode ser correlacionada com probabilidade e ser usada como critério para inferir homologia. Quanto maior a similaridade entre duas sequências, menor a probabilidade de que tenham se tornado similares por acaso e, logo, maior a probabilidade de que tenham uma origem comum. No entanto, em alguns casos, genes homólogos podem ser muito pouco similares e, nestes casos, deve ser feita uma análise cautelosa, buscando identificar, por exemplo, se a região similar é funcionalmente importante e se os resíduos conservados são essenciais para o funcionamento da indícios que podem revelar homologia (Koonin e Galperin, 2002).

Uma das principais ferramentas de bioinformática é o BLAST (*Basic Local Alignment Search Tool*), um algoritmo amplamente utilizado para alinhar sequências de DNA e proteínas com sequências depositadas em bancos de dados (Altschul et al., 1990). Os algoritmos utilizados antes do BLAST comparavam sequências analisando-as resíduo por resíduo, o que exigia muito tempo e computadores com grande capacidade de processamento, dificuldades que aumentavam ainda mais com o crescimento exponencial dos bancos de dados. O BLAST tornou este processo mais rápido, ao comparar sequências por pequenos segmentos (“*words*”), contendo um número ajustável de resíduos.

No entanto, a inovação mais interessante do BLAST é que esta ferramenta calcula a significância estatística da similaridade para cada resultado de alinhamento entre duas

sequências, permitindo uma análise mais objetiva dos resultados obtidos. A significância estatística é indicada pelo valor *e* que corresponde a probabilidade de obter o mesmo alinhamento com uma outra sequência aleatória presente no banco de dados. O valor *e* indica, portanto, as chances de que a similaridade tenha sido encontrada por acaso e que não denote uma relação real entre as sequências analisadas. Desta forma, quanto menor o valor *e* obtido para um alinhamento, mais significativo é o resultado.

1.3. Relações entre genes e ambiente.

Um organismo precisa manter o seu meio interno estável com relação às variações do meio externo. Este processo é chamado homeostase e em última instância é controlado pelos genes. As células monitoram sinais externos e internos e os traduzem em respostas biológicas, frequentemente, através de alterações na expressão de genes.

A forma mais frequente de interação entre genes e ambiente é mediada pela expressão gênica: quais genes são expressos, em quais células, em que momento e com que intensidade; são variáveis diretamente influenciadas por fatores ambientais.

A exposição a substâncias tóxicas pode alterar a homeostase de um organismo resultando na ativação de mecanismos de resposta ao estresse (e.g. destoxificação e reparo), que envolvem a ativação e repressão de grupos de genes específicos, para o seu reestabelecimento (Moggs e Orphanides, 2003; Goldstone, 2008). As relações entre expressão gênica e toxicidade são bem estabelecidas como é descrito, de forma clara, por Nuwaysir et al., (1999):

“A toxicidade é geralmente precedida por (ou resulta em) alterações na expressão gênica. Em muitos casos, estas alterações são muito mais sensíveis, características e mensuráveis do que a própria toxicidade”.

Desta forma, com técnicas de quantificação dos níveis de expressão gênica em grande escala, e.g. microarranjos de DNA, é possível detectar a toxicidade de forma precoce, ainda em seus estágios iniciais. Devido à sensibilidade com que a expressão gênica responde a estímulos, mesmo os efeitos de exposições a doses baixas de uma determinada substância podem ser assim detectados. Além disso, por ser uma abordagem global, onde muitos parâmetros (genes) são avaliados ao mesmo tempo, múltiplos efeitos causados por um ou mais compostos tóxicos podem ser identificados e diferenciados. Ao determinar quais genes foram ligados ou desligados, pode-se inferir quais processos biológicos foram afetados positiva ou negativamente identificando os mecanismos de toxicidade envolvidos na quebra da homeostase.

1.3.1. Organismos modelo para o estudo das relações entre genes e ambiente.

“Ainda não foi construído nenhum instrumento que possa medir toxicidade. Concentração química pode ser medida com um instrumento, mas apenas material vivo pode avaliar toxicidade” (Cairns & Mount 1990).

A escolha de um organismo modelo para avaliação da toxicidade envolve várias questões práticas e éticas. É importante considerar a facilidade de manutenção em laboratório, controle das condições experimentais, conhecimento da biologia e fisiologia, assim como da toxicocinética e do metabolismo do xenobitótico no organismo.

Para a investigação de processos celulares e moleculares mais amplamente conservados, além dos tradicionais modelos animais utilizados em pesquisa biomédica (e.g. o rato *Rattus norvegicus* e o camundongo *Mus musculus*), podem ser utilizados diversos organismos eucarióticos, mesmo espécies unicelulares, consideradas um modelo “mínimo” de célula eucariótica. Alguns exemplos são a levedura *Saccharomyces cerevisiae*, o nematódeo *Caenorhabditis elegans*, o crustáceo *Daphnia spp.*, os moluscos *Mytilus edulis* e *Crassostrea gigas*, o peixe *Danio rerio* e o anfíbio

Xenopus spp.

Os moluscos bivalves (ostras, mexilhões, vieiras, etc) são considerados bons modelos toxicológicos, pois devido ao hábito alimentar filtrador acumulam xenobióticos presentes na água, sendo a toxicocinética de muitos destes compostos muito bem descrita nestes animais. Como podem ser facilmente mantidos em laboratório e são sésseis, a exposição a compostos tóxicos pode ser controlada e conduzida através da água. Além disso, bivalves são amplamente distribuídos sendo, portanto, de fácil obtenção e o seu tamanho é grande o bastante para garantir quantidades de tecidos suficientes para análises biológicas.

A possibilidade de trabalho em todas as etapas do ciclo de vida de moluscos bivalves, desde embriões até adultos sexualmente maduros, é um aspecto importante, uma vez que algumas substâncias podem não ser tóxicas em adultos, mas causar danos durante o desenvolvimento. Por fim, muitas das vias de controle de processos químicos e biológicos em moluscos são altamente conservadas, permitindo a generalização de resultados não somente entre moluscos, mas também para espécies de táxons superiores (Rittschof e McClellan-Green, 2005).

Um dos fatores limitantes ao estudo de mecanismos moleculares em moluscos bivalves, principalmente em espécies tropicais, é a carência de sequências. Este quadro começou a mudar quando, recentemente, a identificação de genes em bivalves passou a ser feita em maior escala, através da geração e sequenciamento de bibliotecas genômicas e de cDNA (Venier et al., 2003; Cunnighamn et al., 2006; Fleury et al., 2009a; Venier et al., 2009; Wang et al., 2009). O sequenciamento de um grande número de clones de cDNA começa a possibilitar o estudo de perfis de expressão gênica em algumas espécies, como o mexilhão zebra, *Dreissena polymorpha* (Xu e Faisal, 2009), as ostras *Crassostrea gigas* e *Crassostrea virginica* (Jenny et al., 2007; Fleury et al., 2009b) e os mexilhões *Mytilus edulis* e *Mytilus galloprovincialis* (Dondero et al., 2006b; Venier et al., 2006).

No entanto, estes esforços estão ainda concentrados em poucas espécies, nativas de regiões temperadas. No *genbank*, das mais de 220 mil sequências de bivalves, a maior

parte (67,7%) está concentrada em apenas duas famílias: *Ostreidae* (ostras) e *Mytilidae* (mexilhões). Juntas, a ostra japonesa *C. gigas* e a ostra americana *C. virginica* respondem por 97% das sequências da família *Ostreidae*. Os pectinídeos (vieiras) respondem por 11,5% das sequências de bivalves, distribuídas entre três principais espécies: *Chlamys farreri* (Ásia), *Argopecten irradians* (EUA) e *Mizuhopecten yessoensis* (Japão). Em contraste, há apenas 26 sequências para a ostra *Crassostrea rhizophorae* (família *Ostreidae*) e 4 para a vieira *Nodipecten nodosus* (família *Pectinidae*), dois bivalves nativos do Brasil e bem representados na literatura (**qua. 3**), mas ainda pouco estudados em nível molecular (Metanálise no *genbank*, em Janeiro de 2010).

Com a motivação de incrementar estes números e a fim de viabilizar a implementação da técnica de microarranjos de DNA no estudo de mecanismos moleculares de toxicidade em bivalves tropicais, iniciamos o sequenciamento e anotação funcional do transcriptoma destes dois bivalves tropicais, a ostra *Crassostrea rhizophorae* e a vieira *Nodipecten nodosus*.

Quadro 3 – *Crassostrea rhizophorae* e *Nodipecten nodosus*.

C. rhizophorae é localmente conhecida como ostra de mangue, pois ocorre principalmente neste ecossistema. Sua distribuição, no Brasil, é bastante ampla, indo do Paraná ao Rio Grande do Norte (Ignácio et al., 2000). *N. nodosus*, também chamado “coquille”, ocorre em águas mais profundas. No Brasil, já foi encontrado entre 6 e 30 m de profundidade, ocorrendo em águas tropicais, mas principalmente, subtropicais entre os cabos de São Tomé (RJ) e Santa Marta (SC) (Rupp e Parson, 2006).

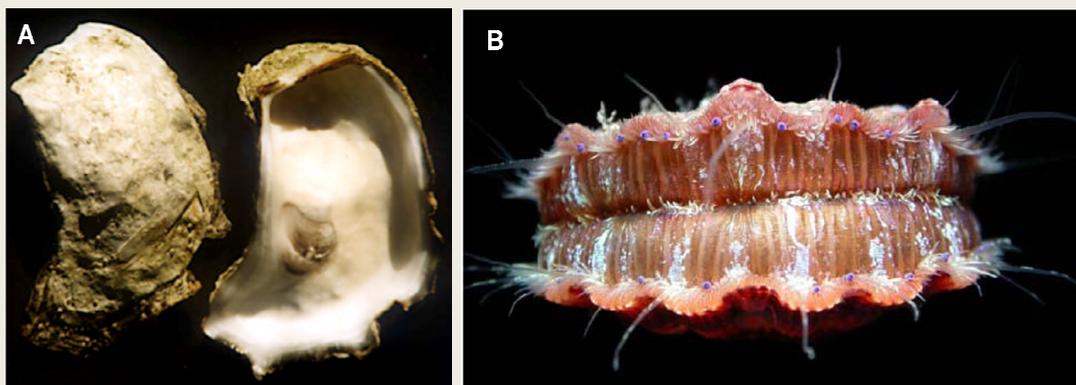


Figura 3 – (A) Valvas de *Crassostrea rhizophorae* (Guilding 1828). Foto: Mauro Rebelo. **(B)** *Nodipecten nodosus* (Linnaeus, 1758). Foto: José Zaganelli.

O acúmulo e a toxicocinética de poluentes ambientais, em particular metais pesados, são bem caracterizados nestes dois bivalves (Rebelo et al., 2003a; Lacerda e Molisani, 2006; Galvão, 2007). Na ostra, a indução/repressão de atividades enzimáticas, proteínas e genes específicos em função da exposição a poluentes ambientais também têm sido avaliados, porém ainda em pequena escala, com um ou poucos destes parâmetros analisados por experimento (Rebelo et al., 2003b; Wanick, 2007; Silva et al., 2008; Zanette et al., 2008).

Estas duas espécies são ainda cultivadas e comercializadas regionalmente. Devido a esta importância econômica, aspectos da biologia destes dois bivalves são estudados, como, por exemplo, a infecção por patógenos e a reprodução (Sabry e Magalhães, 2005; Ferreira et al., 2006). Alguns estudos genéticos também já foram conduzidos como a determinação do cariótipo de *N. nodosus* (Pauls e Affonso, 2000) e identificação molecular da ostra *C. rhizophorae* (Varela et al., 2007).

2. OBJETIVOS

Este trabalho teve como objetivo geral a identificação e anotação funcional de genes expressos na ostra de mangue *Crassostrea rhizophorae* e na vieira *Nodipecten nodosus*.

2.1. OBJETIVOS ESPECÍFICOS

- Construir e validar bibliotecas de cDNA como fontes não redundantes de cDNA para o sequenciamento de genes expressos na glândula digestiva de *Nodipecten nodosus* e na glândula digestiva e nos hemócitos de *Crassostrea rhizophorae*.
- Sequenciar, identificar e caracterizar funcionalmente, através de ferramentas de bioinformática, genes expressos na glândula digestiva de *Nodipecten nodosus* e na glândula digestiva e nos hemócitos de *Crassostrea rhizophorae*.
- Depositar as sequências obtidas no *genbank*, banco de dados de sequências nucleotídicas do NCBI.

3. MATERIAIS E MÉTODOS

Bibliotecas de cDNA foram construídas para a glândula digestiva e hemócitos de *C. rhizophorae* e para a glândula digestiva de *N. nodosus*. A escolha destes tecidos se deu em função da sua importância na resposta imune, acúmulo e metabolismo de xenobióticos.

Para a obtenção de bibliotecas de cDNA tecido-específicas com sequências completas, baixa redundância e alta representatividade, ou seja, com ao menos uma cópia de cada molécula de cDNA/RNA presente inicialmente no RNA original, utilizamos uma combinação de técnicas descritas na **figura 4**. De forma geral, através do mecanismo SMART (*Switching Mechanism At 5' end of RNA Transcript*) (Zhu et al., 2001), o cDNA foi sintetizado com adaptadores em ambas as extremidades, permitindo a amplificação seletiva das sequências de cDNA completas por PCR (**fig. 4, etapa 2**). Esta e todas as etapas de amplificação do cDNA por PCR foram feitas com um único *primer* de forma a regular o tamanho médio dos produtos de cDNA e evitar a amplificação preferencial de sequências curtas (**fig. 4, etapa 2**) (Shagin et al., 1999).

A normalização DSN (*Duplex-Specific Nuclease*) (Zhulidov et al., 2004) removeu o excesso das sequências de cDNA mais frequentes, referentes a genes altamente expressos, resultando em um cDNA onde as diferentes sequências estão presentes em concentrações semelhantes, ou seja, em um cDNA pouco redundante (**fig. 4, etapa 3**). O cDNA normalizado foi então amplificado (**etapa 4**), digerido com a enzima de restrição *SfiI* e fracionado por cromatografia de filtração em gel (**etapa 5**) para remoção das moléculas de cDNA com menos de 250 pb (**fig. 4, etapa 6**).

A fração de cDNA normalizado com mais de 250 pb foi clonada no plasmídeo pDNR-LIB (**fig.5**, clontech), entre os sítios de restrição *SfiIA* e *SfiIB*. As bibliotecas de cDNA resultantes foram em seguida inseridas em *Escherichia coli*, de onde clones recombinantes foram selecionados aleatoriamente para sequenciamento.

A construção das bibliotecas de cDNA (da 2ª a 8ª etapa, fig.4) foi realizada nas

instalações do Departamento de Ciências do Ambiente e da Vida da Università del Piemonte Orientale Amedeo Avogadro (Itália), em colaboração com os professores Francesco Dondero e Aldo Viarengo.

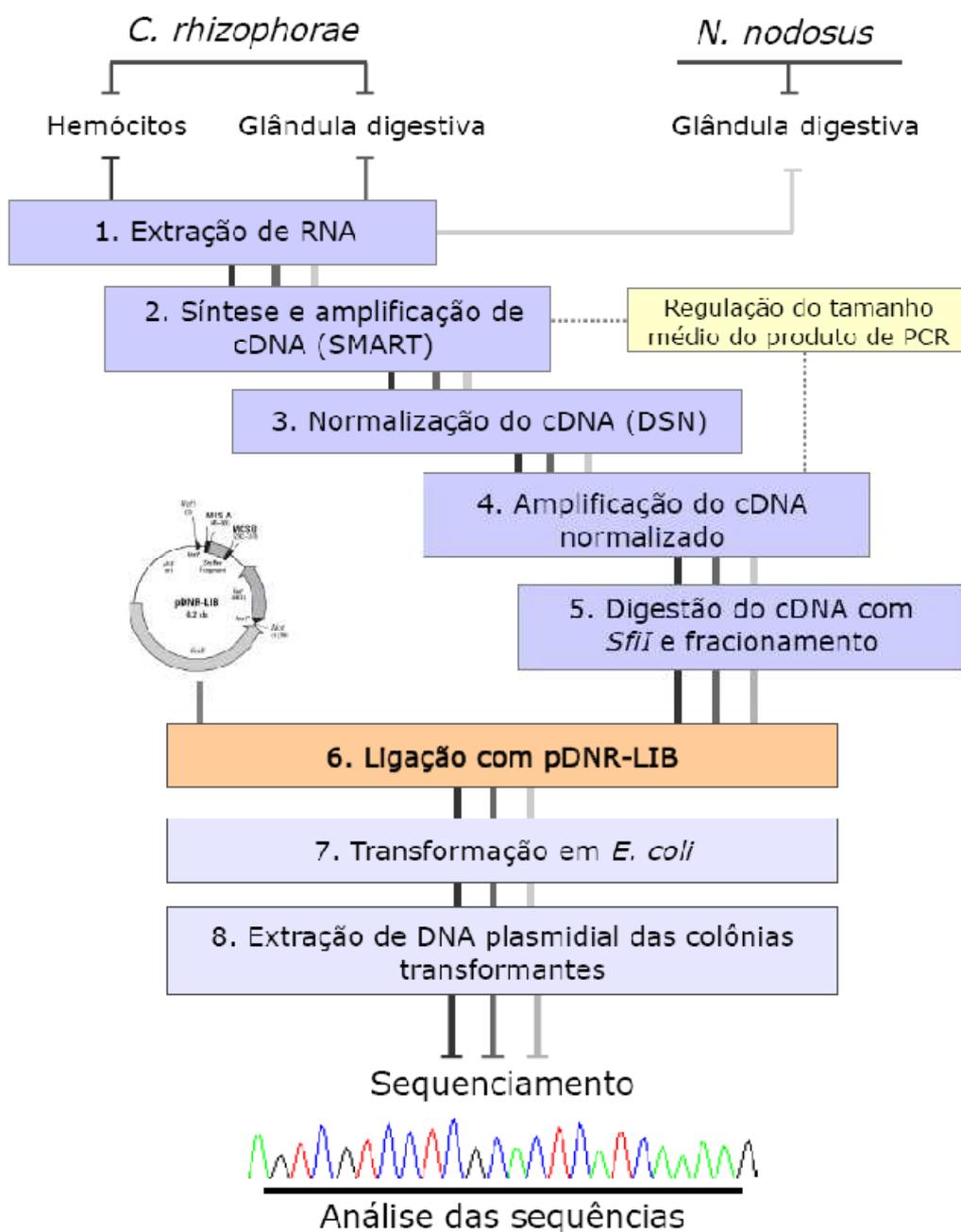
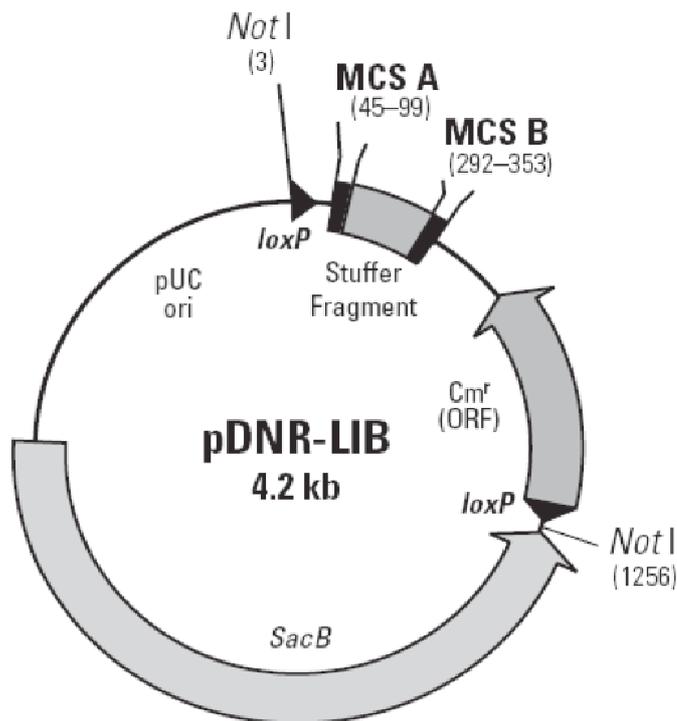


Figura 4 - Principais etapas da construção das bibliotecas de cDNA. Ao longo da seção materiais e métodos, cada um dos passos aqui enumerados será detalhado. Este fluxograma serve de referência para a contextualização e compreensão da sequência dos itens que serão descritos. O mapa do plasmídeo pDNR-LIB é mostrado com maiores detalhes na figura 5.



MCS A

45
loxP •
 TTA TCA GTC GAC GGT ACC GGA CAT ATG CCC GGG AAT TCG GCC ATT ACG GCC TGC AGG ATC C **Stuffer fragment**
 SalI *NdeI* *SmaI* *EcoRI* *SfiIA* *PstI* *BamHI*

MCS B 292

•
Stuffer fragment GG ATC CGG CCG CCT CGG CCC TCG AGA AGC TTT CTA GAC CAT TCG TTT GGC GCG CCG GCC CAG TAG GTA AGT GAA
 BamHI *SfiIB* *XhoI* *HindIII* *XbaI* *Bsp120I*
 ApaI

Figura 5 – Representação esquemática do plasmídeo pDNR-LIB. Em destaque, os sítios múltiplos de clonagem (**MCS A** e **MCS B**). Retirado de <<http://www.clontech.com>>.

3.1. Extração e purificação de RNA

Na primeira etapa da sequência experimental (**fig. 4, etapa 1**), o RNA total foi extraído e purificado dos hemócitos (*C. rhizophorae*) e da glândula digestiva (*C. rhizophorae* e *N. nodosus*). O RNA poli-A foi purificado apenas das amostras de RNA total da glândula digestiva que apresentavam quantidade suficiente.

Quadro 4 – Animais. As ostras utilizadas neste trabalho foram coletadas no Rio São João, em Barra de São João (RJ), onde são cultivadas. As vieiras foram cedidas pelo projeto POMAR (Projeto de Repovoamento Marinho da Baía da Ilha Grande) coordenado pelo Instituto de Ecodesenvolvimento da Baía da Ilha Grande (IED-BIG), em Angra dos Reis (RJ).

Para a extração de RNA total de hemócitos, a hemolinfa foi extraída do músculo adutor de diferentes indivíduos, formando um *pool* de 25 mL, que foi armazenado em igual volume de solução para preservação de RNA ((NH₄)₂SO₄ 55%; EDTA 1 mM; pH 5.2) a -20°C. Imediatamente antes da extração, a solução com hemolinfa foi centrifugada (10'@4000g@4°C). Após o descarte do sobrenadante, o *pellet* de hemócitos foi lavado em 25 mL de tampão PBS 1X (Salina tamponada com fosfato, Sigma) e a suspensão celular, novamente centrifugada nas mesmas condições da etapa anterior. Após o descarte do sobrenadante, o *pellet* foi totalmente resuspenso em 1 mL de QIAzol Lysis Reagent (Qiagen).

Para extração de RNA total das amostras de glândula digestiva da ostra de mangue e das vieiras, entre 50-100 mg de tecido foram totalmente homogeneizados em 1 mL de TRIZOL (Invitrogen). A partir desta etapa, todas as amostras, de hemolinfa e glândula digestiva, foram processadas de acordo com as instruções do fabricante. Ao final do processo, o *pellet* de RNA foi resuspenso em água destilada livre de nucleases (GIBCO). O RNA total foi ainda purificado com cloreto de lítio. Para cada 100 µL de RNA total resuspenso em água, foram adicionados 20 µL de cloreto de lítio (8M). Após 30 min a -20°C, as amostras foram centrifugadas (20 min@16.000xg) e o *pellet* resultante foi lavado duas vezes com 1 mL de etanol 75%. Após a última etapa de lavagem e

evaporação de todos os resíduos de etanol, o *pellet* de RNA foi ressuspensão em água destilada livre de nucleases.

Finalmente, o RNA mensageiro (poli-A) foi isolado com o kit GenElute mRNA miniprep (Sigma Aldrich) conforme as recomendações do fabricante a partir do RNA total extraído das amostras de glândula digestiva.

Quadro 5 – Quantificação e controle de qualidade do RNA. A quantidade de RNA foi avaliada pela absorvância em 260 nm. A possível contaminação por proteínas e compostos usados durante a extração (sais, fenol, clorofórmio, etc) foi avaliada pela absorvância em 280 e 230 nm, respectivamente. A integridade do RNA total foi avaliada por meio do fracionamento por eletroforese em gel de agarose em condições desnaturantes conforme descrito por Masek et al. (2005), porém em tampão TBE (Tris/Borato/EDTA) 0.5X (GIBCO). Estes e os demais géis deste trabalho foram visualizados e fotografados no sistema de fotodocumentação Molecular Imager Gel Doc XR+ System (BIORAD).

3.2. Síntese e amplificação de cDNA

O RNA isolado dos hemócitos e da glândula digestiva foi utilizado como molde para a síntese do DNA complementar (cDNA). Porém, com algumas particularidades. A primeira fita de cDNA foi sintetizada de forma a adicionar sequências adaptadoras contendo os sítios de restrição *SfiIA* e *SfiIB* através do mecanismo SMART – *Switching Mechanism At 5' end of RNA Transcript* (qua.6 e 7) (Zhu et al., 2001), sendo esta a primeira etapa para a construção das bibliotecas.

Quadro 6 – Síntese da primeira fita de cDNA.

Aproximadamente 1 µg de RNA poli-A da glândula digestiva (*C. rhizophorae* e *N. nodosus*) e aproximadamente 400 ng de RNA total dos hemócitos (*C. rhizophorae*) foram utilizados. Para cada reação, foram adicionados RNA, 1 µL do adaptador 5' SMART IV oligonucleotide 12 µM (Clontech) e 1 µL do oligodT CDS-3M adapter 12 µM (Evrogen) em um volume final de 5 µL. Todos os oligonucleotídeos utilizados encontram-se descritos na **tabela 1**. A mistura foi incubada a 72°C por 2 min e imediatamente incubada em gelo, por mais 2 min. Após esta etapa, foram adicionados 2 µL de tampão de primeira fita 5X (Clontech), 1 µL de DTT (20 mM), 1 µL de dNTPs (10 µM) e 1 µL de RevertAid M-MuL-V reverse transcriptase (Fermentas). As reações foram incubadas por 1 hora e 30 min a 42°C e armazenadas a -20°C até a próxima etapa. Componentes dos kits TRIMMER-DIRECT cDNA normalization (Evrogen) e Creator SMART cDNA library construction (Clontech) foram usados na síntese de cDNA e nas etapas seguintes da construção das bibliotecas.

Tabela 1: Oligonucleotídeos utilizados durante a construção das bibliotecas de DNA. Em negrito, o sítio de restrição **SfiIA**. Sublinhado, o sítio de restrição **SfiIB**.

Oligonucleotídeo	Sequência
SMART IV oligonucleotide	5'-AAGCAGTGGTATCAACGCAGAGT GGCCATTACGGCCGGG -3'
CDS-3M adapter	5'AAGCAGTGGTATCAACGCAGAGT <u>GGCCGAGGCGGCC</u> (T)20VN*3'
5' PCR Primer	5'-AAGCAGTGGTATCAACGCAGAGT-3'
PCR primer M1	5' AAGCAGTGGTATCAACGCAGAGT 3'
PCR primer M2	5' AAGCAGTGGTATCAACGCAG 3'
M13 Foward	5' GTAAAACGACGGCCAGT 3'
M13 Reverse	5' AAACAGCTATGACCATGTTCA 3'

*N = A, C, G or T; V = A, G or C

Quadro 7 – Mecanismo SMART – Switching Mechanism At 5' end of RNA Transcript

Este mecanismo se baseia na adição de três citosinas (C) na extremidade 3' do cDNA, ao final da retrotranscrição da molécula molde de RNA utilizada pela retrotranscriptase MML-V (vírus Maloney da leucemia- **figura 6**). Na presença de um oligonucleotídeo que contenha três guaninas (G) que anelem com a cauda de citosina do cDNA, a retrotranscriptase MML-V troca de suporte e continua a polimerizar DNA usando o oligonucleotídeo como molde. Desta forma, em uma única etapa, o cDNA é sintetizado a partir do oligodT CDS-3M e o adaptador SMART IV oligonucleotide é adicionado a extremidade 3' da fita de cDNA recém sintetizada. Como as citosinas são adicionadas somente quando a retrotranscriptase MML-V atinge o final do transcrito, o adaptador é adicionado somente às moléculas de cDNA completas e não àquelas onde a enzima tenha interrompido a polimerização prematuramente, algo frequente durante reações de retrotranscrição. O cDNA com adaptadores em ambas extremidades é amplificado por PCR resultando em um cDNA enriquecido com sequências completas e flanqueado pelos sítios de restrição apropriados para a clonagem direcional do cDNA no plasmídeo pDNR-LIB (**Fig. 4, etapa 6**).

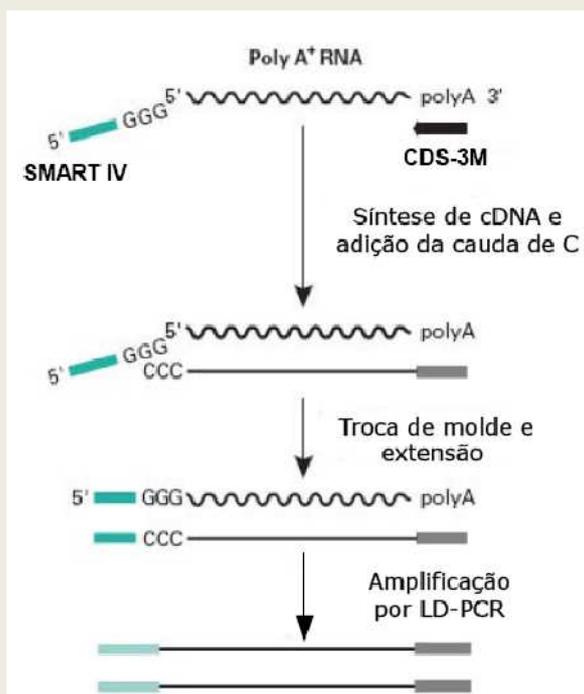


Figura 6 – Representação esquemática da “troca de molde” (“template switching”) pela retrotranscriptase MML-V durante a síntese da primeira fita de cDNA. Adaptado de <http://www.clontech.com>

3.2.1. Amplificação do cDNA por LD-PCR e regulação do tamanho médio dos produtos de PCR

A primeira fita de cDNA flanqueada pelos adaptadores CDS-3M e SMART IV (**fig. 6**) foi amplificada por PCR de longa distância (LD-PCR) para aumentar a quantidade de cDNA disponível para a etapa seguinte de normalização.

Sabe-se que durante a amplificação de cDNA por PCR, as sequências mais curtas são amplificadas com maior eficiência que as maiores. Como o cDNA é uma mistura complexa, com fragmentos de tamanhos diferentes, os menores fragmentos são amplificados com maior eficiência resultando na perda das seqüências raras e mais longas, assim como na redução do comprimento médio dos fragmentos de cDNA. Para resolver este problema foi utilizada a estratégia de regulação do tamanho médio do produto de PCR proposta por Shaginet al. (1999). Esta solução consiste em fazer um PCR de longa distância (LD-PCR) com um único *primer*.

O PCR de longa distância é otimizado para amplificar fragmentos de DNA longos com maior eficiência e fidelidade a sequênciã original. Para isso, foram utilizados reagentes do kit Advantage 2 PCR (Clontech) que contém uma mistura de duas enzimas: uma DNA polimerase de alta sensibilidade e rendimento; e uma DNA polimerase com atividade exonuclease 3'-5', para correção dos pares de bases mal pareados (qua.8). Os adaptadores SMART IV e CDS-3M adicionados ao cDNA contém um trecho onde suas sequências são iguais e ao usarmos um único primer que anela nesta região, as extremidades do produto de PCR nos ciclos iniciais conterão sequências invertidas e complementares (**fig. 7**). Estas extremidades irão hibridar uma com a outra, porém mais facilmente nas sequências mais curtas impedindo o anelamento do *primer* e, portanto, inibindo a amplificação das sequências seletivamente em função do seu tamanho. Assim, a amplificação das sequências curtas é inibida e fragmentos maiores são amplificados preferencialmente. Segundo Matz (2002), para que a biblioteca de cDNA seja representativa, ou seja, contenha ao menos uma cópia de cada molécula de RNAm presente inicialmente, o número de ciclos das etapas de PCR não deve ser maior

que 25. Um número de ciclos maior poderia resultar na amplificação preferencial de algumas sequências em detrimento de outras, reduzindo a complexidade do cDNA original. Por isso, o número de ciclos usados neste primeiro LD-PCR não ultrapassou 18 (tab.1).

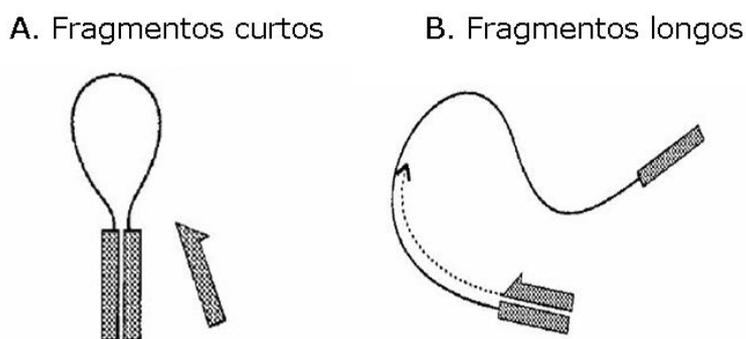


Figura 7 – Regulação do tamanho médio dos produtos do LD-PCR. Nos fragmentos de cDNA curtos (A), a amplificação é inibida devido a hibridação das extremidades do cDNA que são complementares, o que impede o acesso do primer. Nos fragmentos maiores (B), este efeito não ocorre na maior parte das moléculas resultando no anelamento do primer em sua região alvo e consequente amplificação. Adaptado de Shagin et al., (1999).

Quadro 8: Amplificação de cDNA por LD-PCR. Para cada reação de foram adicionados: 2 μ l de cDNA (produto da reação de síntese da primeira fita de cDNA), tampão 10X Advantage 2 PCR (clontech), 0,2 μ M de cada dNTP (Invitrogen), 0,4 μ M de 5' PCR primer (Tab.1, Clontech), 2 μ l de 50X Advantage 2 Polymerase Mix (Clontech) e água livre de nucleases suficiente para completar 100 μ l. As reações de LD-PCR foram incubadas em um termociclador (*GeneAmp PCR System 2400* – Perkin Elmer)de acordo com os ciclos descritos na **tabela 2**. O cDNA obtido foi purificado para remoção de primers, dNTPs e proteínas com o kit *GenElute PCR Clean-up* (Sigma Aldrich), conforme protocolo recomendado pelo fabricante.

Para observação dos perfis de cDNA obtidos, os produtos da amplificação do cDNA por LD-PCR foram analisados por eletroforese em gel de agarose. A **figura 8** mostra três diferentes perfis de cDNA: a glândula digestiva de *C. rhizophorae* com fragmentos aproximadamente entre 300 pb e 3 Kb, os hemócitos deste mesmo organismo, com fragmentos entre 200 pb e 3 Kb e a glândula digestiva de *N. Nodosus* com fragmentos desde 200 pb até maiores que 3 Kb. As amostras de cDNA de *C. rhizophorae* resultaram também na amplificação de uma banda mais intensa (entre 700 e 800 pb) enquanto que a de *N. Nodosus* resultou na amplificação de duas bandas mais intensas (de 1 e 1,5 Kb, aproximadamente) provavelmente referentes a genes constitutivos altamente expressos e, portanto, mais abundantes no cDNA original.

Quadro 9 – Eletroforese em gel de agarose. O fracionamento de cDNA por eletroforese em gel agarose (concentrações variáveis entre 1.1-1.4%, indicadas nas legendas) foram feitos com tampão TBE 0.5X (GIBCO). Brometo de etídeo foi adicionado ao gel de agarose na concentração final de 17 µg/mL.

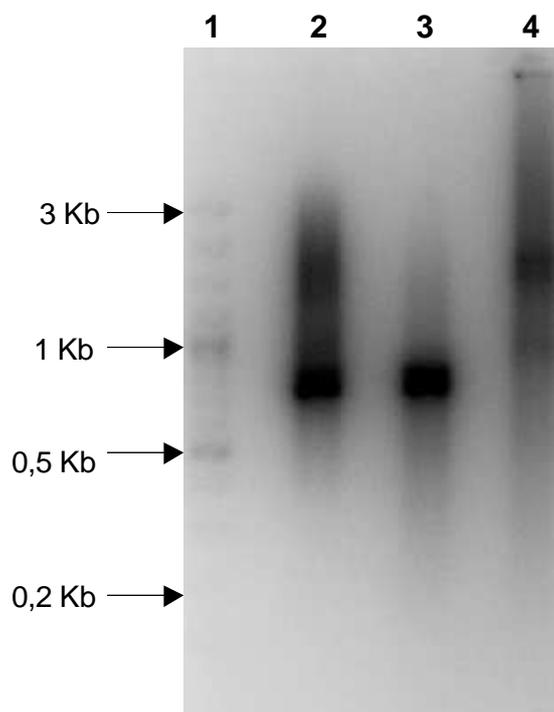


Figura 8 – cDNA amplificado por LD-PCR e fracionado por eletroforese em gel de agarose 1.4%. 1 - padrão GeneRuler 100 pb DNA ladder plus (Fermentas). 2: produto do LD-PCR da glândula digestiva de *C. rhizophorae*. 3 – produto do LD-PCR dos hemócitos de *C. rhizophorae*. 4 – Produto do LD-PCR da glândula digestiva de *N. nodosus*.

Tabela 2 – Ciclos de PCR utilizados ao longo da construção das bibliotecas de cDNA.

PCR: Descrição/Etapa	Desnaturação inicial	Número de ciclos	Desnaturação	Anelamento	Extensão	Extensão final
Amplificação de cDNA (1ª Fita) – Etapa 2	95°C (1 min)	13 (G. digestiva/ <i>C.</i> <i>rhizophorae</i>)	15" @95°C	20" @66°C	6' @72°C	
		18 (Hemócitos/ <i>C.</i> <i>rhizophorae</i>)				
		15 (G. digestiva/ <i>C.</i> <i>rhizophorae</i>)				
1ª amplificação de cDNA normalizado - Etapa 4	95°C (1 min)	20 (cDNA normalizado)	15" @95°C	20" @66°C	3' @72°C	
		17 (cDNA não normalizado)				
2ª amplificação de cDNA normalizado - Etapa 4	95°C (1 min)	12	15" @95°C	20" @64°C	3' @72°C	20" @64°C 3' @72°C
PCR de colonia - Etapa 7	95°C (1 min)	30	15" @95°C	20" @55°C	1' @72°C	5' @72°C

3.3. Normalização de cDNA

Na **etapa 3** da sequência experimental, o cDNA amplificado por LD-PCR foi normalizado pelo método descrito por Zhulidov et al., (2004), revisado por Bogdanova et al., (2008) (**qua.10**). A normalização DSN (*Duplex-Specific Nuclease*) remove as espécies de cDNA mais abundantes, evidenciadas pelas bandas mais intensas presentes no cDNA (**fig. 8**), e iguala a concentração dos diferentes cDNAs transcritos de mRNA.

Este método se baseia na cinética de reassociação do cDNA. Após a desnaturação, a taxa de hibridação de cada sequência de cDNA é proporcional ao quadrado da sua concentração na mistura. Desta forma, quanto maior a concentração de uma sequência, mais moléculas hibridarão e formarão novamente moléculas fita dupla. Conseqüentemente, durante a reassociação, o cDNA forma duas frações (**fig. 9**): **(1)** fração fita dupla onde a maior parte do cDNA contém os transcritos mais abundantes e **(2)** fração fita simples que contém cDNAs individuais em concentrações aproximadamente iguais (Young e Anderson, 1985; Galau et al., 1977).

Os métodos de normalização diferem na maneira como separam a fração dupla fita da fração fita simples (normalizada). A normalização DSN consiste na digestão enzimática da fração fita dupla durante a reassociação do cDNA. A enzima *DSN* é estável em temperaturas elevadas e digere ácidos nucléicos dupla fita (DNA:DNA ou DNA:RNA) a 68°C. Essa temperatura é elevada o suficiente para evitar tanto a formação de estruturas secundárias no cDNA fita simples como a hibridação das regiões adaptadoras, prevenindo a clivagem da fração normalizada.

Quadro 10 – Digestão do cDNA com DSN.

Para cada uma das três amostras de cDNA a serem normalizadas, 600-1200 ng de cDNA, 4 μ L de Hybridization buffer (Evrogen) e água deionizada livre de nucleases suficiente para completar o volume de 16 μ L foram misturados em um microtubo. Em seguida, alíquotas de 4 μ L foram dispensadas em 4 tubos independentes denominados tubos 1, 2, 3 e 4 (tab. 3) e cobertas com uma gota de óleo mineral. Todos os tubos foram incubados a 98°C por 2 min para total desnaturação do cDNA e, em seguida, hibridadas a 68°C por um período de 5 h (fig. 9).

Após o período de hibridação, o cDNA dupla fita foi digerido com a enzima *DSN*. Para tanto, aos tubos ainda incubados a 68°C, foi adicionado 5 μ L de DSN master buffer (Evrogen) pré-aquecido a 68°C. Após 10 min, a cada tubo experimental (tubos 1, 2 e 3) foi adicionado 1 μ L de *DSN* (foram testadas três diluições da enzima *DSN*, como descrito na tabela 3) e 1 μ L de DSN storage buffer (Evrogen) foi adicionado aos tubos de controle negativo (tubo 4). Os tubos experimentais e controle foram incubados a 68°C por mais 25 min para digestão do DNA dupla fita. Após este período, 10 μ L de DSN stop solution (Evrogen) foram adicionados em cada tubo e após mais 10 min de incubação a 68°C, todas as reações foram diluídas com 20 μ L de água destilada livre de nucleases e armazenadas a -20°C até a próxima etapa.

Tabela 3 – Descrição das diferentes diluições da enzima DSN utilizadas na etapa de normalização do cDNA.

Diluições de enzima (DSN)	Tubos			
	Experimentais			Controle negativo
	1	2	3	4
DSN	1 μ l	-	-	-
½ DSN	-	1 μ l	-	-
¼ DSN	-	-	1 μ l	-
storage buffer	-	-	-	1 μ l

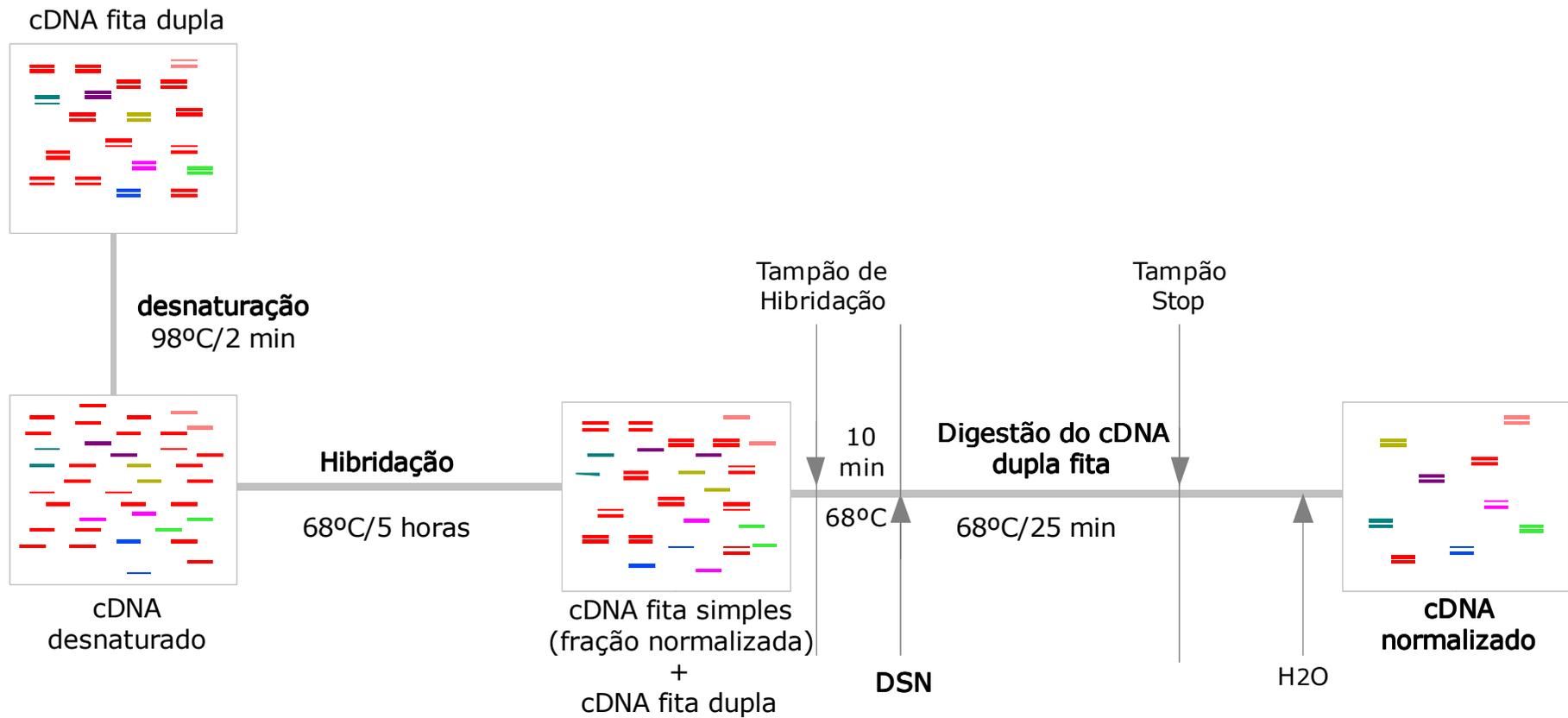


Figura 9 – Representação esquemática da normalização DSN. As sequências de cDNA em vermelho representam os cDNAs mais abundantes, enquanto que as demais cores representam sequências de cDNA individuais (fração normalizada).

3.4. Amplificação do cDNA normalizado

O cDNA resultante de todas as reações de digestão com *DSN* (tubos 1, 2, 3) e do controle negativo (tubo 4) foi também amplificado por PCR de longa distância com um único *primer* (fig.4, etapa 4), para aumentar a quantidade de cDNA disponível sem alterar a representatividade e a complexidade do cDNA normalizado (etapa 2.2).

Quadro 11 – Primeira amplificação do cDNA normalizado.

Para cada reação, adicionou-se: 1 µL de cDNA, 10X Advantage 2 PCR (clontech), 0,2 µM de cada dNTP (Invitrogen), 0,3 µM de PCR primer M1 (Tab. 1, Evrogen), 1 µL de 50X Advantage 2 Polymerase Mix (Clontech) e água deionizada suficiente para completar 50 µL. As reações foram submetidas a 7 ciclos de amplificação (tab. 2), em seguida, mantidas em gelo enquanto se determinou o número ótimo de ciclos para as amostras experimentais, como descrito a seguir.

Uma alíquota da reação de amplificação (LD-PCR) de cada cDNA controle foi retirada ao término do 7º ciclo e outras alíquotas foram retiradas a cada 2 ciclos até o 13º ciclo, conforme descrito na **figura 10**. Estas alíquotas foram analisadas por eletroforese em gel de agarose para determinar o número ótimo de ciclos adicionais a serem utilizados na amplificação das amostras experimentais (digeridas com *DSN*), de forma a não ultrapassar a fase exponencial do PCR (1 ou 2 ciclos antes de atingir o platô) e assim manter a proporção original entre os diferentes genes (normalização).

Com base nos resultados obtidos para as amostras de cDNA controle (**fig. 11**), determinou-se que estas amostras alcançaram o platô no 13º ciclo e assim, calculou-se a quantos ciclos adicionais cada amostra de cDNA experimental deveria ser submetida (de acordo com as fórmulas descritas na **figura 10**, indicadas pelo fabricante do kit - Evrogen). Assim, todas as amostras de cDNA normalizadas com *DSN* foram submetidas a mais 13 ciclos, num total de 20 ciclos, enquanto as amostras de cDNA controle (tubo 4) foram submetidas a 4 ciclos adicionais, totalizando 17 ciclos.

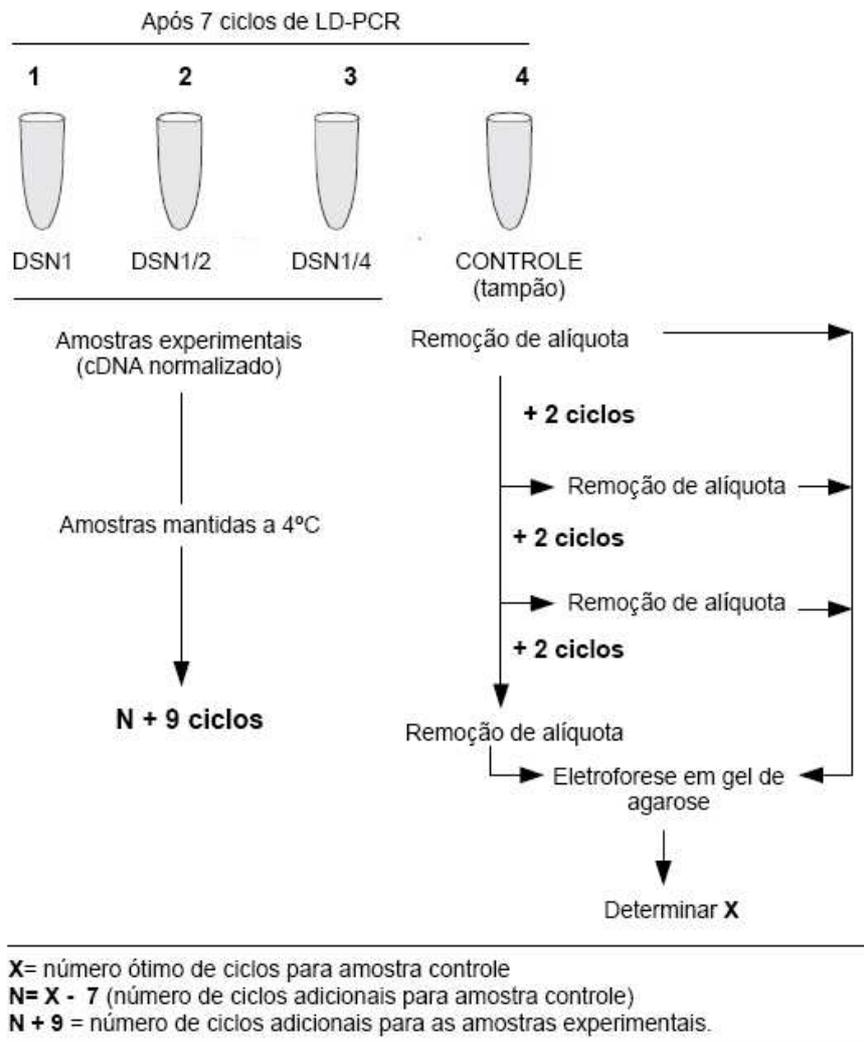


Figura 10 – Representação esquemática das etapas de otimização do número de ciclos de LD-PCR para amplificação do cDNA normalizado.

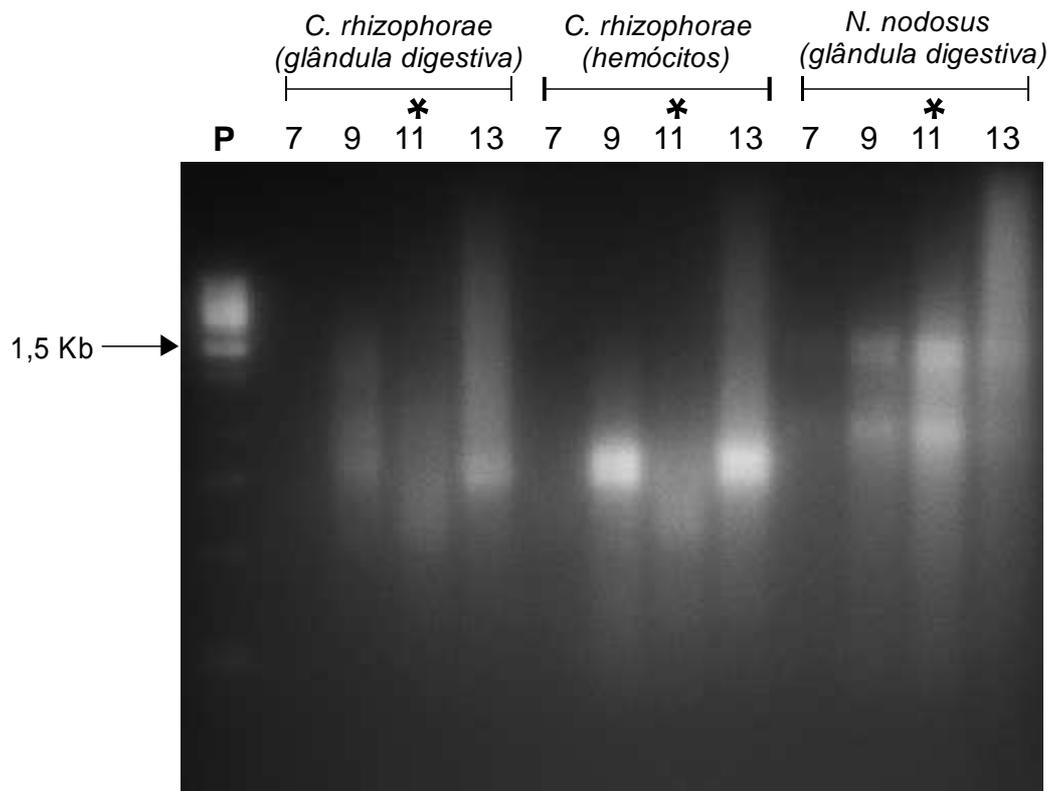


Figura 11 –Otimização do número de ciclos de LD-PCR com cDNA controle. Eletroforese em gel de agarose 1.4%. **1** – padrão GeneRuler 1 kb DNA ladder plus (Fermentas). **7; 9; 11 e 13** – alíquotas do LD-PCR das amostras controle (cDNA não digerido com DSN) após 7, 9, 11 e 13 ciclos, respectivamente. *número de ciclos ótimo para cada amostra controle.

Os produtos finais da amplificação do cDNA normalizado e dos controles negativos foram analisados por eletroforese em gel de agarose para determinar qual concentração de *DSN* forneceu os melhores resultados de normalização (**fig. 12**). Para as três amostras de cDNA, foi selecionada a reação nas quais foi utilizado 1 μ L de *DSN* não diluída (**tubo 1**) e que resultou na remoção das bandas mais intensas presentes no cDNA antes da normalização (**fig. 8**) e que menos diminuiu a faixa de tamanho dos fragmentos de cDNA em comparação com o cDNA não digerido com *DSN*. O cDNA normalizado que apresentou os melhores resultados (tubos 1) foi submetido a uma segunda amplificação por LD-PCR com um único primer (qua.12).

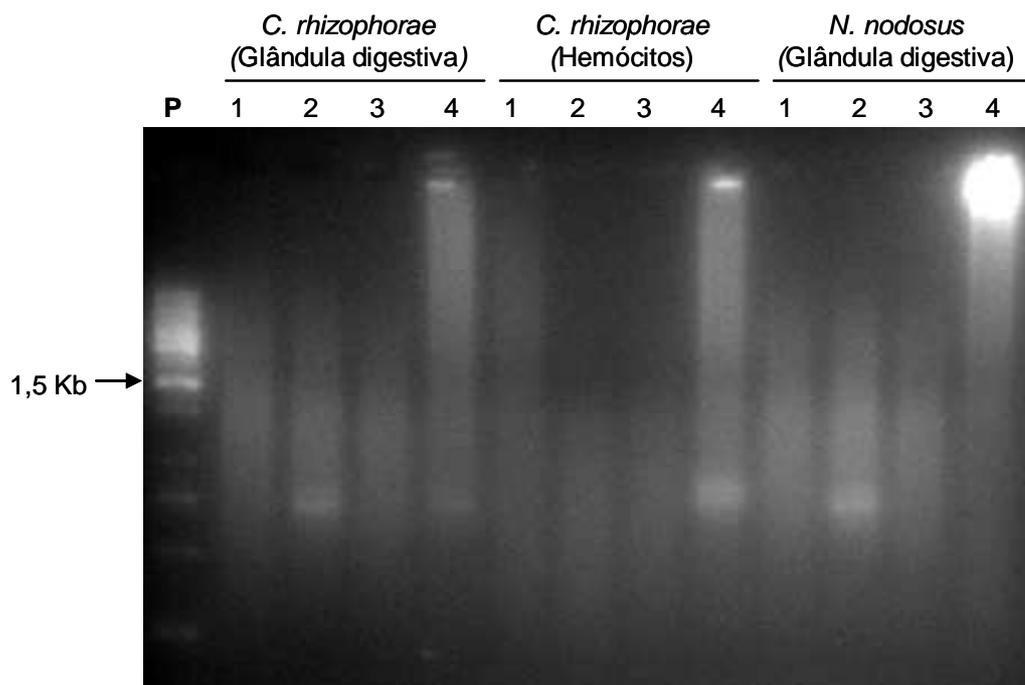


Figura 12 – Primeira amplificação do cDNA normalizado. Eletroforese em gel de agarose 1.4% P-padrão Gene Ruler 1 Kb plus DNA ladder. 1 – cDNA digerido com DSN não diluída; 2 – cDNA tratado com DSN ½; 3 – cDNA digerido com DSN ¼; 4 -cDNA não digerido com DSN.

Quadro 12. 2ª amplificação do cDNA normalizado. Para cada reação, adicionou-se: 2 µl de cDNA (diluído 10X), tampão 10X Advantage 2 PCR (clontech), 0,2 µM de cada dNTPs (Invitrogen), 0,4 µM de PCR primer M2 (Tab. 1, Evrogen), 2 µl de 50X Advantage 2 Polymerase Mix(Clontech) e água deionizada suficiente para completar 100 µl. As reações foram submetidas a 12 ciclos de amplificação (**tabela 2**) e uma alíquota de cada uma das reações foi analisada por eletroforese em gel de agarose (**figura 13**). Por fim, o cDNA foi digerido com *proteinase K* e purificado com Fenol/Clorofórmio conforme o protocolo recomendando pelo fabricante (clontech). Ao final deste processo, cada *pellet* de cDNA foi ressuspenso em 79 µl de água livre de nucleases.

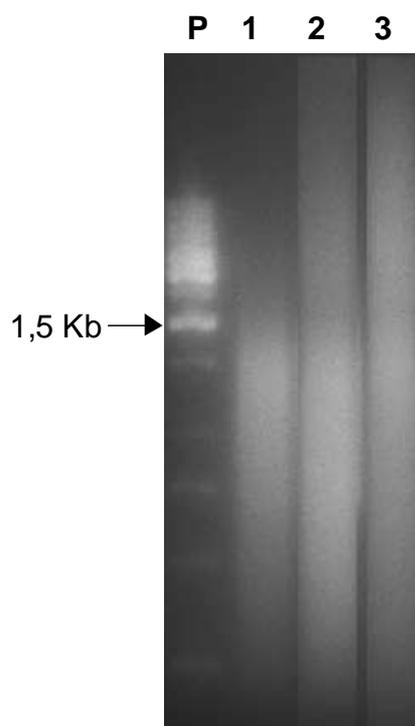


Figura 13 - Produtos da 2ª amplificação do cDNA normalizado fracionados por eletroforese em gel de agarose. P – padrão Gene Ruler 1 Kb DNA ladder plus (fermentas); 1 – cDNA normalizado da glândula digestiva de *C. rhizophorae*; 2 – cDNA normalizado dos hemócitos de *C. rhizophorae*; 3 – cDNA normalizado da glândula digestiva de *N. nodosus*.

3.5. Digestão com enzima de restrição (*Sfi*) e fracionamento do cDNA normalizado

O cDNA normalizado e purificado foi digerido com a enzima de restrição *Sfi* (clontech) (fig. 4, etapa 4).

Quadro 13 – Reação de digestão de cDNA com *Sfi*.

Aos 79 μ L de cDNA adicionou-se: 10 μ L de *Sfi* buffer 10X (clontech), 1 μ L de BSA (albumina do Soro Bovino) 100X (clontech), 10 μ L de *Sfi* (20 unidades/ μ L, clontech). As reações foram incubadas a 50°C por 2 h. Após este período, 2 μ L de Xylene Cyanol FF (1%) foram adicionados a cada reação.

Cada uma das amostras de cDNA digeridas com *Sfi* foi fracionada em colunas CHROMA SPIN 400 (Clontech), conforme recomendações do fabricante, para remoção dos fragmentos de cDNA pequenos (<250pb) e que seriam clonados com maior eficiência que os maiores na etapa subsequente. Em cada fracionamento, 16 frações foram coletadas e analisadas por eletroforese em gel de agarose para identificação das frações que continham cDNA na faixa de tamanho de interesse. As três frações que continham fragmentos de cDNA maiores que 250 pb, indicadas por asteriscos na **figura 14**, foram misturadas e precipitadas.

Quadro 14 – Precipitação do cDNA fracionado.

Em cada amostra, foi adicionado de 200 μ L de etanol absoluto, 1.3 μ L de glicogênio (20 μ g/ μ L) e 8 μ L de acetato de sódio 3 M. As soluções precipitantes foram mantidas a -20°C por 1 hora e, em seguida, o cDNA foi coletado por centrifugação (20 min@14000xg) e os *pellets* ressuspensos em 7 μ L de água livre de nucleases.

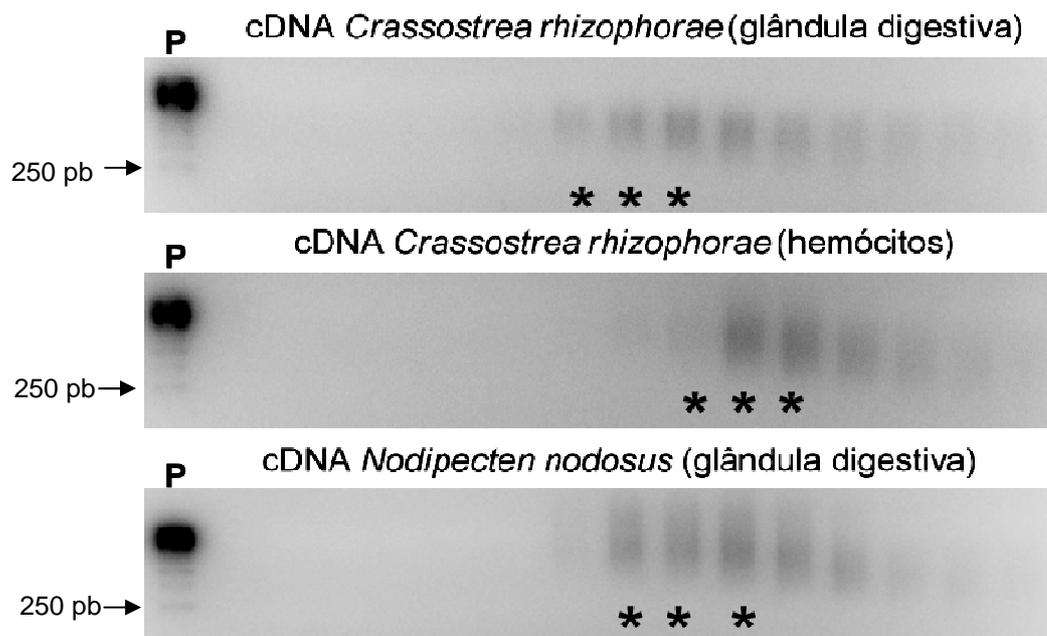


Figura 14 – Eletroforese em gel de agarose 1.1% das alíquotas de cDNA obtidas após fracionamento. **P** – padrão Gene Ruler 1 Kb DNA ladder. As frações que apresentaram cDNA com tamanho superior a 250 pb (marcadas com asteriscos) foram selecionadas para clonagem.

3.6. Reação de Ligação do cDNA ao vetor de clonagem pDNR-LIB.

Após as etapas de normalização, digestão com *SfiI* e fracionamento, o cDNA obtido foi ligado ao plasmídeo pDNR-LIB (**fig. 4, etapa 6**). A clonagem foi feita entre os sítios de restrição *SfiIA* e *SfiIB* do pDNR-LIB (**fig. 5**) de forma direcional, ou seja, com uma única orientação possível para a ligação dos insertos de cDNA no plasmídeo.

Quadro 15– Reações de Ligação cDNA-pDNR-LIB. Para cada amostra de cDNA, três reações de ligação independentes foram feitas com três volumes de cDNA distintos: 0,5; 1,0 e 1,5 μL . Desta forma, três proporções de cDNA/vetor foram utilizadas, como descrito na tabela 4. Além do cDNA, em cada reação também foi adicionado: 0,5 μL de 10X ligation buffer (clontech); 0,5 μL de ATP (10 mM), 0,5 μL de *T4 DNA Ligase* (clontech); 1 μL de pDNR-LIB (100 ng/ μL) e água deionizada suficiente para completar o volume de 5 μL . As reações foram incubadas cerca de 18 h a 16°C. Após este período o DNA plasmidial foi precipitado e os *pellets* foram ressuspensos em 3 μL de água livre de nucleases.

Tabela 4 – Bibliotecas de cDNA e as respectivas proporções de cDNA/vetor utilizadas na reação de ligação.

Organismo/Tecido	Nome da Biblioteca	cDNA/pDNR-LIB (ng/ng)
<i>C. rhizophorae</i> / Glândula digestiva	GD.Cr.1	70/100
	GD.Cr.2	140/100
	GD.Cr.3	210/100
<i>C. rhizophorae</i> / Hemócitos	H.Cr.1	90/100
	H.Cr.2	180/100
	H.Cr.3	270/100
<i>N. nodosus</i> / Glândula digestiva	GD.Nn.1	70/100
	GD.Nn.2	140/100
	GD.Nn.3	210/100

3.7. Transformação das bibliotecas de cDNA em *Escherichia coli*, titulação e avaliação inicial da proporção de clones recombinantes.

As bibliotecas de cDNA, produtos das reações de ligação descritas na **tabela 3**, foram inseridas em células de *Escherichia coli* (DH5- α) por eletroporação (**etapa 7**).

Quadro 16 – Transformação. Os produtos das reações de ligação (3 μ L) foram misturados com 40 μ L de células eletrocompetentes em um microtubo gelado e incubado em gelo por 30 s. A mistura de DNA e células competentes foi transferida para uma cubeta de 1 mm gelada e submetida a um pulso elétrico (1800V). Em seguida, se adicionou 1 mL de meio 2XTY (1.6% de Triptona, 1% de extrato de levedura, 0.5% de cloreto de sódio) e incubou-se a suspensão celular a 37°C com agitação durante 45 min.

Para determinar o título de cada uma das bibliotecas, ou seja, o número de clones independentes, diluições seriadas (10^{-4} , 10^{-5} , 10^{-6} e 10^{-7}) das suspensões celulares foram plaqueadas em meio LB-agar (Invitrogen) com 30 μ g/mL de clorafenicol (Sigma-Aldrich), para seleção das células transformantes com pDNR-LIB. As placas foram incubadas a 37°C por cerca de 18 h para contagem das colônias transformantes. As bibliotecas foram armazenadas a -80°C após adição de glicerol (na concentração final de 20%) nas suspensões celulares.

Entre 8 e 16 colônias foram aleatoriamente selecionadas de cada uma das bibliotecas para determinar a proporção de clones recombinantes (com insertos de cDNA) e o tamanho médio dos insertos por PCR de colônia (qua.15). Cada colônia foi dissolvida em 25 μ L de água destilada estéril e as células foram lisadas a 98°C por 2 min. Os lisados foram centrifugados a velocidade máxima por 2 min em uma microcentrifuga de bancada e 20 μ L do sobrenadante foram transferidos para um novo tubo.

Quadro 17 – PCR de colônia. Para cada reação adicionou-se em um microtubo: 1 μ L de DNA plasmidial (sobrenadante do lisado), 17.9 μ L de água livre de nucleases, tampão de PCR 10X, 1.5 mM de $MgCl_2$, 0.2 μ M de cada dNTP (Invitrogen), 0.4 μ M de M13 Forward, 0.4 μ M de M13 Reverse, 0.14 μ L de BSA (Fermentas) e 0.2 μ L de DNA polimerase. As reações foram incubadas no termociclador onde foram submetidos a sequência de ciclos descritos na **tabela 2**. Os produtos dos PCRs de colônia foram analisados por eletroforese em gel de agarose, a exemplo do gel da figura 15.

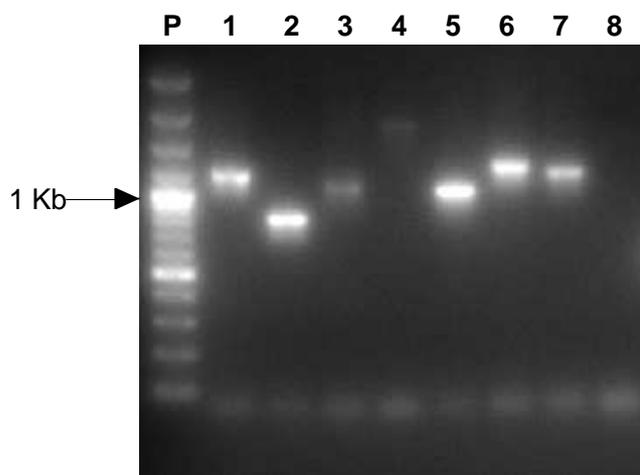


Figura 15 – Produtos do PCR de colônia fracionados por eletroforese em gel de agarose. P – padrão Gene Ruler 100 pb plus DNA ladder (Fermentas). 1-8 Produtos do PCR de 8 colônias isoladas da biblioteca GD.Cr.1.

3.8. Extração de DNA plasmidial e sequenciamento.

Noventa e seis colônias foram selecionadas aleatoriamente para sequenciamento (qua. 16), sendo 33 da biblioteca **GD.Cr.2**, outras 33 da biblioteca **GD.Nn.3** e 30 da biblioteca **H.Cr.1**.

Quadro 18 –As colônias foram inoculadas em 2 mL de meio LB (Invitrogen) com 30 µg/µL de cloranfenicol e mantidas a 37°C durante cerca de 18 h com 250 RPM de agitação. O DNA plasmidial foi então extraído com Kit QuickLyse Miniprep (QIAGEN), conforme as recomendações do fabricante. As 96 amostras de DNA plasmidial foram sequenciadas pela empresa Macrogen (Coréia) com *primer* M13forward em um sequenciador modelo 3730XL (Applied Biosystems).

3.9. Processamento das Sequências e anotação funcional de genes.

As sequências obtidas foram analisadas seguindo a sequência de etapas descritas na figura 16.

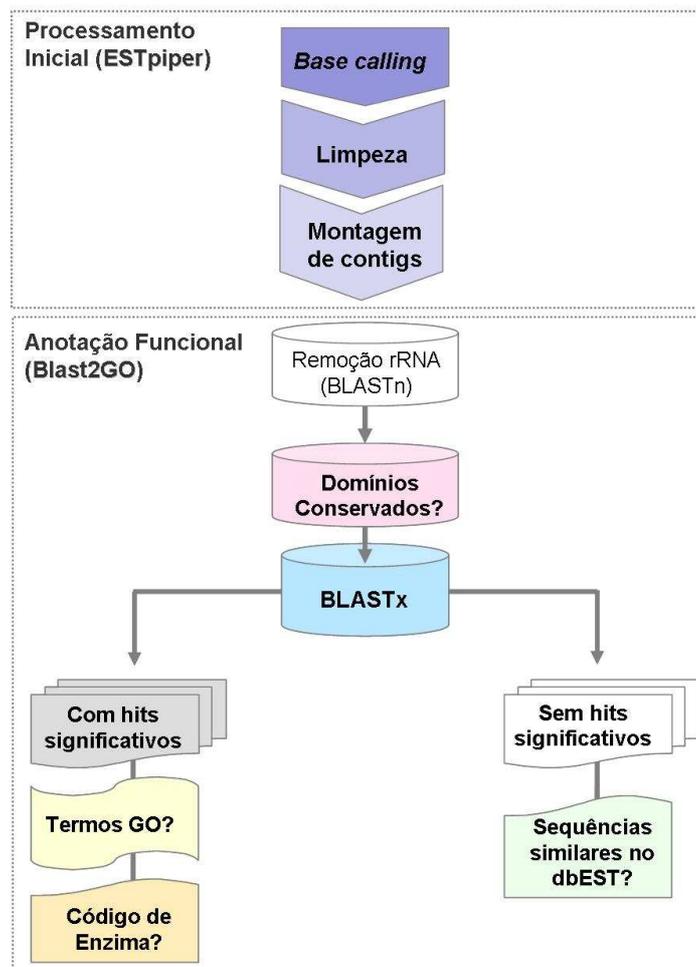


Figura 16 – Principais etapas da análise de sequências. A ferramenta online “ESTpiper” (Tang et al., 2009) foi utilizada para as etapas de conversão dos cromatogramas em sequências de nucleotídeos (“*base calling*”), limpeza das sequências (remoção dos trechos de baixa qualidade de sequenciamento e referentes ao vetor pDNR-LIB) e montagem de *contigs*. O programa Blast2GO (Conesa et al., 2005) foi utilizado na identificação de sequências similares através de algoritmos BLAST e na anotação funcional com termos de ontologia gênica (GO), códigos de enzimas e domínios funcionais conservados.

Durante o processamento inicial, as sequências foram extraídas a partir dos cromatogramas e trechos de baixa qualidade e regiões referentes ao plasmídeo usado na clonagem, removidos. Antes da etapa de anotação funcional, as sequências referentes a um mesmo gene foram ainda agrupadas em *contigs*

As sequências já processadas foram submetidas a buscas por similaridade contra o banco de dados não redundante de nucleotídeos (NCBI) usando o algoritmo BLASTn. O objetivo desta etapa foi identificar e remover os possíveis clones de RNAs ribossomais. Em seguida, com as sequências remanescentes, foram realizadas buscas por similaridade contra o banco de dados não redundante de proteínas (NCBI) usando o algoritmo BLASTx. Os '*hits*', termo usado para indicar as sequências dos bancos de dados similares às sequências submetidas, que apresentaram valor $e < -5$ foram considerados significativos. O programa Blast2GO (Conesa et al., 2005) mostrou os termos de ontologia gênica e códigos de enzima (**qua. 19**) associados aos *hits*. Para considerar a associação das anotações funcionais dos *hits* válidas também para as sequências submetidas, este programa considera não só a similaridade e valor e entre as sequências, como também as evidências que foram utilizadas na anotação de cada termo GO nos respectivos *hits*. Evidências experimentais são privilegiadas enquanto análises computacionais e outras evidências indiretas são preteridas. Essas variáveis podem ser ajustadas para dar maior especificidade e estringência ao processo. Neste trabalho, utilizamos os parâmetros descritos por Götz et al. (2009) como o "estilo de anotação" "Default (GOw:5)", limitando a anotação entre sequências similares com valor $e < 10^{-6}$.

Todas as sequências, com ou sem *hits* significativos, foram analisadas quanto a presença de domínios conservados através do InterproScan do EMBL-EBI (*European Molecular Biology Laboratory - European Bioinformatics Institute*), acessado pelo Blast2GO, e por buscas manuais no banco de domínios conservados do NCBI (CDD, Conserved Domains Database). As sequências sem *hits* obtidos através do BLASTx ou BLASTn e sem domínios conservados foram submetidas a buscas por similaridade com o algoritmo BLASTn contra o banco de sequências expressas (EST_others) do NCBI.

O programa SignalIP (disponível em <http://www.cbs.dtu.dk/services/SignalIP/>) foi

utilizado para identificar peptídeos sinais e sítios de clivagem nas sequências de proteínas, quando necessário.

Quadro 19 - Termos de Ontologia Gênica (GO) e Códigos de Enzima.

Os termos de ontologia gênica obedecem a um vocabulário controlado e hierarquizado, elaborado pelo Consórcio de Ontologia Gênica (<<http://www.geneontology.org/>>). Cada termo está associado a outros mais gerais (“termos parentais”) e outros mais específicos (“termos filhos”). Os termos GO são ainda organizados em três categorias, descrevendo o processo celular (P), o componente celular (C) e a função molecular (F) associadas a um gene/proteína (Ashburner et al, 2000).

O código de enzima ou número EC identifica enzimas pelas reações que catalizam. Com este número é possível visualizar, através do banco de vias “Kegg pathways” (<<http://www.genome.jp/kegg/pathway.html>>), as reações catalizadas pela enzima e as vias metabólicas das quais participa.

Por fim, todas as 82 sequências de alta qualidade obtidas foram depositadas na divisão dbEST do genbank, por exceção daquelas identificadas como RNA ribossomais, depositadas na divisão de nucleotídeos deste mesmo banco de dados. Os códigos de acesso das sequências estão listados nos Apêndices A, B, C, D e E.

4. RESULTADOS E DISCUSSÃO

4.1. Avaliação inicial de parâmetros de qualidade das bibliotecas de cDNA normalizadas

Bibliotecas de cDNA normalizadas foram geradas a partir da glândula digestiva de *Nodipecten nodosus* e da glândula digestiva e hemócitos de *Crassostrea rhizophorae*. Para cada um destes três tecidos, foram construídas 3 bibliotecas, a partir de quantidades diferentes de cDNA. Para selecionar aquelas com os melhores resultados para a etapa de sequenciamento, alguns parâmetros de qualidade, descritos na **tabela 4**, foram analisados. De forma geral, o número de clones independentes foi da ordem de 10^6 a 10^7 cfu (unidades formadoras de colônias) em cada biblioteca. O tamanho médio dos insertos variou de 730 pb em uma das bibliotecas de cDNA dos hemócitos da ostra (H.Cr.3) até 1458,3 pb em uma das bibliotecas do coquile (DG.Nn.3). Estes resultados demonstram o enriquecimento das bibliotecas com cDNAs longos (maiores que 250 pb) possivelmente completas. O tamanho da sequência de cDNA é especialmente importante para a sua posterior identificação com base na sua similaridade com outras sequências conhecidas, de forma que quanto maior a sequência, maiores as chances de se identificá-la e caracterizá-la (Das et al., 2001).

A proporção de clones recombinantes, ou seja, a quantidade de clones que contém um inserto de cDNA, é de crucial importância para o sequenciamento em grande escala de uma biblioteca para que não se perca tempo e reagentes com a análise de clones sem insertos. Como observado na **tabela 5**, esta proporção variou entre 62,5 e 93,75% entre as diferentes bibliotecas segundo as estimativas feitas por PCR de colônia. Com base nesses resultados, as bibliotecas **Gd.Cr.2**, **H.Cr.1** e **Gd.Nn.3** foram selecionadas para sequenciamento por terem apresentado as maiores proporções de clones recombinantes e/ou tamanho médio de insertos.

Tabela 5 – Avaliação inicial dos principais parâmetros de qualidade das bibliotecas de cDNA. As bibliotecas de cDNA selecionadas para sequenciamento encontram-se destacadas em negrito.

Biblioteca	Número de Clones Independentes (cfu)	Proporção de Clones Recombinantes (%)	Tamanho* médio dos insertos (pb)	Intervalo de tamanho* dos insertos (pb)	Número de clones analisados
GD.Cr.1	2x10 ⁷	87,5	1164,3	800-1750	8
GD.Cr.2	1,2x10 ⁷	93,75	1267,6	600-2000	16
GD.Cr.3	1,2x10 ⁷	62,5	1160	800-1750	8
H.Cr.1	9,6x10 ⁶	75	1233,3	1000-1750	16
H.Cr.2	8x10 ⁶	80	953,8	500-1500	8
H.Cr.3	1x10 ⁷	62,5	730,0	250-1200	8
GD.Nn.1	7x10 ⁶	87,5	1021,4	300-1750	8
GD.Nn.2	7x10 ⁶	75	1458,3	1200-1750	8
GD.Nn.3	2x10 ⁶	93,75	1406,7	250-2500	16

cfu: unidades formadoras de colônia

*Tamanho aproximado estimado por eletroforese em gel de agarose dos insertos amplificados por PCR de colônia

4.2. Limpeza, agrupamento e montagem das sequências

A **tabela 6** mostra o número de clones isolados para sequenciamento a partir de cada uma das bibliotecas de cDNA selecionadas. Um total de 96 clones foram sequenciados e, após a remoção dos trechos de baixa qualidade de sequenciamento e das sequências do vetor e adaptadores, 85 sequências de alta qualidade foram obtidas. Todas as sequências das bibliotecas de cDNA da glândula digestiva de *C. rhizophorae* e de *N. nodosus* continham insertos de cDNA, enquanto que a biblioteca de cDNA dos hemócitos da ostra de mangue apresentou 3 (11,1%) sequências sem insertos. No entanto, em todas elas, a proporção de clones recombinantes foi superior àquela determinada pelo PCR de colônia, descrita na **tabela 5**.

Tabela 6 – Estatísticas do processamento inicial (limpeza e montagem de *contigs*) das sequências

	Biblioteca			Total
	Gd.Cr.2	H.Cr.1	Gd.Nn.3	
Número total de clones sequenciados	33	30	33	96
Número de ESTs de alta qualidade	30	27	28	85
Número de <i>singlets</i>	30	23	28	81
Número de <i>contigs</i>	0	1	0	1
Número de clones sem inserto	0	3	0	3

Após a etapa de limpeza (**fig. 17**), a maior parte das sequências obtidas apresentaram entre 700 e 800 pb. Porém, como o limite de leitura do sequenciador utilizado é em torno de 850 pb e o tamanho médio dos insertos determinado por PCR de colônia era superior a este valor, na maior parte dos casos, as sequências obtidas não devem representar o inserto de cDNA completo. Nestes casos, seria necessário sequenciar os mesmos clones utilizando também o *primer* antisense para, então, obter a sequência completa do inserto.

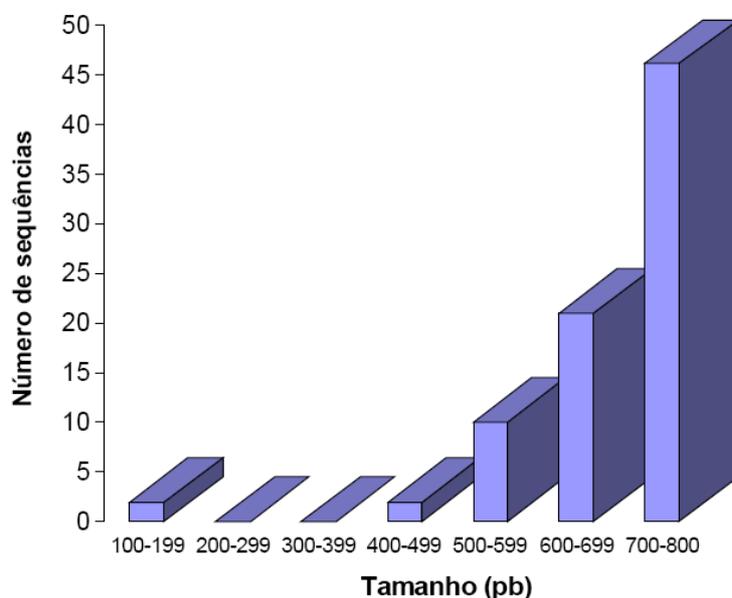


Figura 17 – Distribuição das sequências obtidas em função do tamanho em pares de base (pb). Foram considerados os tamanhos das sequências após a etapa de limpeza das sequências resultantes das 3 bibliotecas de cDNA selecionadas para sequenciamento.

Com os clones sequenciados até o momento, a biblioteca de cDNA dos hemócitos da ostra de mangue foi a única a apresentar alguma redundância (14,8%) com 4 sequências (montadas em um *contig*) referentes ao RNA ribossomal 16S. As bibliotecas da glândula digestiva da ostra e da vieira resultaram apenas em sequências únicas (nenhuma redundância). Também não houve redundância entre as duas bibliotecas da ostra e, portanto, para esta espécie foi obtido um total de 54 sequências únicas. Outros trabalhos que consistiram na construção de bibliotecas de cDNA de bivalves resultaram em bibliotecas com redundâncias tão elevadas quanto 55% (Roberts et al., 2009), 72.7% (Gestal et al., 2007), 78.25% (Pallavicini et al., 2008) e até 100% (Venier et al., 2009). As menores redundâncias foram obtidas em trabalhos que também utilizaram a normalização DSN durante a construção das bibliotecas. Tanguy et al. (2008) utilizaram esta abordagem de normalização para construir bibliotecas de cDNA para quatro espécies de bivalves e obtiveram bibliotecas com em média 7% de redundância. Da mesma forma, Quilang et al. (2007) construíram uma biblioteca de cDNA normalizada de *C. virginica* e obtiveram 15.4% de redundância.

A redundância das bibliotecas será determinada com maior precisão com o sequenciamento de um maior número de clones, no entanto, os resultados obtidos indicam que a normalização do cDNA foi efetiva. A baixa redundância em conjunto com o elevado número de clones independentes obtidos garantem a representatividade das bibliotecas de cDNA obtidas em relação ao conjunto de moléculas de mRNA/cDNA presentes na amostra original.

4.3. Caracterização e anotação funcional das sequências

4.3.1. Identificação de sequências similares

As 82 sequências únicas e de alta qualidade obtidas foram submetidas a buscas por similaridade usando o algoritmo BLASTx contra o banco de dados não redundante de proteínas e BLASTn, contra o banco de dados não redundante de nucleotídeos e o

banco de dados de ESTs do NCBI. Os resultados de BLAST de cada uma das sequências que apresentaram *hits* significativos, incluindo valor *e*, similaridade e descrição do melhor *hit*, estão descritos nos APÊNDICES A e B, para *N. nodosus* e C e D, para *C. rhizophorae*.

A **figura 18** mostra a proporção de sequências que exibiram similaridade significativa (valor $e < -5$) com outras sequências em cada uma das bibliotecas de cDNA analisadas. A biblioteca de cDNA da glândula digestiva da ostra apresentou a maior proporção de sequências com *hits* de BLASTx (70%). Mais da metade das sequências de *N. nodosus* apresentaram *hits* (64%), sendo 43% deles obtidos por BLASTx contra o banco de dado de proteínas. A biblioteca de cDNA dos hemócitos da ostra apresentou a menor proporção de sequências com *hits* de BLASTx (17%) e o maior número de sequências sem *hits* contra todos os bancos de dados (50%).

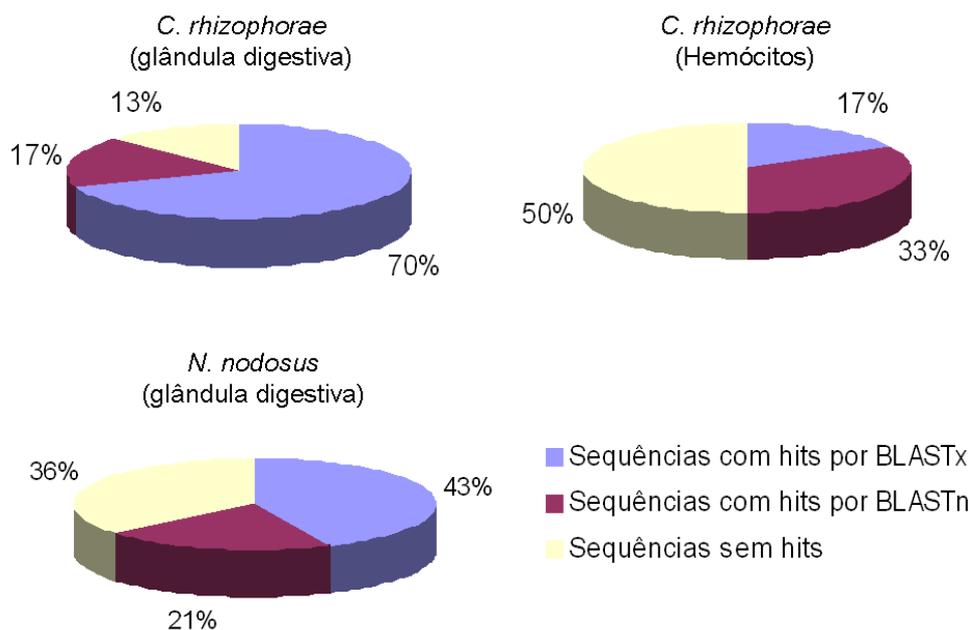


Figura 18 – Proporção de sequências com *hits* significativos obtidos por BLAST em cada uma das três bibliotecas de cDNA sequenciadas.

Um grande número de sequências sem *hits* é comum na análise de sequências de bivalves. Como exemplos, Roberts et al. (2009) construíram uma biblioteca de cDNA a partir de hemócitos plaqueados de *C. gigas* e encontraram 66% de sequências sem *hits*. Quilang et al. (2007) encontraram 56% de sequências sem *hits* em uma biblioteca de cDNA de diferentes tecidos de *C. gigas*. Em parte, isso se deve ao fato de que este grupo ainda está pouco representado nos bancos de dados de sequências nucleotídicas e protéicas. Como muitos genes ainda não foram descritos em bivalves, são identificados somente aqueles que são conservados em outros organismos presentes nos bancos de dados, em grande parte pertencentes a grupos taxonômicos distantes evolutivamente e divergentes quanto as sequências de aminoácidos de invertebrados. Mas esta não é a única razão.

O sequenciamento de genomas eucarióticos tem mostrado que em todas as espécies há sempre uma proporção de genes, em torno de 10-20%, que não apresenta similaridade com nenhum outro gene já descrito, mesmo entre organismos evolutivamente muito próximos (Khalturin et al., 2009). Tais genes têm sido chamados “genes órfãos” e recebido pouca atenção, sendo suas funções ainda pouco conhecidas. Wilson et al., (2005) introduziram um termo mais adequado: “genes taxonomicamente restritos” (TRGs - “*Taxonomically-restricted genes*”), uma vez que eles estariam restritos a um grupo taxonômico ou mesmo a uma única espécie. Os TRGs seriam particularmente importantes evolutivamente por estarem relacionados a novidades evolutivas que foram selecionadas em cada grupo ou espécie ao longo da evolução. Desta forma, os TRGs estariam relacionados com a geração da diversidade biológica que permite aos organismos estarem adaptados a diferentes nichos ecológicos (Khalturin et al., 2009).

Espécies de diferentes táxons apresentam um número significativo de genes “órfãos” que codificam potenciais peptídeos antimicrobianos (Khalturin et al., 2009). Estes peptídeos apresentam grande diversidade estrutural e de sequência de aminoácidos e são definidos, de forma geral, pelo caráter catiônico, tamanho pequeno e hidrofobicidade relativamente alta (Powers e Hancock, 2003; Yount e Yeaman, 2003), o que torna a identificação dos mesmos através de buscas por similaridade entre

sequências difícil, especialmente considerando-se que ainda são poucos os peptídeos antimicrobianos identificados em bivalves. Pallavicini et al., (2008) construíram bibliotecas de cDNA a partir de hemócitos de *Mytilus galloprovincialis* imuno estimulados, simulando infecções por bactérias e vírus e encontraram entre 25.9 e 43.4% de peptídeos antimicrobianos entre as sequências obtidas, inclusive uma nova forma, altamente polimórfica, chamada miticina C. Em um trabalho posterior, Costa et al., (2009) detectaram uma elevada diversidade do gene da miticina C entre indivíduos de *M. galloprovincialis* em comparação com um gene não relacionado com o sistema imune (β -actina). Recentemente, outros genes altamente polimórficos possivelmente envolvidos com o reconhecimento de padrões moleculares associados a patógenos foram identificados em invertebrados, como por exemplo, as proteínas relacionadas ao fibrinogênio de *Biomphalaria glabrata* (Zhang et al., 2004) e o gene altamente polimórfico FuHC (Ligante de histocompatibilidade secretado) de *Botryllus schlosseri* (De Tomaso, 2009). Esta diversificação de alguns genes do sistema imune inato pode ser explicada pelo fato de que cada indivíduo, espécie ou grupo taxonômico é exposto a uma microbiota específica de acordo com o ambiente em que vivem, onde devem estar aptos a se defenderem contra o ataque de diferentes grupos de microorganismos que acabam por moldar a evolução destes genes (Khalturin et al., 2009). Este pode ser um dos motivos para a observação de uma maior proporção de genes sem *hits* observadas no grupo de sequências isoladas de hemócitos da ostra neste trabalho, assim como em outros estudos semelhantes.

4.3.2. Anotação funcional das sequências

As sequências que exibiram similaridade com outras pelo BLASTx tiveram seus *hits* mapeados quanto a anotação de termos de ontologia gênica (*GO – gene ontology*). O programa Blast2GO foi utilizado para validar a transferência dos termos GO mapeados para as sequências analisadas. Também foram anotados os códigos de enzima (EC, *enzyme code*) existentes. Todas as sequências, com ou sem *hits*, foram submetidas a

buscas por domínios conservados e outras assinaturas de proteínas usando a ferramenta InterproScan do EMBL-EBI e a ferramenta de buscas do banco de domínios conservados (CDD) do NCBI.

A **figura 19** mostra a proporção de sequências que foram anotadas com termos GO e/ou assinaturas de proteínas ou que permaneceram sem anotações funcionais em cada uma das bibliotecas de cDNA analisadas. As sequências isoladas da glândula digestiva da ostra apresentaram a maior proporção de genes anotados (63,33%), sendo a maior parte das sequências (43%) anotadas com termos GO. Entre as sequências da glândula digestiva de *N. nodosus*, 28,57% foram anotadas funcionalmente, sendo 17,85% do total de sequências, anotadas com termos GO. A biblioteca de cDNA dos hemócitos de *C. rhizophorae* teve somente 12,5% das sequências anotadas com termos GO e assinaturas de proteínas, refletindo o baixo número de sequências com *hits* no BLASTx. Um total de 105 termos GO foram atribuídos às sequências das três bibliotecas de cDNA analisadas com uma média de 4,7 termos GO por sequência. Todas as sequências com anotações funcionais estão descritas nos **APÊNDICES A** (*N. nodosus*) e **C** (*C. rhizophorae*).

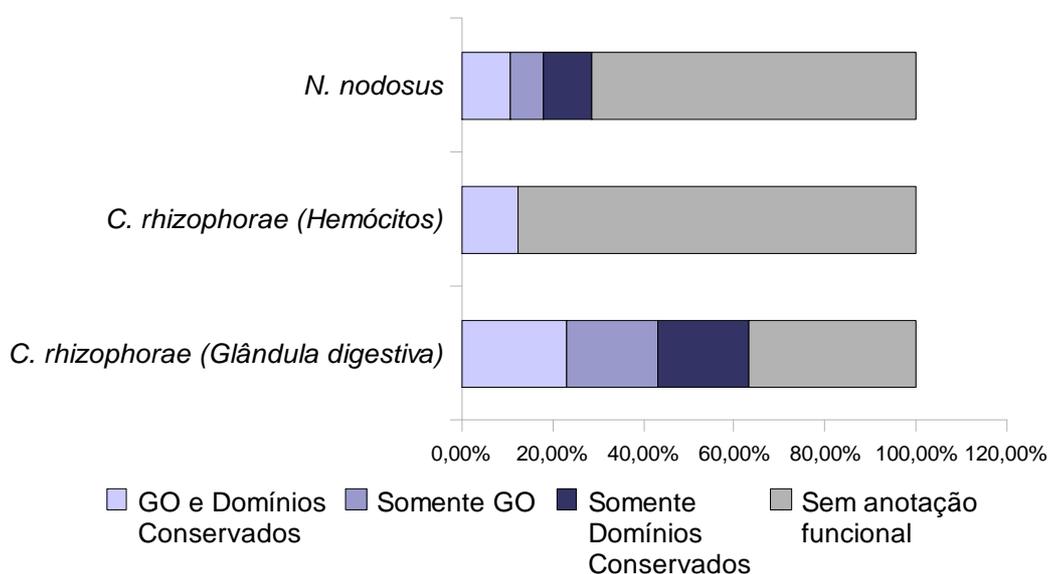


Figura 19 – Proporção de genes com anotações funcionais (termos GO e/ou domínios conservados) em cada biblioteca de cDNA sequenciada.

De forma geral, para que duas proteínas possam ser consideradas homólogas é

necessário pelo menos 30% de similaridade. Noventa por cento das sequências com similaridade maior que esse valor se revelam de fato homólogas, contra apenas 10% das sequências com similaridade inferior a 25% (Rost, 1999). Com exceção da sequência do clone D12, discutido na seção seguinte, todas as sequências aqui identificadas com anotações funcionais apresentaram similaridade superior a 30%, com valores entre 52,3% e 96,2% (valores *e* entre -6 e -81) para *C. rhizophorae* e 25% e 88,3% (valores *e* entre -7 e -93) para *N. nodosus* (**fig. 20**).

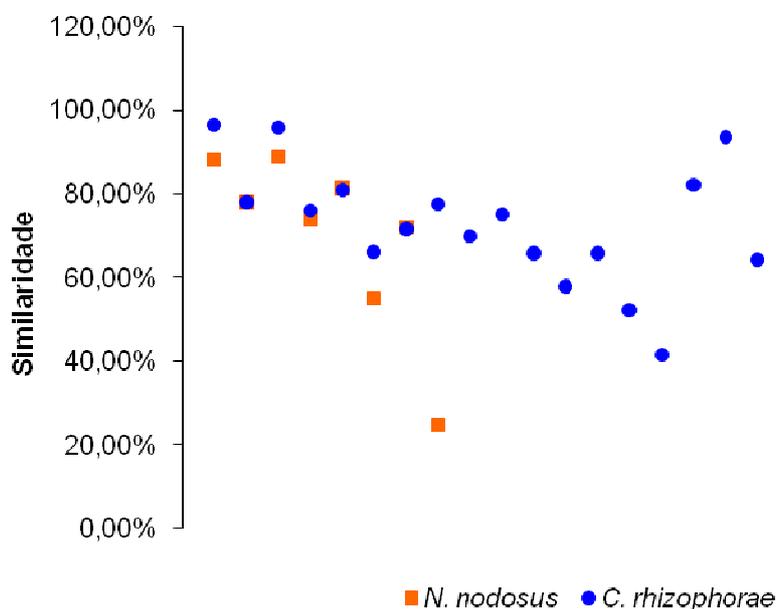


Figura 20 - Distribuição de valores de similaridade encontrados entre as sequências de *C. rhizophorae* e *N. nodosus* com anotações funcionais e seus melhores *hits*.

4.3.3. Identificação de genes de *Nodipecten nodosus*

Oito novos genes de *N. nodosus* foram identificados funcionalmente através da abordagem utilizada. As anotações funcionais destes genes (termos GO, números EC e domínios conservados) estão descritas no **APÊNDICE A** em conjunto com os resultados de BLAST (descrição do melhor *hit*, similaridade e valor *e*).

Entre as sequências identificadas, estão dois genes relacionados ao metabolismo de lipídeos: glicoproteína sináptica 2 (clone H12) e a gliceraldeído-3-fosfato desidrogenase (GPD, Clone C10). A GPD é uma enzima citoplasmática que participa da glicólise produzindo um dos importantes precursores metabólicos produzidos nesta via e que é substrato para a acetilação de ácidos graxos. Na vieira *Pecten maximus*, a expressão de GPD foi associada com o metabolismo de lipídeos em células da glândula digestiva. Em pectinídeos, a glândula digestiva acumula e libera lipídeos sazonalmente e a transferência de lipídeos deste órgão para a gônada durante a gametogênese já foi observada em algumas espécies. Variações sazonais na expressão de GPD também foram observadas nestas células, diminuindo e aumentando, respectivamente, antes e após a liberação de gametas em *P. maximus* (Le Pennec e Le Pennec, 2002). A GPD tem, portanto, o potencial de ser usada como um marcador funcional em estudos do metabolismo de lipídeos na glândula digestiva e da sua relação com a reprodução destes organismos, processos de particular interesse para a maricultura.

Uma lectina do tipo C (clone D12) foi identificada pela anotação do domínio de ligação (CTLD – *C-type Lectin-like Domain*) característico destas proteínas. A sequência deste clone inicialmente não apresentou *hits* significativos através do algoritmo BLASTx. No entanto, além da identificação do domínio CTLD, um quadro aberto de leitura (ORF) de 630 pb foi identificado. A sequência de aminoácidos prevista corresponde a uma proteína com 209 aminoácidos contendo um único domínio CTLD e um peptídeo sinal N-terminal entre os aminoácidos 1 e 29. A sequência de aminoácidos prevista foi comparada contra o banco de dados de proteínas através do algoritmo BLASTp e exibiu similaridade, apesar de baixa, significativa com a lectina do tipo de C de *Tribolium castaneum* (resultados descritos no **Apêndice A**).

As lectinas do tipo C exercem papéis importantes na defesa contra a infecção por patógenos, reconhecendo-os por meio da sua habilidade de se ligar a componentes, em geral carboidratos, presentes na superfície dos microrganismos e atuando, portanto, como moléculas reconhecedoras de padrões moleculares associados a patógenos (PMAPs) (McGrealet al., 2004). Pelo menos uma lectina do tipo C identificada no pectinídeo *Chlamys farreri* (Wang et al., 2007) demonstrou ser capaz de

agregar bactérias e inibir seu crescimento. Neste mesmo organismo, foram posteriormente identificadas 4 tipos de lectinas do tipo C contendo entre 1 e 4 cópias do domínio de ligação a carboidratos. A última proteína deste tipo identificada, Cflc-4, contém 4 domínios de ligação e sua expressão é induzida por patógenos, nos hemócitos de *C. farreri* (Zhang et al., 2009a). No camarão *Panaeus monodon* e no bivalve *Ruditapes philippinarum*, foram identificadas Lectinas do tipo C que aumentam a atividade fagocítica dos hemócitos após a agregação de bactérias, funcionando, portanto, como opsoninas (Luo et al., 2006, Kim et al., 2006).

Mais recentemente, Yamaura et al., (2008) avaliaram a expressão de uma lectina do tipo C (CgCLec-1) em diferentes tecidos de *Crassostrea gigas* e mostraram que este gene é expresso não apenas pelos hemócitos como pelas células basofílicas da glândula digestiva, um tecido altamente suscetível a infecções. As lectinas do tipo C são, portanto, parte importante da linha de defesa do sistema imune inato de invertebrados que permite a estes organismos resistirem a infecções mesmo vivendo em ambientes com elevadas densidades de microrganismos.

A sequência do clone F10 foi identificada como uma proteína contendo o domínio FAS1 (*Fasciclin-like domain*) e um peptídeo sinal N-terminal localizado entre os aminoácidos 1 e 25. O domínio FAS1 está presente em uma variedade de proteínas secretadas ou ancoradas por GPI (Glicosilfosfatidilinositol) na superfície extracelular e podem estar presente em uma ou múltiplas cópias em *tandem* ou ainda intercalados com outros domínios. No clone F10 foi identificada uma cópia completa deste domínio seguida por outra cópia incompleta, sugerindo que este clone represente uma sequência parcial de uma proteína com ao menos dois domínios FAS1. De fato, este mesmo clone apresentou similaridade à proteína β -Ig-H3/fasciclin de *Octadecabacter antarcticus* 238, proteínas que caracteristicamente contêm quatro cópias do domínio FAS1 e têm sido amplamente descritas como moléculas mediadoras da adesão celular em mamíferos, plantas, bactérias, leveduras e ouriços-do-mar (Kim et al., 2000).

A sequência do clone D11, embora pequena (127 pb), apresentou similaridade de 81,5% (valor $e 5,0 \times 10^{-26}$) com a sequência do gene *HSP90* de *Crassostrea gigas*. Além disso, esta mesma sequência foi anotada com o domínio conservado característico

desta mesma família de proteínas. Algumas HSPs (proteínas de choque térmico) são proteínas constitutivas que atuam na síntese e transporte de proteínas, garantindo o correto enovelamento destas moléculas durante estes processos, enquanto outras são induzidas sob condições de estresse, auxiliando na manutenção da integridade estrutural das proteínas assim como prevenindo a agregação protéica. Embora a temperatura tenha sido o primeiro fator descrito capaz de induzir a expressão das HSPs, atualmente sabe-se que estas proteínas são induzidas por uma série de outras condições como estresse osmótico, anoxia e a exposição à xenobióticos (Gupta et al., 2010). Neste sentido, já foi demonstrado que a expressão de *HSP90* é induzida por cádmio, cobre e chumbo em *C. farreri* (Gao et al., 2007) e pela exposição à patógenos nos hemócitos de *Argopecten irradians* (Gao et al., 2008). Desta forma, esta proteína exerce papéis importantes tanto na defesa contra patógenos, como na exposição à xenobióticos em moluscos bivalves, sendo um potencial biomarcador de estresse ambiental.

Em *N. nodosus*, também foram identificados dois genes, *nadh-ubiquinone oxidoreductase fe-s protein 7* (clone C12) e *dehydrogenase reductase sdr family member 4* (clone B12) que codificam proteínas mitocondriais que participam de reações de oxirredução nesta organela. Entre as sequências que apresentaram *hits* pelo BLASTn, somente um clone foi identificado como referente a um RNA ribossomal (Clone F12, sequência parcial do RNA ribossomal 28S). Os demais clones com *hits* de BLASTx e BLASTn, descritos no **APÊNDICE B**, não puderam ser anotados funcionalmente e permanecem classificados como “proteínas de função desconhecida”. Por fim, 10 sequências de *N. nodosus* não apresentaram nenhum *hit* ou domínio conservado e podem se tratar de novos genes táxon-específicos (TRGs).

4.3.4. Identificação de genes de *Crassostrea rhizophorae*

Foram identificados e funcionalmente caracterizados 22 novos genes de *C. rhizophorae*, incluindo genes envolvidos em processos celulares, como a resposta ao estresse, síntese de proteínas, catabolismo e processamento de RNA, transcrição, catabolismo de esteróides, entre outros. As anotações funcionais destes genes (termos GO, números EC e domínios conservados) estão descritas no **APÊNDICE C** juntamente com os resultados de BLAST (descrição do melhor *hit*, similaridade e valor *e*).

Uma das sequências identificadas codifica uma proteína *Bcl-2 like* (clone C3, isolado da biblioteca de cDNA da glândula digestiva). Os membros da família de proteínas Bcl-2 atuam na regulação da permeabilidade da membrana mitocondrial e podem prevenir ou induzir a liberação de fatores apoptóticos, atuando como reguladores centrais da apoptose (morte celular programada) (Chao e Korsmeyer, 1998). A apoptose é um mecanismo de morte celular rigidamente controlado e organizado, conservado em todos os metazoários e ativado em inúmeros processos biológicos para eliminar células desnecessárias ao organismo adulto ou durante o seu desenvolvimento e ainda células danificadas, infectadas ou em senescência (Ellis et al., 1991). Em moluscos, entre outros contextos, a ocorrência de apoptose tem sido investigada em função da exposição a metais pesados, como o cádmio, um conhecido indutor de apoptose em vertebrados (Sokolova et al., 2004).

As proteínas Bcl-2 caracteristicamente apresentam de 1 a 4 domínios BH ("*Bcl-2 Homology*") e alguns membros podem ainda apresentar um domínio transmembrana C-terminal. Os domínios BH1 e BH2 estão presentes em todos os membros anti-apoptóticos, enquanto que o domínio BH3 está presente em todos os membros pró-apoptóticos, uma vez que é necessário para a dimerização com outros membros Bcl-2 e consequente indução da apoptose. (Chao et al., 1998). Na sequência obtida a partir do clone C3 de *C. rhizophorae* foram identificados pelo menos dois domínios BH (BH1 e BH2) e foram anotados termos GO que indicam localização na membrana mitocondrial, função anti-apoptótica e participação nos processos de resposta ao

estresse e a substâncias orgânicas.

Um domínio FRED (*Fibrinogen-related domain*) foi identificado na sequência do clone B5, isolado da biblioteca de cDNA dos hemócitos. Estes domínios estão presentes em uma série de proteínas como as ficolinas, proteínas que funcionam como lectinas, a tenascina e a glicoproteína microfibrila associada 4 (MFAP4), envolvidas na interação entre tecidos e na adesão celular (Zhao et al., 1995; Jones e Jones, 2000; Matsushita e Fujita, 2004). Em moluscos, entre as proteínas contendo domínios FRED já identificadas, estão uma lectina ligante de ácido siálico no gastrópode *Limax flavus* (Kurachi et al., 1998), uma família de proteínas no caramujo *Biomphalaria glabrata* (Zhang et al., 2004) e outra na lesma *Aplysia californica* (Gorbushin et al., 2010) e, mais recentemente, uma proteína denominada AiFREP no pectnídeo *Argopecten irradians* (Zhang et al., 2009b).

Apesar de muitas proteínas contendo estes domínios já terem sido descritas em invertebrados, suas funções ainda são pouco conhecidas, mas evidências indicam que estejam envolvidas principalmente com o reconhecimento de padrões moleculares associados a patógenos. Por exemplo, Zhang et al., (2009b) demonstraram que o nível de mRNA de AiFREP é fortemente aumentado após estimulação com *Listonella anguillarum* e que a proteína AiFREP recombinante apresenta atividade aglutinante dependente de cálcio contra bactérias gram positivas e negativas e contra eritrócitos de galinha e humanos.

Genes envolvidos no metabolismo de RNA, proteínas, oligossacarídeos e esteróides também foram identificados. O clone G4 codifica a proteína *MPDU1* ("*mannose-p-dolichol utilization defect 1*"), envolvida na síntese de oligossacarídeos, mas cuja função exata ainda não é clara. No entanto, em humanos, mutações neste gene foram associadas a uma desordem congênita de glicosilação, indicando sua participação nesta via (Schenk et al., 2001).

A sulfatase esteróide (clone G2) exerce papel importante na produção de esteróides biologicamente ativos, pois é responsável pela conversão de sulfato de estrona em estrona e estradiol. A deficiência desta enzima, em humanos, causa a Ictiose ligada ao

cromossomo X e a sulfatase esteróide também tem sido considerada um alvo para terapias contra o câncer de mama, uma vez que sua expressão é especificamente aumentada nestes tumores e os esteróides por ela produzidos induzem o seu crescimento (Reed et al., 2005).

Entre os genes associados ao metabolismo de proteínas, estão os de duas proteínas ribossomais, s3 (clone C4) e l6 (clone B6), partes constituintes do ribossomo e, portanto, envolvidas no processo de tradução de proteínas. A KKR1, proteína que participa da montagem do ribossomo e do seu transporte para o núcleo, também foi identificada (clone B1) (Gromadka et al., 2004). A sequência do clone A3 também foi associada à tradução, tendo sido identificada como o fator de alongamento EF-TS.

O clone H4 corresponde a uma das subunidades (tipo 1- α) do proteasomo, complexo envolvido no catabolismo de proteínas, responsável pela degradação de proteínas danificadas ou com enovelamento incorreto, evitando que gerem efeitos tóxicos. O proteasomo está também envolvido na regulação da concentração de proteínas regulatórias, fazendo com que estas sejam degradadas no tempo correto (Alberts et al., 2004). Por fim, o clone B2, foi anotado com o domínio desacilase da família YbaK que inclui o domínio de edição de aminoácidos da prolil tRNA sintetase do tipo II (ProRS) e seus homólogos YbaK, ProX, e PrdX. A principal função deste domínio é hidrolisar tRNAs que deveriam carregar um aminoácido prolina, mas que foram erroneamente carregados com uma cisteína ou alanina, auxiliando desta forma na fidelidade do processo de tradução de proteínas (Ruan e Söll, 2005).

Um grupo de 4 genes foi associado ao metabolismo de RNA. O fator de iniciação da transcrição IIB ou TFIIB (clone E2) é um dos fatores de transcrição gerais necessários para a iniciação da transcrição pela *RNA polimerase II*, enzima responsável pela transcrição de todos os genes que codificam proteínas, snoRNAs (*small nucleolar RNAs*) e alguns snRNAs (*small nuclear RNAs*) (Alberts et al., 2004). A sequência do clone H6 codifica a subunidade 2 do complexo CCR4-NOT ("*carbon catabolite repression*"), um regulador global da expressão gênica, conservado de leveduras a humanos. Evidências indicam que, em leveduras, este complexo atue como uma plataforma regulatória de inúmeros processos como a transcrição e a degradação de

RNA e que funcione como um sensor de nutrientes, estresse e, possivelmente, outros sinais extracelulares. O complexo CCR4-NOT seria, portanto, importante para a manutenção da homeostase celular em resposta a sinais extracelulares (Collart, 2003).

O clone E3 codifica um dos componentes com atividade exonuclease que constituem o exossomo, um complexo multiprotéico responsável pela degradação, no núcleo, de RNAs danificados, incompletos ou que não foram corretamente processados, impedindo que sejam traduzidos em proteínas (Alberts et al., 2004). A proteína Magoshashi (clone F8) está envolvida em um mecanismo alternativo a este para o controle de qualidade de RNA chamado NMD (*Nonsense-mediated mRNA decay*) e também conhecido como “vigilância do RNA”. Este processo elimina RNAs mensageiros que contenham códons de terminação em locais errados e desta forma previne a produção de proteínas truncadas que podem não funcionar corretamente ou ainda passar a exercer atividades deletérias para a célula (Maquat, 2004).

Duas sequências estão relacionadas com a divisão celular: a proteína ligante de Ran 1 (RanBP1, clone D3) e a proteína *Yippee-like* (clone B4). A RanBP1 é uma das proteínas regulatórias de Ran GTPase, proteína implicada em uma série de processos como o transporte entre núcleo e citoplasma e a progressão do ciclo celular. A RanBP1 regula a proteína Ran ligando-se a sua forma acoplada a GTP (Guanosina trifosfato) e favorecendo a sua conversão em Ran-GDP. Esta atividade regulatória de RanBP1 é necessária para a organização e funcionamento do fuso mitótico e evidências indicam que esta proteína também regule a coesão do centrôssomo durante a mitose, em células de mamíferos (Guarguaglini et al., 2000; Di Fiori et al., 2003).

As proteínas *Yippee-like* (YPEL) foram primeiramente identificadas em *Drosophila* e posteriormente proteínas homólogas foram encontradas também em pelo menos 68 outras espécies, incluindo, animais, plantas e fungos. As localizações subcelulares destas proteínas, no núcleo, durante a interfase, e em associação ao centrôssomo e ao fuso mitótico, durante a mitose, indicam que elas exerçam alguma função, ainda desconhecida, na divisão celular. Dada sua ampla conservação entre eucariontes, tal função possivelmente é essencial (Hosono et al., 2004).

Duas repetições do módulo tiroglobulina (Thyroglobulin-TY) do tipo I foram identificadas na sequência do clone H3, isolado da glândula digestiva. Estes módulos protéicos ricos em cisteína foram primeiramente caracterizados na proteína tiroglobulina humana, precursora dos hormônios tireoidianos e que possui 11 repetições deste módulo. A ocorrência de 1 a 3 repetições TY foi detectada também em 29 outras proteínas não relacionadas com a tiroglobulina, como proteínas que atuam na adesão celular, antígenos de superfície celular e proteínas solúveis (Molina et al., 1996). Embora estas proteínas exerçam funções diferentes, grande parte delas tem a propriedade de se ligar a outras proteínas ou a pequenos ligantes, o que pode indicar a participação das repetições TY na mediação destas interações.

Uma cópia de um motivo estrutural que sabidamente media interações proteína-proteína, o tetratricopeptídeo (TPR), foi identificada na sequência do clone A4. Os TPRs facilitam a interação entre proteínas parceiras e ocorrem em múltiplas cópias em uma grande variedade de proteínas de animais, plantas, bactérias e fungos, das quais muitas fazem parte de complexos multiprotéicos. De fato, evidências indicam que os TPRs são necessários para o correto funcionamento de complexos protéicos como o da transcrição, de chaperonas e que atuam no ciclo celular e no transporte de proteínas (Blatch e Lässle, 1999).

Entre as sequências remanescentes, estão o clone F4, anotado com uma região $\alpha\beta$ -hidrolase presente principalmente em muitas lipases; uma proteína transmembrana de função desconhecida (clone C2); uma proteína envolvida em reações de oxiredução (clone G1) e uma proteína com um domínio de ligação de FAD (Flavina adenina dinucleotídeo) (clone F3), presente em uma série de enzimas que usam esta molécula como cofator.

Sequências referentes a RNA ribossomais foram encontradas somente na biblioteca de cDNA dos hemócitos que apresentou 4 sequências referentes ao RNA ribossomal mitocondrial 16S (Clones C7, H7, D5 e H5). 15 sequências apresentaram *hits*, principalmente contra os bancos de dados de nucleotídeos, mas não puderam ser funcionalmente anotadas. As demais 16 sequências não apresentaram *hit* algum e, portanto, não possuem genes homólogos conhecidos e podem representar novos

genes táxon-específicos (TRGs), em sua grande maioria, genes expressos pelos hemócitos.

5. CONCLUSÕES

Neste trabalho, iniciamos o sequenciamento do transcriptoma de dois bivalves tropicais, a ostra *C. rhizophorae* e a vieira *N. nodosus*, espécies de particular interesse para o estudo de mecanismos de toxicidade de poluentes ambientais no Brasil. Para tanto, geramos bibliotecas de cDNA normalizadas tecido-específicas para estas duas espécies.

A combinação de técnicas utilizadas se mostrou eficaz para gerar bibliotecas de cDNA com baixa redundância, insertos longos, possivelmente contendo sequências gênicas completas, além de uma alta proporção de clones recombinantes. O sequenciamento de clones selecionados aleatoriamente a partir destas bibliotecas possibilitou a identificação de genes envolvidos em uma diversidade de processos, demonstrando a aplicabilidade destas bibliotecas para o isolamento e identificação de novos genes da ostra *C. rhizophorae* e da vieira *N. nodosus*.

Frente aos objetivos pré-estabelecidos e em função dos resultados obtidos e discutidos ao longo desta dissertação, concluímos que:

- Em relação à construção e validação de bibliotecas de cDNA não redundantes:
 - Três bibliotecas de sequências expressas na glândula digestiva de *N. nodosus* foram construídas, entre as quais uma (Gd.Nn.3) foi validada como fonte não redundante de cDNA para o isolamento e sequenciamento de novos genes desta espécie.
 - Seis bibliotecas de cDNA foram construídas para *C. rhizophorae*, sendo metade com sequências expressas nos hemócitos e as demais, com sequências expressas na glândula digestiva. Para cada um destes

tecidos, uma biblioteca de cDNA foi validada como fonte de cDNA não redundante (glândula digestiva) ou pouco redundante (hemócitos) para o isolamento de novos genes da ostra.

- Em relação à identificação e anotação funcional de genes por ferramentas de bioinformática:
 - foram identificados e anotados com termos de ontologia gênica e/ou domínios conservados 22 novos genes de *C. rhizophorae* e 8, de *N. nodosus*
 - foram isoladas, mas não puderam ser caracterizadas funcionalmente 31 novas sequências da ostra e 19, da vieira e, portanto, suas funções são desconhecidas.
 - Todas as sequências produzidas foram depositadas no genbank (NCBI).

6. PERSPECTIVAS

Sequenciamento e anotação funcional de genes.

Em um primeiro momento, daremos continuidade ao processo de sequenciamento e anotação funcional de genes de *C. rhizophorae* e *N. nodosus*, isolando e sequenciando clones em maior escala a partir das bibliotecas de cDNA geradas neste trabalho.

Desenvolvimento de microrranjos de DNA e estudo de perfis de expressão gênica.

A construção de microrranjos de DNA para estas duas espécies poderá ser viabilizada através do uso das sequências produzidas para o desenho de oligonucleotídeos, assim como através da produção de sondas por PCR, a partir da coleção de clones isolados das bibliotecas. Em princípio, estes microrranjos serão aplicados para investigar os mecanismos moleculares envolvidos na toxicidade causada pelo metal pesado cádmio em ostras *Crassostrea rhizophorae* expostas em laboratório.

7. REFERÊNCIAS BIBLIOGRÁFICAS

- Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P. 2004. *Biologia Molecular da Célula*. Editora Artes Médicas, Porto Alegre, 1584p.
- Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W. & Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol.* 215(3):403-10.
- Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; Harris, M.A.; Hill, D.P.; Issel-Tarver, L.; Kasarskis, A.; Lewis, S.; Matese, J.C.; Richardson, J.E.; Ringwald, M.; Rubin, G.M. & Sherlock G. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet.* 25(1):25-9.
- Adams, J. 2008. DNA sequencing technologies. *Nature Education* 1(1).
- Barrett, J.C. & Kawasaki, E.S. 2003. Microarrays: the use of oligonucleotides and cDNA for the analysis of gene expression. *Science*, 8(3), 134-141.
- Bishop, J.O.; Morton, J.G.; Rosbash, M. & Richardson, M. 1974. Three abundance classes in HeLa cell messenger RNA. *Nature* 250: 199-204.
- Blatch, G.L. & Lässle, M. 1999. The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. *Bioessays*. 21(11):932-9.
- Bogdanova, E.A., Shagin, D.A. & Lukyanov, S.A. 2008. Normalization of full-length enriched cDNA. *Mol Biosyst* 4(3):205-12.
- Bonaldo, M.F.; Lennon, G. & Soares, M.B. 1996. Normalization and subtraction: two approaches to facilitate gene discovery. *Genome Research* 6: 791-806.
- Brown, M.M.; Williams, T.D.; Chipman, J.K.; Katsiadaki, I.; Sanders, M. & Craft, J.A. 2008. Construction of subtracted EST and normalised cDNA libraries from liver of chemical-exposed three-spined stickleback (*Gasterosteus aculeatus*) containing pollutant-responsive genes as a resource for transcriptome analysis. *Marine Environmental Research* 66: 127-130.
- Cairns, J., Jr. and D.I. Mount. *Aquatic Toxicology. Env. Sci. Tech.*24(2):154-191.
- Carthew, R.W. & Sontheimer, E.J. 2009. Origins and Mechanisms of miRNAs and siRNAs. *Cell* 136, 642–655
- Chao, D.B & Korsmeyer, S.J. 1998. BCL-2 FAMILY: Regulators of Cell Death. *Annual Review of Immunology* 16:395-419
- Collart, M.A. 2003. Global control of gene expression in yeast by the Ccr4-Not complex. *Gene* 313: 1-16.
- Conesa, A.; Götz, S.; García-Gómez, J.M.; Terol, J.; Talón, M. & Robles, M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics Applications Note* 21(18): 3674-3676.
- Costa, M.M.; Dios, S.; Romero, A.; Novoa, B. & Figueras, A. 2009. Evidence of high individual diversity on mytilin C in mussel (*Mytilus galloprovincialis*). *Developmental and Comparative Immunology* 33:162–170.
- Cunningham, C.; Hikima, J.; Jenny, M.J.; Chapman, R.W.; Fang, G.; Sasaki, C.; Lundqvist, M.L.; Wing, R.A. Cupit, P.M.; Gross, P.S.; Warr, G.W. & Tomkins, J.P. 2006. New Resources for Marine Genomics: Bacterial Artificial Chromosome Libraries for the Eastern and Pacific Oysters (*Crassostrea virginica* and *C. gigas*). *Marine biotechnology* 8: 521-533.
- Dardel, F. & Képès, F. 2006. *Bioninformatics – genomics and post genomics*. John Wiley & Sons, 252p.

- Das, M.; Harvey, I.; Chu, L.L.; Sinha, M. & Pelletier, J. 2001. Full-length cDNAs: more than just reaching the ends. *Physiol Genomics* 6: 57–80.
- De Tomaso, A.W. 2009. Sea squirts and immune tolerance. *Dis Model Mech* (9-10):440-5.
- Di Fiori, D.; Ciciarello, M.; Mangiacasale, R.; Palena, A.; Tassin, A.; Cundari, E. & Lavia, P. 2003. Mammalian RanBP1 regulates centrosome cohesion during mitosis. *Journal of Cell Science* 116(16): 3399-3411.
- Dondero, F.; Piacentini, L.; Marsano, F.; Rebelo, M.; Vergani, L.; Venier, P. & Viarengo, A. 2006b. Gene transcription profiling in pollutant exposed mussels (*Mytilus* spp.) using a new low-density oligonucleotide microarray. *Gene* 376: 24-36.
- Ellis, R.E.; Yuan, J. & Horvitz, R. 1991. Mechanisms and functions of cell death. *Annu. Rev. Cell Bioi.* 7: 663-98.
- Ferreira, V.S.; Rocchetta, I.; Conforti, V.; Bench, S.; Feldman, R. & Levin, Mariano, J. 2007. Gene expression patterns in *Euglena gracilis*: Insights into the cellular response to environmental stress. *Gene* 389: 136-145.
- Ferreira, M.A.P.; Paixão, L.F.; Alcântara-Neto, C.P.; Santos, S. S.D. & Rocha, R.M. 2006. Morphological and Morphometric Aspects of *Crassostrea rhizophorae* (Guilding, 1828) Oocytes in Three Stages of the Gonadal Cycle. *Int J Morphol* 24(3):437-442.
- Fleury E.; Moal, J.; Boulo, V.; Daniel, J.; Mazurais, D.; Hénaut, A.; Corporeau, C.; Boudry, P.; Favrel, P.; & Huvet, A. 2009a. Microarray-Based Identification of Gonad Transcripts Differentially Expressed Between Lines of Pacific Oyster Selected to Be Resistant or Susceptible to Summer Mortality. *Marine biotechnology* Oct.
- Fleury, E., Huvet, A., Lelong, C., Lorgeril, J. D., Boulo, V., Gueguen, Y. Bachère, E.; Tanguy, A.; Moraga, D.; Fabioux, C.; Lindeque, P.; Shaw, J.; Reinhardt, R.; Prunet, P.; Davey, G.; Lapègue, S.; Sauvage, C.; Corporeau, C.; Moal, J.; Gavory, F.; Wincker, P.; Moreews, F.; Klopp, C.; Mathieu, M.; Boudry, P. & Favrel, F. 2009b. Generation and analysis of a 29,745 unique Expressed Sequence Tags from the Pacific oyster (*Crassostrea gigas*) assembled into a publicly accessible database: the GigasDatabase. *BMC Genomics* 15: 1-15.
- Frith, M.C.; Pheasant, M. & Mattick, J.S. 2005. The amazing complexity of the human transcriptome. *Eur J Hum Genet.* 13(8):894-7.
- Galau, G.A.; Klein, W.H.; Britten, R.J. & Davidson, E.H. 1977. Significance of rare mRNA sequences in liver. *Arch Biochem Biophys* 179:584–599.
- Galvão PMA. 2007. Dinâmica de Zn²⁺ e Cd²⁺ em *Nodipecten nodosus*. 2005. Dissertação de Mestrado Ecologia. Universidade Federal do Rio de Janeiro.
- Gao, Q., Song, L., Ni, D., Wu, L., Zhang, H., Chang, Y., et al. 2007. cDNA cloning and mRNA expression of heat shock protein 90 gene in the haemocytes of Zhikong scallop *Chlamys farreri*. *Comparative Biochemistry and Physiology, Part B* 147: 704–715.
- Gao, Q.; Zhao, J.; Song, L.; Qiu, L.; Yu, Y.; Zhang, H. & ni, D. 2008. Molecular cloning, characterization and expression of heat shock protein 90 gene in the haemocytes of bay scallop *Argopecten irradians*. *Fish Shellfish Immunol.* 24(4): 379-85.
- Gestal, C.; Costa, M.; Figueiras, A. & Novoa, B. 2007. Analysis of differentially expressed genes in response to bacterial stimulation in hemocytes of the carpet-shell clam *Ruditapes decussatus*: identification of new antimicrobial peptides. *Gene* 406:134–143.
- Götz, S.; García-Gómez, J.M.; Terol, J.; Williams, T.D.; Nagaraj, S.H.; Nueda, M.J.; Robles, M.; Talón, M.; Dopazo, J. & Conesa, A. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research* (2), 1-16.
- Goffeau, A.; Barrell, B. G.; Bussey, H.; Davis, R. W.; Dujon, B.; Feldmann, H.; Galibert, F.; Hoheisel, J. D.; Jacq, C.; Johnston, M.; Louis, E.J.; Mewes, H.W.; Murakami, Y.; Philippsen, P.; Tettelin, H. & Oliver,

- S.G. 1995. Life with 6000 Genes. *Science* 274: 546-567.
- Goldstone, J.V. 2008. Environmental sensing and response genes in Cnidaria: the chemical defensome in the sea anemone *Nematostella vectensis*. *Cell Biol Toxicol* 24(6): 483–502.
- Gorbushin, A.M.; Panchin, Y.V. & Iakovleva, N.V. 2010. In search of the origin of FREPs: characterization of *Aplysia californica* fibrinogen-related proteins. *Dev Comp Immunol.* 34(4):465-73.
- Griffiths, A.J.F.; Miller, J.H.; Suzuki, D.T.; Lewontin, R.C.; Gelbart, W.M.. 1999. An introduction to genetic analysis. W. H. Freeman & Co., New York, 860p.
- Gromadka, R.; Karkusiewicz, I.; Rempola, B. & Rytka, J. 2004. Functional and physical interactions of Krr1p, a *Saccharomyces cerevisiae* nucleolar protein. *Acta Biochimica Polonica* 51(1): 173-187.
- Guarguaglini, G.; Renzi, L.; D'ottavio, F.; Fiore, D. Casenghi, M; Cundari, E. & Lavia, P. 2000. Regulated Ran-binding protein 1 activity is required for organization and function of the mitotic spindle in mammalian cells in vivo. *Cell Growth Differ.* 11: 455-465.
- Gupta, S.C.; Sharma, A.; Mishra, M.; Mishra, R.K. & Chowdhuri, D.K. 2010. Heat shock proteins in toxicology: How close and how far? *Life Sciences* (In press).
- Hernandez, L.M. & Blazer, D.G. 2006. Genes, Behavior, and the Social Environment: Moving Beyond the Nature/Nurture Debate. The National academies press, Washington, 384 p.
- Hosono, K.; Sasaki, T.; Minoshima, S. & Shimizu, N. 2004. Identification and characterization of a novel gene family YPEL in a wide spectrum of eukaryotic species. *Gene* 340: 31 - 43.
- Ignácio, B.L.; Absher, T.M.; Lazoski, C. & Solé-Cava, AM. 2000. Genetic evidence of the presence of two species of *Crassostrea* (*Bivalvia*: *Ostreidae*) on the coast of Brazil. *Marine Biology* 135: 987-991.
- Jenny, M.J.; Chapman, R.W.; Mancina, A.; et ally. 2007. A cDNA Microarray for *Crassostrea virginica* and *C. gigas* *Marine biotechnology* 9: 577-591.
- Jones, F.S. & Jones, P.L. 2000. The Tenascin Family of ECM Glycoproteins: Structure, Function, and Regulation During Embryonic Development and Tissue Remodeling. *Developmental dynamics* 218:235–259.
- Kandpal, R. P.; Saviola, B., & Felton, J. 2009. The era of 'omics unlimited. *BioTechniques* 46:351-355.
- Kang, Y.; Kim, Y.; Park, K.; Cho, S.K.; Cho, K. & Cho, M. 2006. Analysis of EST and lectin expressions in hemocytes of Manila clams (*Ruditapes philippinarum*) (*Bivalvia*: *Mollusca*) infected with *Perkinsus olseni*. *Developmental and Comparative Immunology* 30: 1119-1131.
- Khalturin, K.; Hemmrich, G.; Fraune, S.; Augustin, R. & Bosch, T.C.G. 2009. More than just orphans: are taxonomically-restricted genes important in evolution? *Trends in Genetics* 25 (9): 404-413.
- Kim, J.; Kim, S.; Lee, B.; Park, R.; Kim, K. & Kim, I. 2000. Identification of Motifs for Cell Adhesion within the Repeated Domains of Transforming Growth Factor- β -induced Gene, big-h3*. *The journal of biological chemistry* 275 (40) 6: 30907–30915.
- Kim, Y.M.; Park, K.I., Alvarez, R.A., Cummings, R.D. & Cho, M. 2006. Lectin from the Manila clam *Ruditapes philippinarum* is induced upon infection with the protozoan parasite *Perkinsus olseni*. *J Biol Chem.* 15;281(37):26854-64.
- Koonin, E.V. & Galperin, M.Y. 2002. Sequence - Evolution – Function. *Computational Approaches in Comparative Genomics*. Hardcover, 488 p.
- Koonin, E.V. 2005. Orthologs, Paralogs, and Evolutionary Genomics. *Annual Review of Genetics* 39: 309-338.
- Kurachi, S.; Song, Z.; Takagaki, M.; Yang, Q.; Winter, H.C.; Kurachi, K. & Goldstein IJ. 1998. Sialic-acid-binding lectin from the slug *Limax flavus*--cloning, expression of the polypeptide, and tissue localization. *Eur J Biochem.* 254(2):217-22.
- Kuznetsov, V.A.; Knott, G.D.; & Bonner, R.F. 2002. General Statistics of Stochastic Process of Gene

- Expression in Eukaryotic Cells. *Gene Expression*, 1332: 1321-1332.
- Lacerda L.D. & Molisani M.M. 2006. Three decades of Cd and Zn contamination in Sepetiba Bay, SE Brazil: Evidence from the mangrove oyster *Crassostreaa rhizophorae*. *Mar Pollut Bull* 52(8):974-7.
- Lander, E.S.; Linton, L.M. et al. 2001. Initial sequencing and analysis of the human genome. *Nature* 15;409(6822):860-921.
- Le Pennec, G. & Le Pennec, M. 2002. Molecular analysis of the seasonal expression of genes coding for different functional markers of the digestive gland of the bivalve mollusk *Pecten maximus* (L.) *Comparative Biochemistry and Physiology Part B* 133: 417–426.
- Li, R.; Li, Y.; Zheng, H.; Luo, R. et al. 2010. Building the sequence map of the human pan-genome. *Genome Research*, 28(1), 57-63.
- Liu, E.T. 2005. Systems Biology, Integrative Biology, Predictive Biology. *Cell* 121: 505-506.
- Li, T., & Brouwer, M. 2009. Bioinformatic analysis of expressed sequence tags from grass shrimp *Palaemonetes pugio* exposed to environmental stressors. *Comparative Biochemistry and Physiology, Part D* 4: 187-195..
- Luo, T.; Yang, H.; Zhang, X. & Xu, X. 2006. Purification, characterization and cDNA cloning of a novel lipopolysaccharide-binding lectin from the shrimp *Penaeus monodon*. *Dev Comp Immunol*. 30(7):607-17.
- Macek, B.; Mann, M. & Olsen, J.V. 2009. Global and Site-Specific Quantitative Phosphoproteomics: Principles and Applications. *Annual Review of Pharmacology and Toxicology* 49: 199-221
- Maquat, L.E. 2004. Nonsense-mediated mRNA Decay: splicing, translation and mRNP dynamics. *Nature Reviews Molecular Biology* 5: 89-99.
- Masek, T.; Vopalensky, V.; Suchomelova, P. & Pospisek, M. 2005. Denaturing RNA electrophoresis in TAE agarose gels. *Anal Biochem*. 336(1):46-50.
- Matsushita, M. & Fujita, T. 2002. The Role of Ficolins in Innate Immunity. *Immunobiol*. 205: 490–497.
- Matz, M.V. 2002. Amplification of representative cDNA samples from microscopic amounts of invertebrate tissue to search for new genes. *Methods Mol Biol* 183:3-18.
- McGreal, E.P.; Martinez-pomares, L. & Gordon, S. 2004. Divergent roles for C-type lectins expressed by cells of the innate immune system. *Molecular Immunology* 41: 1109-1121.
- Metzker, M.L. 2010. Sequencing technologies — the next generation. *Nature Reviews* 11: 31-46.
- Moggs, J.G & Orphanides, G. 2003. Genomic analysis of stress response genes. *Toxicology Letters*: 140--141:
- Molina, F.; Bouanani, M.; Pau, B. & Granier, C. 1996. Characterization of the type-1 repeat from thyroglobulin, a cysteine-rich module found in proteins from different families. *Eur. J. Biochem*. 240, 125-133.
- Nuwaisir, E.F.; Bittner, M.; Trent, J.; Barrett, J.C. & Afshari, C.A. 1999. Microarrays and Toxicology: The Advent of Toxicogenomics. *Molecular carcinogenesis* 24:153–159.
- NCBI. Statistics for NCBI Resources. Disponível em: <<http://www.ncbi.nlm.nih.gov/Sitemap/Summary/statistics.html>>. Acesso em 01/2010.
- Oresic, M. 2009. Metabolomics, a novel tool for studies of nutrition, metabolism and lipid dysfunction. *Nutrition, Metabolism & Cardiovascular Diseases* 19: 816-824.
- Pallavicini, A.; Costa, M.M.; Gestal, C.; Dreos, R.; Figueiras, A.; Venier, P. & Novoa, B. 2008. High sequence variability of myctin transcripts in hemocytes of immune-stimulated mussels suggests ancient host-pathogen interactions. *Dev Comp Immunol* 32:213–226.
- Pauls, E. & Affonso, P.R.A.M. 2000. The karyotype of *Nodipecten nodosus* (Bivalvia:

- Pectinidae). *Hydrobiologia* 420: 99-102.
- Ponting, C.P.; Oliver, P.L. & Reik, W. Evolution and Functions of Long Noncoding RNAs. *Cell* 136, 629–641.
- Powers, J.P.S. & Hancock, R.E.W. 2003. The relationship between peptide structure and antibacterial activity. *Peptides* 24: 1681–1691.
- Quilang, J.; Wang, S.; Li, P.; Abernathy, J.; Peatman, E.; Wang, Y.; Wang, L.; Shi, Y.; Wallace, R.; Guo, X. & Liu, Z. 2007. Generation and analysis of ESTs from the eastern oyster, *Crassostrea virginica* Gmelin and identification of microsatellite and SNP markers. *BMC Genomics* 8:157.
- Rebello, M.F.; Pfeiffer, W.C.; Silva, H. & Moraes, M.O. 2003a. Cloning and detection of metallothionein mRNA by RT-PCR in mangrove oysters (*Crassostrea rhizophorae*). *Aquatic Toxicology* 64: 359-362.
- Rebello, M.F.; Amaral, M.C.R. & Pfeiffer, W.C. 2003b. High Zn and Cd accumulation in the oyster *Crassostrea rhizophorae*, and its relevance as a sentinel species. *Marine Pollution Bulletin* 46(10): 1354-1358.
- Reed, M.J.; Purohit, A.; Woo, L.W.; Newman, S.P. & Potter, B.V. 2005. Steroid Sulfatase: Molecular Biology, Regulation, and Inhibition. *Endocrine Reviews* 26: 171–202.
- Rittschof, D. & McClellan-Green P. 2005. Molluscs as multidisciplinary models in environment toxicology. *Mar Pollut Bull* 50(4):369-73.
- Robbens, J.; Ven, K.V.; Maras, M.; Blust, R.; & Coen, W.D. 2007. Ecotoxicological risk assessment using DNA chips and cellular reporters. *Trends in biotechnology* 25(10): 460-466.
- Roberts, S.; Goetz, G.; White, S. & Goetz, F. 2009. Analysis of Genes Isolated from Plated Hemocytes of the Pacific Oyster, *Crassostrea gigas*. *Mar Biotechnol* 11:24–44.
- Rost B. 1999. Twilight zone of protein sequence alignments. *Protein Eng* 12(2):85-94.
- Ruan, B. & Söll, D. 2005. The Bacterial YbaK Protein Is a Cys-tRNA^{Pro} and Cys-tRNA^{Cys} Deacylase. *The Journal of Biological chemistry* 280(7): 25887–25891.
- Rupp, G.S. & Parsons, G.J. 2006. Scallop aquaculture and fisheries in Brazil. In: Shumway, S.E. & Parsons, G.J. *Scallops: biology, ecology and aquaculture*. Elsevier, Oxford: 1225-12-48.
- Sabry, R.C.; Magalhães, A.R.M. 2005. Parasitas em Ostras de Cultivo (*Crassostrea rhizophorae* e *Crassostrea gigas*) da Ponta do Sambaqui, Florianópolis, SC. *Arquivo Brasileiro de Medicina Veterinária e Zootecnia* 57: 194 -202.
- Schena, M.; Shalon, D.; Davis, R.W. & Brown, P. 1995. Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA microarray. *Science* 270: 467-470.
- Schena, M.; Heller, R.A.; Thieriault, T.P.; Konrad, K.; Lachenmeier, E. & Davis, R.W. 1998. Microarrays: biotechnology's discovery platform for functional genomics. *Trends Biotechnol.* 16(7): 301-306.
- Schenk, B.; Imbach, T.; Frank, C.G.; Grubenmann, C.E.; Raymond, G.V.; Hurvitz, H.; Raas-Rotschild, A.; Luder, A.A.; Kaeken, J.; Berger, E.G.; Matthijs, G.; Hennet, T. & Aebi, M. 2001. MPDU1 mutations underlie a novel human congenital disorder of glycosylation, designated type J. *Clin. Invest.* 108:1687–1695.
- Shagin, D.A.; Lukyanov, K.A.; Vagner, L.L. & Matz, M.V. 1999. Regulation of average length of complex PCR product. *Nucleic Acids Res.* 27(18):e23.
- Silva, C.A.; Smith, B.D. & Rainbow P.S. 2008. Comparative biomonitoring of coastal trace metal contamination in tropical South America (N. Brazil). *Mar Environ Res* 61(4):439-55.
- Soares, M.B.; Bonaldo, M.F.; Jelene, P.; Su, L.; Lawton, L. & Efstratiadis, A. 1994. Construction and characterization of a normalized cDNA library. *Proc Natl Acad Sci U S A.* 27;91(20):9228-32.
- Sokolova I.M.; Evans, S. & Hughes, F.M. 2004. Cadmium-induced apoptosis in oyster hemocytes involves disturbance of cellular energy balance but no mitochondrial permeability transition. *J. Exp. Biol.*

207: 3369-3380.

- Tang, Z; Choi, J.; Hemmerich, C.; Sarangi, A.; Colbourne, J.K. & Dong, Q. 2009. ESTPiper – a web-based analysis pipeline for expressed sequence tags. *BMC Genomics* 10:174.
- Tanguy, A.; Bierne, N.; Saavedra, C.; Pina, B.; Bachère, E.; Kube, M.; Bazin, E.; Bonhomme, F.; Boudry, P.; Boulo, V.; Boutet, I.; Cancela, L.; Dossat, C.; Favrel, P.; Huvet, A.; Jarque, S.; Jollivet, D.; Klages, S.; Lapègue, S.; Leite, R.; Moal, J.; Moraga, D.; Reinhart, R.; Samain, J.; Zouros, E. & Canario, A. 2008. Increasing genomic information in bivalves through new EST collections in four species: Development of new genetic markers for environmental studies and genome evolution. *Gene* 408: 27–36.
- Varela, E.S.; Beasley, C.R.; Schneider, H.; Sampaio, I.; Marques-Silva, N.S. & Tagliaro, C.H. 2007. Molecular Phylogeny of Mangrove Oysters (*Crassostrea*) from Brazil. *Journal Molluscan Studies*: 229-234.
- Venter, J.C.; Adams, M.D.; Myers, E.W. et al. 2001. The sequence of the human genome. *Science* 291(5507):1304-51.
- Velculescu, V.E.; Madden, S.L.; Zhang, L.; Lash, A.E.; Yu, J.; Rago, C.; Lal, A.; Wang, C.J.; Beaudry, G.A.; Ciriello, K.M.; Cook, B.P.; Dufault, M.R.; Ferguson, A.T.; Gao, Y.; He, T.C.; Hermeking, H.; Hiraldo, S.K.; Hwang, P.M.; Lopez, M.A.; Luderer, H.F.; Mathews, B.; Petroziello, J.M.; Polyak, K.; Zawel, L. & Kinzler KW. 1999. Analysis of human transcriptomes. *Nature Genetics* 23: 387–388.
- Venier, P.; Pallavicini, A.; Nardi, B.D. & Lanfranchi, G. 2003. Towards a catalogue of genes transcribed in multiple tissues of *Mytilus galloprovincialis*. *Gene* 314, 29 - 40.
- Venier, P.; De Pittà, C.; Pallavicini, A.; Marsano, F.; Varotto, L.; Romualdi, C.; Dondero, F.; Viarengo, A. & Lanfranchi, G. 2006. Development of mussel mRNA profiling: can gene expression trends reveal coastal water pollution? *Mutat. Res.* 602(1-2): 121-134.
- Venier, P.; De Pittà C.; Bernante F.; Varotto, L.; De Nardi, B.; Bovo, G.; Roch, P.; Novoa, B.; Figueras, A.; Pallavicini, A. & Lanfranchi, G.. 2009. MytiBase: a knowledgebase of mussel (*M. galloprovincialis*) transcribed sequences. *BMC genomics* 10:72.
- Wang, H.; Song, L.; Li, C.; Zhao, J.; Zhang, H.; Ni, D. & Xu, W. 2007. Cloning and characterization of a novel C-type lectin from Zhikong scallop *Chlamys farreri*. *Mol Immunol* 44(5):722–31.
- Wang, L., Song, L., Zhao, J., Qiu, L., Zhang, H., Xu, W.; Li, H.; Li, C.; Wua, L. & Guo, X. 2009. Expressed sequence tags from the zhikong scallop (*Chlamys farreri*): Discovery and annotation of host-defense genes. *Fish & Shellfish Immunology* 26: 744-750.
- Wanick R.C. 2007. Identificação de proteínas metalotioneínas em ostras *Crassostrea rhizophorae* para aplicações ambientais. Dissertação de Mestrado em Geociências/UFF.
- Watson, J.D. & Crick F.H.C. 1953. Molecular structure of nucleic acids. *Nature* 171: 737-738.
- Wilson, G.A.; Bertrand, N.; Patel, Y.; Hughes, J.B.; Feil, E.J. & Field, D. 2005. Orphans as taxonomically restricted and ecologically important genes. *Microbiology Comment* 151: 2499–2501.
- Xu, W. & Faisal, M. 2009. Development of a cDNA microarray of zebra mussel (*Dreissena polymorpha*) foot and its use in understanding the early stage of underwater adhesion. *Gene* 1, 436(1-2):71-80.
- Yamaura, K., Takahashi, K.G. & Suzuki, T. 2008. Identification and tissue expression analysis of C-type lectin and galectin in the Pacific oyster, *Crassostrea gigas*. *Comparative Biochemistry and Physiology, Part B* 149: 168–175.
- Young, B.D. & Anderson, M.L.M. 1985. Quantitative analysis of solution hybridisation. In: Hames, B.D. & Higgins, S.J. (eds) *Nucleic Acid Hybridisation*. IRL Press, Oxford-Washington, pp 47–71
- Yount, N.Y. & Yeaman, M.R. 2004. Multidimensional signatures in antimicrobial peptides. *PNAS* 101:

7363–7368.

- Zanette, J., Nunes, F.N., Medeiros, I.D., Siebert, M.N., Mattos, J.J., Lüchmann, K.H., Melo, C.M.R. & Bainy, A.C.D. 2008. Comparison of the antioxidant defense system in *Crassostrea rhizophorae* and *Crassostrea gigas* exposed to domestic sewage discharges. *Marine Environmental Research* 66: 196–198.
- Zhang, S.; Adema, C.M.; Kepler, T.B. & Loker, E.S. 2004. Diversification of Ig Superfamily Genes in an Invertebrate. *Science* 305: 251-305.
- Zhang, H.; Wang, H.; Wang, L.; Song, L.; Song, X.; Zhao, J.; Li, L. & Qiu, L. 2009a. Cflec-4, a multidomain C-type lectin involved in immune defense of Zhikong scallop *Chlamys farreri*. *Developmental and Comparative Immunology* 33: 780-788.
- Zhang, H.; Wang, L.; Song, L.; Song, X.; Wang, B.; Mu, C. & Zhang, Y. 2009b. A fibrinogen-related protein from bay scallop *Argopecten irradians* involved in innate immunity as pattern recognition receptor. *Fish & Shellfish Immunology* 26: 56–64.
- Zhao, Z.; Lee, C.C.; Jiralerspong, S.; Juyal, R.C.; Lu, F.; Baldini, A.; Greenberg, F.; Caskey, C.T.; & Patel P.I. 1995. The gene for a human microfibril-associated glycoprotein is commonly deleted in Smith-Magenis syndrome patients. *Hum Mol Genet.* 4(4):589-97.
- Zhu, Y.Y.; Machleder, E.M.; Chenchik, A.; Li, R. & Siebert, P.D. 2001. Reverse transcriptase template switching: a SMART approach for fulllength cDNA library construction. *Biotechniques* 30:892-89.
- Zhulidov, P.A.; Bogdanova, E.A., Shceglov, A.S.; Vagner, L.L.; Khaspekov, G.L., Kozhemyako, V.B.; Matz, M.V., Meleshkevitch, E.; Moroz, L.L., Lukyanov, S. A. & Shagin, D.A. 2004. Simple cDNA normalization using kamchatka crab duplex-specific nuclease. *Nucleic Acids Research* 32 (3): e37.

APÊNDICE A - Anotações funcionais de genes de *N. nodosus*

Clone/Nº de Acesso (Genbank)	Descrição	Tamanho (pb)	Termos GO/Número EC	InterproScan/CDD	Melhor <i>hit</i> de BLAST		
					Valor <i>e</i>	Similaridade	Espécie (Código de acesso)
C10/ GW314973	glyceraldehyde-3-phosphate dehydrogenase	787	C:cytoplasm; F:NAD or NADH binding; P:oxidation reduction; P:glycolysis; F:glyceraldehyde-3-phosphate dehydrogenase (phosphorylating) activity EC:1.2.1.12	IPR000173 :Glyceraldehyde 3-phosphate dehydrogenase subfamily ; IPR016040 : NAD(P)-binding domain; IPR020828 : Glyceraldehyde 3-phosphate dehydrogenase, NAD(P) binding domain; IPR020829 : Glyceraldehyde 3-phosphate dehydrogenase, catalytic domain; IPR020832 : Glyceraldehyde 3-phosphate dehydrogenase, catalytic domain, subgroup	1E-93	88.3%	<i>Haliotis discus discus</i> (ABO26632.1)
H12/ GW314974	synaptic glycoprotein sc2	784	C:integral to membrane; C:cytoplasm; F:oxidoreductase activity, acting on the CH-CH group of donors; P:protein modification process; P:lipid metabolic process	IPR002198 : Short-chain dehydrogenase/reductase SDR; IPR002347 : Glucose/ribitol dehydrogenase; IPR016040 : NAD(P)-binding domain	1E-78	78.15%	<i>Branchiostoma floridae</i> (XP_002602630)

Clone/Nº de Acesso (Genbank)	Descrição	Tamanho (pb)	Termos GO/Número EC	InterproScan/CDD	Melhor <i>hit</i> de BLAST		
					Valor <i>e</i>	Similaridade	Espécie (Código de acesso)
C12/ GW314975	nadh-ubiquinone oxidoreductase fe-s protein 7	733	P:mitochondrial electron transport, NADH to ubiquinone; F:NADH dehydrogenase (ubiquinone) activity; F:iron ion binding; C:mitochondrial respiratory chain complex I; F:4 iron, 4 sulfur cluster binding EC:1.6.5.3	IPR001104 : 3-oxo-5-alpha-steroid 4-dehydrogenase, C-terminal, cl00419 , Oxidored_q6, NADH ubiquinone oxidoreductase	1E-70	88.85%	<i>Apis mellifera</i> (XP_392437.2)
B12/ GW314984	dehydrogenase reductase sdr family member 4	788	F:binding; F:oxidoreductase activity; P:oxidation reduction; C:mitochondrion	—	1E-60	74.1%	<i>Xenopus (Silurana) tropicalis</i> (NP_001015784.1)
A11/ GW314977	Heat Shock Protein 90	127	—	PTZ00272 , heat shock protein 83 kDa (Hsp83); Provisional; IPR001404 : Chaperone protein htpG IPR001404 Chaperone protein htpG, PTHR11528:SF20 : Heat Shock Protein 90 (Family)	5,08E-26	81,50%	<i>Crassostrea gigas</i> (EF687776.1)

Clone/Nº de Acesso (Genbank)	Descrição	Tamanho (pb)	Termos GO/Número EC	InterproScan/CDD	Melhor <i>hit</i> de BLAST		
					Valor <i>e</i>	Similaridade	Espécie (Código de acesso)
F10/ GW314978	Fasciclin	771	—	pfam02469 : Fasciclin; IPR000782 : FAS1 domain	2,17E-16	55.1%	<i>Octadecabacter antarcticus</i> 238 (ZP_05065063.1)
H11/ GW392489.2	heterogeneous nuclear ribonucleoprotein r	651	C:heterogeneous nuclear ribonucleoprotein complex; F:RNA binding; F:protein binding; P:nuclear mRNA splicing, via spliceosome; F:nucleotide binding	—	1E-14	72.15%	<i>Pinctada fucata</i> (ABP04054)
D12/ GW314979	C-type lectin	696	—	IPR001304 : C-type lectin; IPR016186 : C-type lectin-like; IPR016187 : C-type lectin fold; cd00037 : C-type lectin (CTL)/C-type lectin-like (CTLD)	2,00E-07	25%	<i>Tribolium castaneum</i> (XP_970274.2)

APÊNDICE B - Genes de *N. nodosus* com *hits* de BLAST e sem anotação funcional.

Clone/Nº de acesso no genbank	Pb	Melhor hit				Valor e	Similaridade
		Descrição	Código de acesso	Algoritmo BLAST e Banco de dados			
F12/ GU574707	637	<i>P. magellanicus</i> internal transcribed spacer and 28s ribosomal rna partial sequence	AF342798.1	BLASTn - nr	0	100,00%	
G11/GW314980	690	Pm_Admus_12G11_T3 Hele and Hannah <i>P. maximus</i> cDNA clone Pm_Admus_12G11 5-, mRNA seq	DN793786.1	BLASTn - dbEST others	0	87%	
D11/GW314981	681	hypothetical protein BRAFLDRAFT_123718 (<i>Branchiostoma floridae</i>)	XP_002606230.1	BLASTx - nr	4,92E-25	56.85%	
G9/GW314982	743	GL13736 [<i>Drosophila persimilis</i>]	XP_002019990.1	BLASTx - nr	3,45E-16	59.4%	
E12/GW314983	518	hypothetical protein BRAFLDRAFT_68822 (<i>Branchiostoma floridae</i>)	XP_002588182.1	BLASTx - nr	1,36E-12	44.7%	
G12/GW314984	228	FP000964 Pasc Fav cDNA library hemocyte (wy0aba) <i>C. gigas</i> cDNA clone wy0aba17g20, mRNA seq	FP000964.1	BLASTn - dbEST others	2,76E-07	74,8%	
E11/GW314985	787	hypothetical protein (<i>Strongylocentrotus purpuratus</i>)	XP_001203942.1	BLASTx - nr	7,29E-07	57.95%	
H10/GW314986	523	EST665450 RAB <i>R. appendiculatus</i> cDNA clone RABBE79 5- end, mRNA sequence	CD794089.1	BLASTn - dbEST others	6,49E-06	60%	
B11/GW314987	773	pm_admus_17e09_t3 helene and hannah <i>pecten maximus</i> cdna clone pm_admus_17e09 5 mrna	DN793786.1	BLASTn - dbEST others	5,42E-05	83,00%	
E10/GW314988	784	hypothetical protein BRAFLDRAFT_113518 (<i>B. floridae</i>)	XP_002611504.1	BLASTx - nr	3,97E-05	66.5%	

APÊNDICE C - Anotações funcionais de genes de *C. rhizophorae*

Clone/Nº de acesso no genbank (Tecido)	Descrição	Tamanho (pb)	Termos GO/ Número EC	InterproScan/CDD	Melhor <i>hit</i> de BLAST		
					Valor e	Similaridade	Espécie (Código de acesso)
C4/ GW314920 (G. digestiva)	ribosomal protein s3	605	C:ruffle membrane; F:protein kinase binding; P:activation of caspase activity; P:translational elongation; P:negative regulation of DNA repair; P:negative regulation of NF-kappaB transcription factor activity; C:dendrite; P:induction of apoptosis; F:NF-kappaB binding; F:endonuclease activity; F:structural constituent of ribosome; C:endomembrane system; C:nucleus; F:DNA-(apurinic or apyrimidinic site) lyase activity; F:mRNA binding; C:cytosolic small ribosomal subunit EC:4.2.99.18	—	1,93E-81	96.25%	<i>Lineus viridis</i> (ABZ04254.1)
B3/ GW314921 (G. digestiva)	ribosomal protein l6	750	C:ribosome; F:structural constituent of ribosome; P:translation EC:3.6.5.3	IPR000915: Ribosomal protein L6E (Family)	1,56E-72	77.95%	<i>Lineus viridis</i> (ABZ04219.1)

Clone/Nº de acesso no genbank (Tecido)	Descrição	Tamanho (pb)	Termos GO/ Número EC	InterproScan/CDD	Melhor <i>hit</i> de BLAST		
					Valor e	Similaridade	Espécie (Código de acesso)
F8/ GW314950 (Hemócitos)	mag0-nashi proliferation-associated	603	P:nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	IPR004023 : Mago nashi protein	1E-69	95.75%	<i>Rattus norvegicus</i> (XP_001074029.1)
E3/GW314922 (G. digestiva)	exosome component 8	728	F:identical protein binding; F:exonuclease activity; P:RNA processing; C:intracellular part	—	3,70E-68	76.0%	<i>Homo sapiens</i> (2NN6_C)

Clone/Nº de acesso no genbank (Tecido)	Descrição	Tamanho (pb)	Termos GO/ Número EC	InterproScan/CDD	Melhor <i>hit</i> de BLAST		
					Valor e	Similaridade	Espécie (Código de acesso)
H4/ GW314923 (G. digestiva)	proteasome subunit beta type-1-a	740	P:anaphase-promoting complex-dependent proteasomal ubiquitin-dependent protein catabolic process; P:positive regulation of ubiquitin-protein ligase activity during mitotic cell cycle; P:negative regulation of ubiquitin-protein ligase activity during mitotic cell cycle; C:proteasome core complex; F:protein binding; C:nucleus; C:cytosol; F:threonine-type endopeptidase activity EC:3.4.25.0	IPR001353: Proteasome, subunit alpha/beta (family); IPR016050: Proteasome, beta-type subunit, conserved site; G3DSA:3.60.20.10: Glutamine Phosphoribosylpyrophosphate, subunit 1, domain 1, PTHR11599: Proteasome subunit alpha/beta (family), PTHR11599: Proteasome subunit beta type 1 (Family); SSF56235: N-terminal nucleophile aminohydrolases (Ntn hydrolases) superfamily	8,85E-65	80.8%	<i>Rattus norvegicus</i> (AAH58455.1)

Clone/Nº de acesso no genbank (Tecido)	Descrição	Tamanho (pb)	Termos GO/ Número EC	InterproScan/CDD	Melhor <i>hit</i> de BLAST		
					Valor e	Similaridade	Espécie (Código de acesso)
E2/ GW314924 (G. digestiva)	transcription initiation factor IIFB	710	F:helicase activity; F:RNA polymerase II transcription factor activity; F:ATP binding; C:transcription factor TFIIF complex; P:transcription initiation from RNA polymerase II promoter	IPR003196: Transcription initiation factor IIF, beta subunit (Family); IPR011039: Transcription Factor IIF, Rap30/Rap74, interaction (Domain); IPHR01199: Winged helix repressor DNA-binding (Domain); SSF46785: "Winged helix" DNA-binding domain (superfamily)	2,83E-49	66.15%	<i>Apis mellifera</i> (XP_623868.1)
C2/ GW314925 (G. digestiva)	transmembrane protein 17	740	C:membrane	IPR019184: Protein of unknown function, transmembrane-17	2,59E-48	71.6%	<i>Branchiostoma floridae</i> (XP_002605870.1)
G1/ GW314926 (G. digestiva)	loc407663 protein	605	F:oxidoreductase activity; P:oxidation reduction; F:binding	_____	5,49E-45	77.55%	<i>Branchiostoma floridae</i> (XP_002613381.1)
D3/ GW314927 (G. digestiva)	ran binding protein 1	746	P:positive regulation of mitotic centrosome separation; P:intracellular transport; P:spindle organization; C:centrosome	_____	3,56E-45	69.95%	<i>Xenopus (Silurana) tropicalis</i> (NP_989020.1)

Clone/Nº de acesso no genbank (Tecido)	Descrição	Tamanho (pb)	Termos GO/ Número EC	InterproScan/CDD	Melhor <i>hit</i> de BLAST		
					Valor e	Similaridade	Espécie (Código de acesso)
A4/ GW314928 (Gl. Digestiva)	tetratricopeptid e repeat domain 21b	735	F:binding	IPR011990: Tetratricopeptide-like helical (Domain); IPR013026: Tetratricopeptide repeat-containing domain; IPR019734: Tetratricopeptide repeat (Repeat); PTHR14699: STI2 PROTEIN-RELATED (Family); SSF48452: TPR-like superfamily (superfamily)	6,33E-39	75.15%	<i>Strongylocentrotus purpuratus</i> (XP_793431.2)
G2/ GW314929 (G. digestiva)	steroid sulfatase	791	C:lysosome; catabolic process; P:epidermis development; C:endosome; C:microsome; F:steryl-sulfatase activity; C:endoplasmic reticulum; C:Golgi apparatus; C:plasma membrane EC:3.1.6.2	—	3,26E-39	65.85%	<i>Branchiostoma floridae</i> (EEN62448.1)

Clone/Nº de acesso no genbank (Tecido)	Descrição	Tamanho (pb)	Termos GO/ Número EC	InterproScan/CDD	Melhor <i>hit</i> de BLAST		
					Valor e	Similaridade	Espécie (Código de acesso)
G4/ GW314930 (G. digestiva)	mannose-p-dolichol utilization defect 1	674	P:oligosaccharide biosynthetic process	—	2,55E-38	57.9%	<i>Rattus norvegicus</i> (AAH58455.1)
B5/ GW314951 (Hemócitos)	FReD containing protein	742	P:neuromuscular development	junction cl00085 : FReD; Fibrinogen-related domains (FReDs): IPR002181 : Fibrinogen, alpha/beta/gamma chain, C-terminal globular; IPR014716 : Fibrinogen, alpha/beta/gamma chain, C-terminal globular, subdomain 1	1E-16	65.8%	<i>Branchiostoma floridae</i> (XP_002233635.1)

Clone/Nº de acesso no genbank (Tecido)	Descrição	Tamanho (pb)	Termos GO/ Número EC	InterproScan/CDD	Melhor <i>hit</i> de BLAST		
					Valor e	Similaridade	Espécie (Código de acesso)
C3/ GW314931 (G. digestive)	bcl2-like 1	779	F:protein binding; P:multicellular organismal development; P:response to organic substance; P:sexual reproduction; P:anti-apoptosis; P:anatomical structure development; C:organelle membrane; C:mitochondrial envelope; P:response to stress; P:reproductive process	cd06845: Bcl-2_like, IPR000712: Apoptosis regulator, Bcl-2, BH (Domain); IPR002475: BCL2-like apoptosis inhibitor (domain); G3DSA:1.10.437.10: Apoptosis Regulator Bcl-x (superfamily), PTHR11256: BCL-2 RELATED (family); SSF56854: Bcl-2 inhibitors of programmed cell death superfamily	1,43E-15	52.3%	<i>Nematostella vectensis</i> (XP_001629527.1)
H3/ GW314932 (G. digestiva)	TY (Thyroglobulin) repeat containing protein	748	P:cellular process	cd00191, TY, Thyroglobulin type I repeats	4,90E-10	41.75%	<i>Branchiostoma floridae</i> (XP_002607132.1)
H6/ GW314952 (Hemócitos)	ccr4-not transcription subunit 2	603	C:cytoplasm	IPR007282: NOT2/NOT3/NOT5	1E-08	82.1%	<i>Xenopus (Silurana) tropicalis</i> (NP_001106600.1)

Clone/Nº de acesso no genbank (Tecido)	Descrição	Tamanho (pb)	Termos GO/ Número EC	InterproScan/CDD	Melhor <i>hit</i> de BLAST		
					Valor e	Similaridade	Espécie (Código de acesso)
B4/ GW314933 (G. digestiva)	Yippee-like protein	643	—	IPR004910 : Yippee-like protein (family) PTHR13847 : FAD NAD BINDING OXIDOREDUCTASES (Family), PTHR13847:SF1 : YIPPEE PROTEIN (Family)	3,74E-52	93.5%	<i>Branchiostoma floridae</i> (XP_002596802.1)
F4/ GW314934 (G. digestiva)	Protein with an ab-hydrolase lipase region	724	—	pfam04083 , Abhydro_lipase, ab-hydrolase associated lipase region (superfamily)	1,00E-27	64.35%	<i>Trichoplax adhaerens</i> (XP_002111054.1)
A3/ GW314935 (G. digestiva)	Elongation Factor TS	677	—	pfam00889 , EF_TS, Elongation factor TS	6,73E-10	61.8%	<i>Branchiostoma floridae</i> (XP_002605109)
B1/ GW314936 (G. digestiva)	KRR1 interacting protein 1	698	—	IPR007851 :KRR1 interacting protein 1, subgroup	1,67E-06	65.5%	<i>Tribolium castaneum</i> (XP_973472.2)

Clone/Nº de acesso no genbank (Tecido)	Descrição	Tamanho (pb)	Termos GO/ Número EC	InterproScan/CDD	Melhor <i>hit</i> de BLAST		
					Valor e	Similaridade	Espécie (Código de acesso)
F3/ GW314937 (G. digestiva)	FAD binding domain containing protein	736	—	cl10516 , FAD_binding_4, FAD binding domain (superfamily)	1,00E-08	63,15%	<i>Penicillium marneffeii</i> ATCC 18224 (XP_002153034.1)
B2/ GW314938 (G. digestiva)	YbaK-like protein	784	—	cl00022 , YbaK_like, YbaK-like (superfamily)	1,00E-68	72,50%	<i>Crassostrea gigas</i> (CU995089.1)

APÊNDICE D - Genes de *C. rhizophorae* com *hits* de BLAST e sem anotação funcional.

Clone/ Nº de acesso no genbank	Tecido	Tamanho (pb)	Melhor hit				Valor e	Similaridade
			Descrição	Código de acesso	Algoritmo BLAST e Banco de dados			
Contig1 (C7, H7, D5 e H5)/ GU574706	Hemócitos	732	<i>crassostrea brasiliana</i> isolate cb20 16s ribosomal rna partial sequence mitochondrial	FJ478028.1	BLASTn - nr	0	93.5%	
D4/ GW314939	Glândula digestiva	572	AUF_104_G04_T7 <i>Crassostrea virginica</i> Gill <i>Crassostrea virginica</i> cDNA 5-, mRNA sequence	CD649023.1	BLASTn - dbEST others	1,00E-111	76,60%	
E5/ GW314970	Hemócitos	652	<i>Crassostrea virginica</i> mitochondrion, complete genome	AY905542.2	BLASTn - nr	1,00E-119	74,85%	
E1/ GW314940	Glândula digestiva	755	au_cv_est_007d_a08 oyster mixed tissue normalized cdna library <i>crassostrea virginica</i> cdna 5 mrna	EH645424.1	BLASTn - dbEST others	8,88E-109	75.5%	
D2/ GW314941	Glândula digestiva	738	au_cv_est_005a_h04 oyster mixed tissue normalized cdna library <i>crassostrea virginica</i> cdna 5 mrna	EH644559.1	BLASTn - dbEST others	4,22E-81	71%	
A8/ GW314953	Hemócitos	479	AM856481 cDN21 <i>Crassostrea gigas</i> cDNA clone cDN21P0002D14 5-, mRNA sequence	AM856481.1	BLASTn - dbEST others	1,00E-80	90%	

Clone/ Nº de acesso no genbank	Tecido	Tamanho (pb)	Melhor hit				
			Descrição	Código de acesso	Algoritmo BLAST e Banco de dados	Valor e	Similaridade
B6/ GW314954	Hemócitos	514	CU995089 Pascal Favrel cDNA library all stages (wy0aaa) <i>Crassostrea gigas</i> cDNA clone wy0aaa35a05, mRNA sequence	CU995089.1	BLASTn - dbEST others	1,00E-75	80%
A7/ GW314955	Hemócitos	640	am869282 cdn37 <i>crassostrea gigas</i> cdna clone cdn37p0036c21 5 mrna	AM869282.1	BLASTn - dbEST others	1,00E-52	77%
B7/ GW314956	Hemócitos	582	AM859065 cdn37 <i>Crassostrea gigas</i> cDNA clone cdn37P0005P24 5-, mRNA sequence	AM859065.1	BLASTn - dbEST others	1,00E-38	80%
D1/ GW314942	Glândula digestiva	762	regeneration-upregulated protein 2 [<i>Enchytraeus japonensis</i>]	BAE93514.1	BLASTx - nr	1,00E-38	47%
F2/ GW314943	Glândula digestiva	699	am865904 cdn37 <i>crassostrea gigas</i> cdna clone cdn37p0026d06 5 mrna	AM865904.1	BLASTn - dbEST others	9,41E-26	78.5%
H1/ GW314944	Glândula digestiva	644	mitochondrial 28S ribosomal protein S18B, putative [<i>Pediculus humanus corporis</i>]	XP_002426373.1	BLASTx - nr	9,92E-21	57.5%
E8/ GW314957	Hemócitos	587	AM859854 cdn37 <i>Crassostrea gigas</i> cDNA clone cdn37P0008F17 5-, mRNA sequence	AM859854.1	BLASTn - dbEST others	1,00E-19	83%
B8/ GW314958	Hemócitos	745	<i>Ostrea edulis</i> microsatellite HA10 sequence	AF297863.1	BLASTn - nr	2,75E-14	81,30%

Clone/ Nº de acesso no genbank	Tecido	Tamanho (pb)	Melhor hit				Valor e	Similaridade
			Descrição	Código de acesso	Algoritmo BLAST e Banco de dados			
D8/ GW314959	Hemócitos	543	hypothetical protein BRAFLDRAFT_68846 [<i>Branchiostoma floridae</i>]	XP_002206201.1	BLASTx - nr	1,00E-10	56%	
E4/ GW314945	Glândula digestiva	682	PREDICTED: similar to glycerol-3-phosphate acyltransferase [<i>Strongylocentrotus purpuratus</i>]	XP_790882.1	BLASTx - nr	2,29E-05	55.75%	

APÊNDICE E – Sequências que não apresentaram *hits*.

Clone	Nº de acesso no genbank
<i>Crassostrea rhizophorae</i> (Hemócitos)	
D7	GW314961
E7	GW314962
F7	GW314963
G7	GW314964
G8	GW314965
H8	GW314966
C6	GW314967
A6	GW314968
D6	GW314969
F5	GW314971
C5	GW314972
<i>Crassostrea rhizophorae</i> (Glândula digestiva)	
C1	GW314946
A2	GW314947
H2	GW314948
G3	GW314949
C8	GW314960
<i>Nodipecten nodosus</i> (glândula digestiva)	
E9	GW314989
B10	GW314990
D9	GW314991
F9	GW314992
H9	GW314993
G10	GW314994
A10	GW314995
C11	GW314996
F11	GW314997
D10	GW392490

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)