

Michael Lorenz Jaramillo Bobadilla

**IDENTIFICAÇÃO DE TRANSCRITOS EM EMBRIÕES DE  
*Macrobrachium olfersii*: ÊNFASE NOS GENES *Hox* E DE  
SEGMENTAÇÃO E SUA EXPRESSÃO DURANTE O  
DESENVOLVIMENTO EMBRIONÁRIO**

Tese submetida ao Programa de Pós-graduação em Biologia Celular e do Desenvolvimento da Universidade Federal de Santa Catarina para a obtenção do Grau de Doutor em Biologia Celular e do Desenvolvimento.

Orientadora: Prof.<sup>a</sup> Dr.<sup>a</sup> Yara Maria Rauh Müller

Co-orientador: Prof. Dr. Dib Ammar

Florianópolis  
2016

Ficha de identificação da obra elaborada pelo autor,  
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Bobadilla, Michael Lorenz Jaramillo  
Identificação de transcritos em embriões de  
Macrobrachium olfersii: ênfase nos genes Hox e de  
segmentação e sua expressão durante o desenvolvimento  
embrionário / Michael Lorenz Jaramillo Bobadilla ;  
orientadora, Yara Maria Rauh Müller ; coorientador, Dib  
Ammar. - Florianópolis, SC, 2016.  
238 p.

Tese (doutorado) - Universidade Federal de Santa  
Catarina, Centro de Ciências Biológicas. Programa de Pós  
Graduação em Biologia Celular e do Desenvolvimento.

Inclui referências

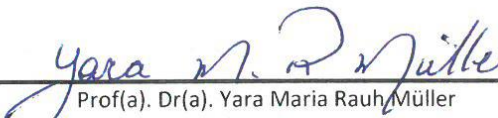
1. Biologia Celular e do Desenvolvimento. 2. Desenvolvimento embrionário de crustáceos. 3. Controle Molecular. 4. Segmentação e identidade segmentar. I. Müller, Yara Maria Rauh. II. Ammar, Dib. III. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Biologia Celular e do Desenvolvimento. IV. Título.

**"Identificação de transcritos em embriões de *Macrobrachium olfersii*: ênfase nos genes *Hox* e de segmentação e sua expressão durante o desenvolvimento embrionário"**

Por

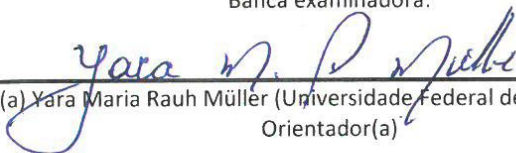
**Michael Lorenz Jaramillo Bobadilla**

Tese julgada e aprovada em sua forma final pelos membros titulares da Banca Examinadora (12/PPGBCD/2016) do Programa de Pós-Graduação em Biologia Celular e do Desenvolvimento - UFSC.

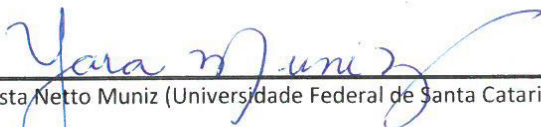


Prof(a). Dr(a). Yara Maria Rauh Müller  
Coordenador(a) do Programa de Pós-Graduação em Biologia Celular e do Desenvolvimento

Banca examinadora:



Dr(a) Yara Maria Rauh Müller (Universidade Federal de Santa Catarina)  
Orientador(a)



Dr(a) Yara Costa Netto Muniz (Universidade Federal de Santa Catarina)



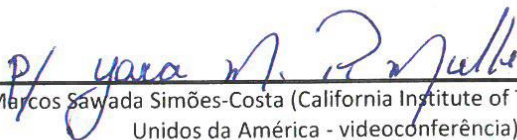
Dr(a) Patrícia Hermes Stoco (Universidade Federal de Santa Catarina)



Dr(a) Afonso Celso Dias Bainy (Universidade Federal de Santa Catarina)



Dr(a) Arnaldo Zaha (Universidade Federal do Rio Grande do Sul)



Dr(a) Marcos Sawada Simões-Costa (California Institute of Technology- Estados Unidos da América - videoconferência)

Florianópolis, 12 de Agosto de 2016.



*A meus pais Lorenzo e Isabel, que me ensinaram desde criança que o conhecimento (educação) é uma ferramenta muito poderosa para avançar. Esta tese é dedicada a eles e é o fruto da educação que me deram.*



## AGRADECIMENTOS

Meus agradecimentos sinceros a todos que contribuíram direta ou indiretamente para que este estudo pudesse ser desenvolvido. Em especial, gostaria de agradecer:

A minha orientadora Dra. Yara Müller por acreditar em minha capacidade e depositar sua confiança em mim para a realização do presente trabalho, pelas sugestões, pela disponibilidade do seu tempo e por me incentivar dizendo que ao final da tese teríamos um bom trabalho.

A meu co-orientador Dr Dib Ammar, pelo apoio nos experimentos, pois cada experimento era um novo desafio. Também, obrigado pelas sugestões e discussões científicas durante o tempo de doutorado que contribuíram para o presente trabalho.

A Professora Dra. Evelise Nazari pela disponibilidade em responder minhas perguntas e pelas suas sugestões. Além disso, sua participação foi fundamental nos artigos científicos produto desta tese.

Aos professores da Pós-graduação de Biologia Celular e do Desenvolvimento, que colaboraram com minha formação. À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela bolsa de Doutorado. Ao Programa de Pós-graduação Biologia Celular e do Desenvolvimento pelo financiamento aos congressos.

Aos Professores e aos colegas dos diferentes laboratórios onde foi realizado parte dos experimentos desta tese, pois sem essas parceiras não teria terminado esta tese. Ao Laboratório de Neuroquímica I, ao Laboratório de Polimorfismos Genéticos, ao Laboratório de Protozoologia e ao Laboratório de Neurogenética do Desenvolvimento. Também, ao pessoal do Laboratório Multiusuário de Estudos em Biologia.

Ao Prof. Dr. Rogerio Margis e ao Dr. Frank Guzman porque me ensinaram e me ajudaram na montagem e análise transcriptômica.

A minha família, pelo carinho, compreensão e incentivo, pois mesmo estando longe sempre foram uma força para me fazer seguir em frente e solucionar diferentes problemas dos experimentos.

A Ruth, meu amor, minha companheira e amiga, que sempre acreditou que poderia contribuir para a ciência. Pela sua compreensão em relação ao grande tempo que dediquei a esta tese. Pelas sugestões, e discussões científicas e ajuda em resolver alguns problemas experimentais desta tese. Obrigado, já que me motivou para fazer o doutorado nesta universidade.

A Ana e Mauro pelos momentos compartilhados aqui em Florianópolis.

Aos meus colegas do Laboratório de Reprodução e Desenvolvimento animal, cujos nomes não mencionarei para não me esquecer de nenhum deles. Pelas trocas de ideias científicas que ajudaram no meu desenvolvimento profissional.

A Deus por dar-me tranquilidade, alegria e por renovar minhas energias para seguir em frente.

A todos aqueles que colaboraram na realização deste trabalho.



*“The important thing is not to stop questioning. Curiosity has its own reason for existing...”*

*Albert Einstein*



## RESUMO

Os estudos dos mecanismos moleculares do desenvolvimento em crustáceos são importantes para entender a diversidade morfológica deste grupo. No entanto, poucos estudos foram realizados, se consideramos a diversidade deste grupo. Dentre os camarões de água doce, *Macrobrachium olfersii* surge como um modelo potencial para estudos de desenvolvimento e de toxicidade. Porém, os mecanismos moleculares durante o desenvolvimento ainda não foram descritos, pois poucas sequências são conhecidas nesta espécie. Neste contexto, o presente estudo teve os seguintes objetivos: 1) identificar transcritos com a finalidade de gerar um banco de sequências e contribuir no entendimento dos mecanismos moleculares nesta espécie; 2) identificar transcritos de genes relacionados ao desenvolvimento, em especial os genes *Hox* e genes de segmentação; 3) e analisar os níveis de seus transcritos durante o desenvolvimento embrionário de *M. olfersii*. Uma montagem *de novo* a partir de 25,6 milhões de *reads* e análise transcriptômica de *pool* de embriões (E4-E8) proporcionou 99.751 unigenes. Do total, 20,95% foram similares às sequências do banco de dados *GenBank*. Uma diversidade de transcritos foi classificada por ontologia gênica (21.845 unigenes) e em 129 vias (6.866 unigenes) do KEGG (*Kyoto Encyclopedia of Genes and Genomes*). Os transcritos de genes codificados pelo DNA mitocondrial (22 tRNA, 2 rRNA e 13 genes codificantes de proteínas) também foram identificados. Na análise do transcriptoma utilizando ontologia gênica, 2.142 unigenes foram classificadas na categoria de processos do desenvolvimento e 553 unigenes na subcategoria do desenvolvimento embrionário, cujas sequências foram mais similares aos artrópodes *Zootermopsis nevadensis* e *Limulus polyphemus*. Além disso, os transcritos de genes do padrão ântero-posterior, do padrão dorso-ventral, de vias de sinalização envolvidas no desenvolvimento (TGF- $\beta$ , Wnt, Notch, MAPK, Hedgehog, Jak-STAT, VEGF) e dos receptores nucleares induzíveis por ecdiesteróides foram identificados. Através da PCR com iniciadores degenerados foram identificados transcritos dos genes *Hox* (*Lab*, *Dfd*, *Antp*, *Ubx*, *AbdA*, *AbdB*), genes de segmentação (*Eve* e *Wg*) e da família *Wnt* (*Wnt2*, *Wnt4*, *Wnt5*, *Wnt7*, *Wnt11*). Estes resultados indicam que os mecanismos moleculares em *M. olfersii* podem ser similares a outros crustáceos estudados. Além disso, parálogos dos genes *En* (*En1* e *En2*) e *Eve* (*Eve1* e *Eve2*) foram identificados, sendo que os de *Eve* foram pela primeira vez identificados em crustáceos. Ainda, isoformas de *En* (*En1a* e *En1b*) e de 4 genes *Hox* (*Dfd1a* e

*Dfd1b*, *Scr1a* e *Scr1b*, *Ubx1a* e *Ubx1b*, *AbdA-1a* e *AbdA-1b*) foram identificados provavelmente resultantes de *splicing* alternativo. Para analisar os níveis de transcritos de genes *Hox* e de segmentação no desenvolvimento, primeiramente foi avaliada a estabilidade de 6 genes de referência utilizando programas computacionais e qPCR. A expressão dos genes *Rpl8* e *Rps6* foi mais estável durante o desenvolvimento e nos tecidos de animais adultos, respectivamente. Enquanto que, a expressão dos genes *Gapdh* e *Ef-1a* foi menos estável no desenvolvimento. Os genes *Rpl8* ou *Rps6* e suas combinações são mais indicados para serem utilizados como genes de referência, na análise durante o desenvolvimento, por RT-qPCR. Da análise dos níveis de transcritos de genes relacionados ao desenvolvimento, os transcritos de 3 genes *Hox* (*Lab*, *Dfd* e *Ftz*) diminuíram com o avanço do desenvolvimento, enquanto que os transcritos de 6 genes *Hox* (*Pb*, *Scr*, *Antp*, *Ubx*, *AbdA* e *AbdB*) aumentaram com o avanço do desenvolvimento, o que estaria relacionado com a função da identidade dos segmentos da cabeça, tórax e abdome. Os genes *En1*, *En2* e *Eve2* estariam relacionados com a segmentação, aumentando seus transcritos com a formação de novos segmentos do corpo durante o desenvolvimento. Enquanto que, os transcritos do gene *Eve1* diminuíram, o que estaria relacionado com uma função em estágios mais iniciais. No entanto, *En* e *Eve* também poderiam participar na neurogênese similar a outros artrópodes estudados. Os genes *Wg* e *Hh* funcionariam como genes de polaridade segmentar mantendo seus níveis de transcritos similares com o avanço do desenvolvimento. Os níveis de transcritos do gene *Dll* foi similar entre os estágios e estaria envolvido na formação dos apêndices durante o desenvolvimento como acontece em outros crustáceos. Assim, o presente estudo contribuiu com a primeira identificação em larga escala de transcritos presentes em embriões de *M. olfersii*, em especial os transcritos de 9 genes *Hox* e de outros genes relacionados ao desenvolvimento. Além disso, proporciona informação da avaliação de genes de referência para análises de qPCR nesta espécie, que permitirá resultados mais precisos para entender os mecanismos moleculares durante o desenvolvimento embrionário. Da mesma forma, a identificação destas sequências em *M. olfersii* contribuirá para futuros estudos em diferentes linhas de pesquisa e poderá ainda servir como referência para identificar genes em outros camarões de água doce.

**Palavras-chave:** Transcriptoma. *Hox*. Segmentação. Gene de referência. qPCR. Expressão gênica.

## ABSTRACT

Studies of the molecular mechanisms of development in crustaceans are important to understand the morphological diversity of this group. However, few studies have been conducted if we consider the diversity of this group. Among the freshwater prawns, *Macrobrachium olfersii* emerges as a potential model for development and toxicity studies. However, the molecular mechanisms during development were not described because few sequences are known in this species. In this context, this study had the following objectives: 1) to identify transcripts in order to generate a sequence data bank and contribute to the understanding of the molecular mechanisms in this species, 2) to identify transcripts of gene related to the development, in particular *Hox* genes and segmentation gene, and 3) analyze the levels of their transcripts during embryonic development of *M. olfersii*. *De novo* assembly from 25.6 million reads and transcriptomic analysis of embryos pool (E4-E8) provided 99.751 unigenes. Of the total, 20.95% were similar to sequences of GenBank database. A diversity of transcripts were classified by gene ontology (21.845 unigenes) and 129 KEGG (*Kyoto Encyclopedia of Genes and Genomes*) pathways (6.866 unigenes). The transcripts of genes encoded by mitochondrial DNA (22 tRNA, 2 rRNA and 13 protein-coding genes) were also identified. In the transcriptome analysis using gene ontology, 2,142 unigenes were assigned to the development processes category and 553 unigenes in embryonic development subcategory, whose sequences were more similar to the *Zootermopsis nevadensis* e *Limulus polyphemus* arthropods. Furthermore, transcripts of anterior-posterior pattern genes, of dorsal-ventral pattern genes and signaling pathways involved in the development (TGF- $\beta$ , Wnt, Notch, MAPK, hedgehog, Jak-STAT VEGF) and ecdysteroid-inducible nuclear receptors were identified. Moreover, by PCR with degenerate primers were identified transcripts of *Hox* genes (*Lab Dfd*, *Antp*, *Ubx*, *AbdA*, *AbdB*), segmentation genes (*Eve* and *Wg*) and *Wnt* family (*Wnt2*, *Wnt4*, *Wnt5*, *Wnt7*, *Wnt11*). These results indicate that the molecular mechanisms in *M. olfersii* could be similar to other crustaceans studied. Moreover, paralogs of genes *En* (*En1* and *En2*) and *Eve* (*Eve1* and *Eve2*) were identified, and the *Eve* paralogs were first identified in crustaceans. In addition, isoforms of *En* (*En1a* and *En1b*) and of 4 *Hox* genes (*Dfd1a* and *Dfd1b*, *Scr1a* and *Scr1b*, *Ubx1a* and *Ubx1b*, *AbdA-1a* and *AbdA-1b*) were identified probably resulting from alternative splicing. To analyze transcript levels of *Hox* genes and segmentation genes in development, it was first

evaluated the expression stability of reference genes using computer programs and qPCR. Expression of *Rpl8* and *Rps6* genes was more stable embryonic development and in the adult animal tissues, respectively, whereas the expression of *Gapdh* and *Ef-1 $\alpha$*  genes was less stable in development. *Rpl8* and *Rps6* genes or its combinations are most appropriate for use as reference gene for RT-qPCR analysis in development. In the analysis of transcript levels of genes related to the development, the transcripts of 3 *Hox* genes (*Lab*, *Dfd* and *Ftz*) decreased and the transcripts 6 *Hox* genes (*Pb*, *Scr*, *Antp*, *Ubx*, *AbdA* and *AbdB*) increased its expression with the progress of the development, which is related to its function in the identity of the segments of the head, thorax and abdomen. The genes *En1*, *En2* and *Eve2* would be related to the segmentation, increasing its transcripts with formation of new segments of the body during the development. Whereas the transcripts of *Eve1* gene decreased that would be related to a function in most early stages. However, *En* and *Eve* could also participate in neurogenesis similar to other arthropods studied. The *Wg* and *Hh* gene would function as segment polarity genes, maintaining similar their transcript levels with the progress of development. The transcript levels of gene *Dll* was very similar between the stages and would be involved in the formation of appendages during development as in other crustaceans studied. Thus, this study contributed with the first large-scale identification of transcripts in *M. olfersii* embryos, in particular transcripts of 9 *Hox* genes and other genes related to development. It also provides information of the reference genes evaluation for qPCR analysis in this species, allowing more accurate results to understand the molecular mechanisms during embryonic development. In addition, the identification of these sequences in *M. olfersii* will contribute in future studies in different lines of research and could serve as a reference to identify genes in other freshwater prawns.

**Keywords:** Transcriptome. *Hox*. Segmentation. Reference gene. qPCR. Gene expression.

## LISTA DE FIGURAS

**Figura 1. Tipos de banda germinativa observadas em embriões de insetos.** Os ovos são esquematizados em escala e em vista lateral, com a região anterior na parte superior. A área do embrião rudimentar é sombreada. Orthoptera: a, *Oecanthus pellucens*; b, *Acheta domesticus*. Odonata: c, *Platycnemis pennipes*. Hemiptera: d, *Euscelis plebejus*. Coleoptera: e, *Atrachya menetriesi*; f, *Leptinotarsa decemlineata*; h, *Bruchidius obtectus*. Lepidoptera: g, *Bombyx mori*. Diptera: i, *Smittia* sp.; j, *D. melanogaster*; k, *Calliphora erythrocephala*. Hymenoptera: l, *Apis mellifera*. Fonte: Adaptado de Davis e Patel (2002).....4

**Figura 2. Padronização ao longo do eixo Ântero-Posterior (A/P) do embrião de *Drosophila melanogaster*.** Uma cascata de genes maternos e zigóticos é ativada no embrião sincicial para subdividir o ectoderma em domínios menores. Ocorre intensa proliferação celular após a ativação dos genes *pair-rule*. Os genes de polaridade de segmentos e os genes *Hox* são ativados pelos genes *pair-rule*, mas um subconjunto de genes *gap* também influencia diretamente aos genes *Hox*. Ambos os genes de polaridade de segmento e *Hox* atuam em conjunto para controlar a diferenciação de cada segmento da futura larva (SANSON, 2001). Fonte: Adaptado de Sanson (2001).....5

**Figura 3. Expressão de genes *pair-rule* e de polaridade segmentar em *D. melanogaster*.** Embrião inicial (parte superior) indicando os parasegmentos e os transcritos de *Even-skipped*, *Ftz*, *Engrailed* e *Hedgehog*. Embrião tardio (parte inferior) indicando os segmentos e os transcritos de *Engrailed*, *Hedgehog*, *Wingless* e *Patched*. Durante o desenvolvimento de *D. melanogaster*, *Wg* é transcrita e secretada de uma fileira anterior de células e mantém os níveis de transcritos de *En* em células ectodermiais adjacentes, na região posterior. As células que expressam *En*, por sua vez, transcrevem e secretam o ligante *Hh*, que mantém mutuamente os níveis de transcritos de *Wg* nas células anteriores vizinhas (DINARDO et al., 1988; BEJSOVEC; MARTINEZ ARIAS, 1991). A interface entre estes dois domínios adjacentes define o limite do parasegmento, com transcritos de *En/Hh* na região anterior e *Wg* na região posterior final de cada parasegmento (BAKER, 1987; LEE et al., 1992; MOHLER; VANI, 1992).....8

**Figura 4. Genes *Hox* e sua expressão em crustáceos e insetos.** ocular (Oc); antena (Ant); mandibular (Mn); maxilar (Mx); tronco em crustáceo (T1-T14); telson (Te); lábio (Lb); intercalar (Int); primeiro ao terceiro torácico (T1-T3); abdominal 1 - abdominal 10 em insetos (A1-A10). Pontos de interrogação para *Hox3* e *Ftz* indicam que as funções

destes genes em crustáceos ainda não são conhecidos. Nos insetos, *Hox3* e *Ftz* tem diferentes funções e não são apresentados na figura. Fonte: Adaptado de Hughes e Kaufman (2002).....11

**Figura 5. Padrões de expressão dos genes *Hox* em crustáceos.** Os segmentos do corpo são indicados no eixo ântero-posterior. Baseado nos estudos de alguns genes *Hox*, a expressão de cada gene *Hox* é mostrada nas diferentes espécies de crustáceos. Vista lateral de *A. franciscana* e *P. hawaiiensis*, e vista dorsal de *P. scaber* e *P. clarkii* são mostrados no lado direito. Fonte: Adaptado de Hughes e Kaufman (2002).....13

**Figura 6. Espécime de *M. olfersii* e sua distribuição geográfica.** A) Vista lateral do animal adulto. B) Vista lateral da fêmea transportando uma grande quantidade de ovos na câmara incubadora externa. Os ovos estão indicados pela seta vermelha. Barra de escala: 1cm. C) Distribuição geográfica de *M. olfersii* na costa leste das Américas. Fonte: Acervo do Laboratorio de Reprodução e Desenvolvimento Animal-UFSC.....15

**Figura 7. Desenvolvimento embrionário de *M. olfersii*.** Dia embrionário (E). Vista ventral (E2-E6). Vista lateral (E7-E14). Antênula (an), antena (at), área blastoporal (a), blastômeros (bl), sulcos de clivagem (cf), papila caudal (cp), olho (ey), disco germinativo (gd), mandíbula (mn), lobo óptico (ol), apêndice pós-naupliar (pa), estomodeu (st), télson (te), vitelo (yk), massa de vitelo (ym). Fonte: adaptado de Müller; Nazari; Simões-Costa (2003).....16

**Figura 8. Ciclo de vida de *M. olfersii*.** Alguns estágios do desenvolvimento embrionário e larvais são mostrados. O tamanho dos embriões, larvas, juvenis e adultos não estão em proporção de tamanho. As ilustrações dos embriões foram de Müller et al. (2003) e dos adultos foram cortesia do Dr. Ammar. As ilustrações das larvas foram de Dugger e Dobkin (1975).....18

**Figura 9. Resumo de materiais e métodos utilizados no presente estudo.** As cores azul, rosa, verde e roxo correspondem aos métodos utilizados no Capítulo I, Capítulo II, Capítulo III e Capítulo IV, respectivamente.....24

**Figura 10. Desenho esquemático do processo de montagem *de novo* dos unigenes.** Depois do sequenciamento na plataforma Illumina®, os *reads* (sequências geradas do sequenciamento) são utilizados para gerar os *contigs* (sequências geradas na montagem *de novo* derivadas da coleção de *reads* sobrepostos) com diferentes tamanhos ou valores de k-mer (k-mer é definido como subseqüências geradas por decomposição, o qual é utilizado para procurar sobreposição com outros *reads*). Os *contigs* gerados com os diferentes valores de k-mer são agrupados em



unigenes (sequências finais obtidas da montagem *de novo* que são derivadas da montagem agrupada depois da remoção de *contigs* redundantes). As caixas da cor cinza indicam os programas utilizados.....30

**Figura 11. Distribuição do comprimento dos unigenes montados de *M. olfersii* em função do número de unigenes.** Diferentes comprimentos dos unigenes foram obtidos para os 99.751 unigenes. Aproximadamente, 34,5% dos unigenes contém mais do que 1000 pb.....33

**Figura 12. Distribuição dos melhores *hits* (*Top Hits*) dos unigenes de *M. olfersii* contra as sequências de todas as espécies do banco de dados de proteínas NR.** O algoritmo Blastx ( $E$ -value  $<1e^{-5}$ ) foi utilizado para busca de similaridade com sequências de proteínas NR do NCBI. As espécies com melhores *hits* e o número de unigenes homólogos são mostradas. Das 28 espécies melhor classificadas, quatro são decápodes malacostracos de um total de cinco táxons de crustáceos (17,9%). Outros correspondem à soma de unigenes com *hits* a proteínas em outras espécies.....34

**Figura 13. Histograma da classificação GO de unigenes anotadas do transcriptoma de *M. olfersii*.** Um maior número de unigenes foi obtido para processo celular, núcleo e proteínas de ligação em cada categoria.....37

**Figura 14. Unigenes de *M. olfersii* classificados em vias biológicas do KEGG.** O número de enzimas em cada via é indicado entre parênteses.....38

**Figura 15. Transcritos do genoma mitocondrial de *M. olfersii* e a organização de genoma mitocondrial de Pancrustacea.** As caixas de cor branca indicam os genes que codificam proteínas. As caixas de cor cinza indicam os genes de tRNA. As caixas da cor preta indicam os genes de rRNA. Os genes de tRNA são designados pelo código de uma letra para cada aminoácido específico. As caixas de acima e abaixo na representação do genoma mitocondrial de Pancrustacea indica a localização dos genes em uma ou outra fita do genoma. A localização dos genes no genoma de *M. olfersii* não é conhecida.....40

**Figura 16. Eletroforese em gel de agarose (1,5%) dos produtos da RT-PCR para os transcritos dos genes *Hox* e de segmentação.** Diferentes temperaturas foram testadas para padronizar as condições de PCR. Para alguns genes (*Eve*, *En*, *Dfd*, *Pb*, *Wg* e *Wnt2*) múltiplas bandas foram observadas. Mk: Marcador de peso molecular de DNA (100 pb, Kasvi). As setas indicam as bandas que foram selecionadas para cada

gene. (*) Iniciadores diferentes foram utilizados para a amplificação de transcritos de <i>Dfd</i> .....	56
<b>Figura 17. Via de sinalização Jak-STAT relacionada ao desenvolvimento embrionário em <i>M. olfersii</i>.</b> Transcritos dos genes de cada via identificados em <i>M. olfersii</i> estão indicados pela cor cinza. Os mapas das vias foram obtidos do banco do KEGG.....	60
<b>Figura 18. Via de sinalização MAPK relacionada ao desenvolvimento embrionário em <i>M. olfersii</i>.</b> Transcritos dos genes de cada via identificados em <i>M. olfersii</i> estão indicados pela cor cinza. Os mapas das vias foram obtidos do banco do KEGG.....	61
<b>Figura 19. Via de sinalização Notch relacionada ao desenvolvimento embrionário em <i>M. olfersii</i>.</b> Transcritos dos genes de cada via identificados em <i>M. olfersii</i> estão indicados pela cor cinza. Os mapas das vias foram obtidos do banco do KEGG.....	62
<b>Figura 20. Via de sinalização Hedgehog relacionada ao desenvolvimento embrionário em <i>M. olfersii</i>.</b> Transcritos dos genes de cada via identificados em <i>M. olfersii</i> estão indicados pela cor cinza. Os mapas das vias foram obtidos do banco do KEGG.....	63
<b>Figura 21. Via de sinalização VEGF relacionada ao desenvolvimento embrionário em <i>M. olfersii</i>.</b> Transcritos dos genes de cada via identificados em <i>M. olfersii</i> estão indicados pela cor cinza. Os mapas das vias foram obtidos do banco do KEGG.....	64
<b>Figura 22. Via de sinalização TGF-<math>\beta</math> relacionada ao desenvolvimento embrionário em <i>M. olfersii</i>.</b> Transcritos dos genes de cada via identificados em <i>M. olfersii</i> estão indicados pela cor cinza. Os mapas das vias foram obtidos do banco do KEGG.....	65
<b>Figura 23. Via de sinalização Wnt relacionada ao desenvolvimento embrionário em <i>M. olfersii</i>.</b> Transcritos dos genes de cada via identificados em <i>M. olfersii</i> estão indicados pela cor cinza. Os mapas das vias foram obtidos do banco do KEGG.....	66
<b>Figura 24. Espécies com melhores <i>hits</i> para os unigenes relacionados ao desenvolvimento de <i>M. olfersii</i>.</b> A) Unigenes da categoria processo de desenvolvimento e B) na subcategoria de desenvolvimento embrionário obtidas da análise transcriptômica.....	68
<b>Figura 25. Alinhamento de seqüências de aminoácidos de genes <i>Hox</i> (<i>Lab</i>, <i>Pb</i>, <i>Dfd</i>, <i>Scr</i>) em artrópodes utilizando ClustalW.</b> Regiões de cor preta indicam aminoácidos similares e as regiões de cor cinza indicam aminoácidos idênticos. As linhas indicam os <i>gaps</i> nas seqüências. O motivo YPWM é localizado <i>upstream</i> da região do homeodomínio dos genes <i>Pb</i> , <i>Dfd</i> , <i>Scr</i> . O número de acesso das seqüências são as mesmas que estão indicadas na árvore filogenética	

(Figura 28). A sequência obtida no transcriptoma (\*) ou na estratégia baseada em RT-PCR (\*\*\*) estão indicadas.....70

**Figura 26. Alinhamento de sequências de aminoácidos de genes *Hox* (*Ftz*, *Antp*, *Ubx*, *AbdA*, *AbdB*) em artrópodes utilizando ClustalW.**

Sequências de aminoácidos sublinhados indicam as regiões utilizadas para o desenho de iniciadores degenerados para RT-PCR. Regiões de cor preta indicam aminoácidos similares e as regiões de cor cinza indicam aminoácidos idênticos. As linhas indicam os *gaps* nas sequências. O motivo YPWM é localizado *upstream* da região do homeodomínio dos genes *Hox*: *Antp*, *Ubx* e *AbdA*. O número de acesso das sequências são as mesmas que estão indicadas na árvore filogenética (Figura 28 e 29). A sequência obtida no transcriptoma (\*) ou na estratégia baseada em RT-PCR (\*\*\*) são indicadas. As setas indicam os aminoácidos que são reconhecidos pelo anticorpo FP6.87.....71

**Figura 27. Alinhamento múltiplo e filogenia dos produtos dos genes *Hox* de *M. olfersii*.**

A) Regiões de cor preta no alinhamento indicam aminoácidos similares e as regiões de cor cinza indicam aminoácidos idênticos. As linhas indicam os *gaps* nas sequências. B) Árvore filogenética elaborada com sequências das proteínas *Hox* de *M. olfersii* (Mo), *L vannahmei* (Lv) e *P. hawaiiensis* (Ph). As árvores foram geradas através do programa MEGA v.5.05 pelo método de *Neighbour-Joining* com análises de *bootstrap* de 1000 replicatas. A barra de escala indica a quantidade de mudanças (número de mudanças por 100 sítios). Valores de *bootstrap* de vários nodos são mostrados. Os símbolos indicam um gene específico.....72

**Figura 28. Árvores filogenéticas elaboradas com sequências dos genes *Hox* (*Lab*, *Dfd*, *Scr* e *Ftz*) de várias espécies de artrópodes.**

A) *Lab*, B) *Dfd*, C) *Scr*, E) *Ftz*. Se utilizou o método de *Neighbour-Joining* implementado no programa MEGA v.5.05 com análises de *bootstrap* de 1000 replicatas. Os valores de *bootstrap* de vários nodos são mostrados. A barra de escala indica a quantidade de mudanças (número de mudanças por 100 sítios). As sequências do presente estudo são indicadas em cor preta. Ortólogos destes genes de *Euperipatoides kanangrensis* (Onychophora) foram utilizados como *outgroup*.....73

**Figura 29. Árvore filogenética elaborada com sequências dos genes *Hox* (*Antp*, *Ubx*, *AbdA* e *AbdB*) de várias espécies de artrópodes.**

A) *Antp*, B) *Ubx*, C) *AbdA*, D) *AbdB*. Se utilizou o método de *Neighbour-Joining* implementado no programa MEGA v.5.05 com análises de *bootstrap* de 1000 replicatas. Os valores de *bootstrap* de vários nodos são mostrados. A barra de escala indica a quantidade de mudanças (número de mudanças por 100 sítios). As sequências do

presente estudo são indicadas em cor preta. Ortólogos destes genes de *E. kanangrensis* (Onychophora) foram utilizados como *outgroup*.....74

**Figura 30. Alinhamento múltiplo de sequências de aminoácidos traduzida dos genes *Eve* e *En* de espécies de artrópodes utilizando ClustalW.** A) *Eve* e B) *En*. Regiões de cor preta indicam aminoácidos similares e as regiões de cor cinza indicam aminoácidos idênticos. As linhas indicam os *gaps* nas sequências. O número de acesso das sequências são as mesmas que estão indicadas na árvore filogenética (Figura 33). A sequência obtida no transcriptoma (\*) ou na estratégia baseada na RT-PCR (\*\*) são indicadas. O epítipo de *Eve* e *En* para os anticorpos 2B8 e 4D9 são mostrados no alinhamento.....76

**Figura 31. Alinhamento múltiplo de sequências de Hedgehog em artrópodes utilizando ClustalW.** Regiões de cor preta indicam aminoácidos similares e as regiões de cor cinza indicam aminoácidos idênticos. As linhas indicam os *gaps* nas sequências. Os domínios da proteína são indicados. O número de acesso das sequências são as mesmas que estão indicadas na árvore filogenética (Figura 33).....77

**Figura 32. Alinhamento múltiplo de sequências de aminoácidos traduzidas do gene *Wg* em artrópodes utilizando ClustalW.** Regiões de cor preta indicam aminoácidos similares e as regiões de cor cinza indicam aminoácidos idênticos. O número de acesso das sequências são as mesmas que estão indicadas na árvore filogenética. *Mysidium columbiae* (AAL37756.1), *A. franciscana* (CAO82108.1), *Thamnocephalus platyurus* (ALL53296.1), *Triops longicaudatus* (AAC32377.1), *D. pulex* (EFX86386.1), *D. magna* (BAJ05334.1), *T. castaneum* (EFA04660.1), *C. salei* (CAC87040.1), *Haematopinus tuberculatus* (ADG95819.1), *D. melanogaster* (NP\_523502.1).....78

**Figura 33. Árvores filogenéticas elaboradas com as sequências de aminoácidos ou de nucleotídeos dos genes *Eve*, *En*, *Hh* e da família *Wnt* de espécies de artrópodes.** Sequências nucleotídicas foram usadas para a árvore do gene *En*. Sequências de aminoácidos foram usadas para as árvores dos genes *Eve*, *En*, *Hh* e da família *Wnt*. A) *En*, B) *Eve*, C) *Hh* D) família *Wnt* e. Utilizou-se o método de Neighbour-Joining implementado no programa MEGA v.5.05 com análises de *bootstrap* de 1000 replicatas. Os valores de *bootstrap* de vários nodos são mostrados. A barra de escala indica a quantidade de mudanças (número de mudanças por 100 sítios). As sequências do presente estudo são indicadas em cor preta.....79

**Figura 34. Alinhamento múltiplo de sequências de aminoácidos da família *Wnt* de *M. olfersii* utilizando ClustalW.** Regiões de cor preta indicam aminoácidos similares e as regiões de cor cinza indicam

aminoácidos idênticos. As linhas indicam os *gaps* nas sequências. Mo: *M. olfersii*.....80

**Figura 35. Genes *Hox* em crustáceos.** A caixa de cor preta indica as sequências de transcritos de genes *Hox* identificadas no presente estudo em *M. olfersii*. A caixa de cor cinza indica a informação de transcritos de 56 genes *Hox* em crustáceos decápodes identificados a partir do algoritmo tBlastn utilizando ao banco de dados de sequências *GenBank*. A caixa de cor branca indica as sequências conhecidas em outras espécies de crustáceos. Fonte: Adaptado de Kenny et al. (2014).....85

**Figura 36. Eletroforese em gel de agarose a 1,5% dos produtos de amplificação dos genes de referência.** A) Amplificação por PCR para cada gene a partir de amostra de DNA de embriões (E4) de *M. olfersii*. B e C) Bandas únicas dos produtos de amplificação dos genes que apresenta o tamanho esperado e nenhuma formação de dímero. B) Em estágios de desenvolvimento. C) Em tecidos de animais adultos. D) Curvas de dissociação obtidas das reações de qPCR para cada gene com picos únicos indicando a especificidade das reações. CG: gânglio cerebral; M: músculo; HPT: hepatopâncreas, Mk 100 pb DNA Ladder.....100

**Figura 37. Distribuição do valor de Cq dos genes de referência.** Boxplots dos valores de Cq em seis estágios de desenvolvimento embrionário (A) e três tecidos de animais adultos (B) para cada um dos seis genes de referência são indicados.....102

**Figura 38. Variação na expressão de genes de referência.** Os valores de Cq de cada estágio de desenvolvimento embrionário (A) e em cada tecido de animal adulto (B) de *M. olfersii* para os genes de referencia indicam diferente expressão destes genes.....103

**Figura 39. Método de  $\Delta$ Ct para a seleção de genes de referência.** A) Variabilidade de  $\Delta$ Ct de genes de referência é apresentada como medianas (linhas), percentil 25 ao percentil 75 (caixas) e faixas (bigodes) para amostras de *M. olfersii*. Comparação de seis genes de referência em 24 amostras de estágios de desenvolvimento e 9 amostras de tecido de animais adultos.....105

**Figura 40. Análise de geNorm para a seleção de genes de referência.** Comparação de seis genes de referência em 24 amostras de estágios de desenvolvimento e 9 amostras de tecido de animais adultos. O valor médio de estabilidade de expressão (M) está indicado para cada gene de referência.....106

**Figura 41. Estabilidade geral da expressão de genes candidatos de referência calculado pela RefFinder e com base na análise de um ou dois tecidos de *M. olfersii*.** Genes de referência foram plotados de mais

estável (à esquerda) ao menos estável (à direita). CG-M: Gânglio cerebral e músculo, HPT-M: Hepatopâncreas e músculo, CG-HPT: Gânglio cerebral e hepatopâncreas.....108

**Figura 42. Número ótimo de genes para normalização determinados pelo programa geNorm.** Variação par a par entre dois fatores de normalização subsequentes (NF<sub>n</sub> e NF<sub>n+1</sub>) foi usada para determinar o número ótimo de genes de referência (valor de  $M < 0,15$ ).  $V_n/V_{n+1}$ : Variação média de pares, calculado entre  $n$  e  $n + 1$  genes de referência.....109

**Figura 43. Níveis relativos de transcritos do gene *Dll* nos estágios de desenvolvimento embrionário de *M. olfersii*, utilizando os genes de referência ou suas melhores combinações como normalizadores na análise de RT-qPCR.** Variabilidade nos níveis de transcritos do gene *Dll* é mostrado para cada referência gene testado usado como um normalizador. As barras indicam média  $\pm$  desvio padrão ( $n = 3$  pool de ovos de fêmeas ovígeras / estágio embrionário). (\*) Representa diferença significativa de  $p < 0,05$  entre estágios embrionários.....110

**Figura 44. Morfologia externa da ordem Decapoda.** A) Esquema da morfologia externa de *M. olfersii* adulto mostrando os tagmas, segmentos corporais, principais apêndices cefalotorácicos (cabeça e tórax) e abdominais. Pe: pereiópodo, Pl: pleópodo. Fonte: Adaptado de Bond-Buckup et al. (1989). B) Boca de lagosta indicando os apêndices antenas, mandíbula, maxila, maxilípedes e pereiópodo. Fonte: de <http://www.geocities.ws/CapeCanaveral/Launchpad/4289/diagrams.htm>. .....119

**Figura 45. Perfil temporal de transcritos de genes relacionados ao desenvolvimento de *M. olfersii* por RT-PCR semi-quantitativa.** A) Eletroforese em gel de agarose (1,5%) dos produtos de RT-PCR dos genes relacionados ao desenvolvimento nos estágios embrionários de E3-E10. B) Níveis de transcritos dos genes, relativo ao gene *Rpl8* durante os estágios de E3-E10 em *M. olfersii*. Cada retângulo da matriz indica o valor de expressão do gene em um determinado estágio do desenvolvimento (pools de 3 cDNA/estágio embrionário). A magnitude dos níveis de transcrito é indicada pela escala da cor azul (fraco a forte) presente na parte superior da figura. A figura foi construída através do programa *Multiple Array Viewer* 4.8. Pré-N: Pré-Nauplio, N: Nauplio, Pós-N: Pós-Nauplio.....125

**Figura 46. Níveis de transcritos dos genes *Lab*, *Dfd* e *Ftz* nos estágios de E3 a E10 de *M. olfersii* avaliada pela RT-qPCR.** Os níveis de transcritos dos genes foram normalizados com a média geométrica de *Rpl8*, *Rps6* e *Ak*. (\*\*) indica a maior diferença significativa ( $P < 0,05$ )

entre estágios consecutivos. (\*) indica a diferença significativa ( $p < 0,05$ ) entre o estágio E3 e E10. As barras indicam média  $\pm$  desvio padrão de três experimentos independentes.....127

**Figura 47. Níveis de transcritos dos genes *Pb*, *Scr* e *Antp* nos estágios de E3 a E10 de *M. olfersii* avaliada pela RT-qPCR.** Os níveis de transcritos dos genes foram normalizados com a média geométrica de 3 genes de referência (*Rpl8*, *Rps6* e *Ak*). (\*\*) indica a maior diferença significativa ( $p < 0,05$ ) entre estágios consecutivos. (\*) indica a diferença significativa ( $p < 0,05$ ) entre o estágio E3 e E10. As barras indicam média  $\pm$  desvio padrão de três experimentos independentes.....128

**Figura 48. Níveis de transcritos dos genes *Ubx*, *AbdA* e *AbdB* nos estágios de E3 a E10 de *M. olfersii* avaliada pela RT-qPCR.** Os níveis de transcritos dos genes foram normalizados com a média geométrica de 3 genes de referência (*Rpl8*, *Rps6* e *Ak*). (\*\*) indica a maior diferença significativa ( $P < 0,05$ ) entre estágios consecutivos. (\*) indica a diferença significativa ( $p < 0,05$ ) entre o estágio E3 e E10. As barras indicam média  $\pm$  desvio padrão de três experimentos independentes.....129

**Figura 49. Níveis de transcritos dos genes *Wg*, *Hh* e *Dll* nos estágios de E3 a E10 de *M. olfersii* avaliada pela RT-qPCR.** Os níveis de transcritos dos genes foram normalizados com a média geométrica de 3 genes de referência (*Rpl8*, *Rps6* e *Ak*). (\*\*) indica a maior variação ( $p < 0,05$ ) entre estágios consecutivos. As barras indicam média  $\pm$  desvio padrão de três experimentos independentes.....130

**Figura 50. Detecção de transcritos de genes relacionados ao desenvolvimento em ovário, estágios iniciais e finais do desenvolvimento e larvas de *M. olfersii* avaliado pela RT-PCR.** Eletroforese de gel de agarose (1,5%) dos produtos de RT-PCR dos transcritos dos genes relacionados ao desenvolvimento nos estágios embrionários de E3-E10. O gene *Rpl8* foi utilizado como controle positivo da reação de síntese de cDNA das amostras. As ilustrações das fêmeas foram adaptadas de Nazari et al. (2003). As ilustrações das larvas foram adaptadas de Dugger e Dobkin (1975).....132

**Figura 51. Análise dos transcritos dos genes relacionados ao desenvolvimento embrionário de *M. olfersii* nos estágios de E3 a E10.** A) Modelo hipotético da localização espacial dos transcritos dos genes *Hox* em embriões de E3-E10 de *M. olfersii* e sua relação com a identidade dos segmentos e apêndices. Dia embrionário (E). Antêntula (An), antena (At), papila caudal (Pc), disco germinativo (Dg), mandíbula (Mn), maxílula (Mxl), maxila (Mx), maxilípede (Mxp), pereiópodos (Per), lobo óptico (Lo), abdome (Abd). B) Tendência do

perfil de transcritos de genes *Hox* e genes de segmentação durante o desenvolvimento embrionário de *M. olfersii*.....137



## LISTA DE TABELAS

<b>Tabela 1.</b> Parâmetros dos <i>contigs</i> obtidos utilizando Velvet/Oases com múltiplos k-mers para a montagem do transcriptoma de <i>M. olfersii</i> .....	32
<b>Tabela 2.</b> Resumo de dados de RNA-Seq para a montagem <i>de novo</i> do transcriptoma de <i>M. olfersii</i> utilizando Velvet/Oases com múltiplos k-mers.....	32
<b>Tabela 3.</b> Os 60 domínios/famílias de proteínas conservados do InterPro mais abundantes identificadas nos unigenes de <i>M. olfersii</i> . .....	35
<b>Tabela 4.</b> Similaridade dos transcritos de tRNA do genoma mitocondrial de <i>M. olfersii</i> com as sequências do banco de dados <i>GenBank</i> .....	39
<b>Tabela 5.</b> Similaridade dos transcritos de genes do genoma mitocondrial de <i>M. olfersii</i> com as sequências do banco de dados <i>GenBank</i> .....	40
<b>Tabela 6.</b> Iniciadores degenerados utilizados na RT-PCR para a identificação dos transcritos dos genes relacionados ao desenvolvimento em <i>M. olfersii</i> .....	52
<b>Tabela 7.</b> Similaridade das sequências de aminoácidos traduzidas destes transcritos de genes relacionados ao desenvolvimento de <i>M. olfersii</i> com as sequências do banco de dados <i>GenBank</i> .....	57
<b>Tabela 8.</b> Transcritos de genes relacionados ao desenvolvimento embrionário e vias de desenvolvimento identificados no transcriptoma de <i>M. olfersii</i> .....	59
<b>Tabela 9.</b> Similaridade dos genes de referência de <i>M. olfersii</i> com as sequências disponíveis no banco de dados <i>GenBank</i> .....	98
<b>Tabela 10.</b> Candidatos a genes de referência e sequências de iniciadores para RT-qPCR usadas neste estudo.....	99
<b>Tabela 11.</b> Genes de referência e seus parâmetros derivados da análise de qPCR para determinar a eficiência de amplificação.....	100
<b>Tabela 12.</b> Estatística descritiva dos valores de Cq dos genes de referência nos estágios de desenvolvimento embrionário e tecidos de animais adultos de <i>M. olfersii</i> .....	102
<b>Tabela 13.</b> Valores de estabilidade da expressão de genes de referência no desenvolvimento embrionário e em tecidos de animais adultos com base em vários programas.....	107
<b>Tabela 14.</b> Sequências dos iniciadores dos genes relacionados ao desenvolvimento embrionário de <i>M. olfersii</i> utilizados na RT-PCR semi-quantitativa.....	122
<b>Tabela 15.</b> Sequências dos iniciadores dos genes relacionados ao desenvolvimento embrionário de <i>M. olfersii</i> utilizados na RT-qPCR.....	124



## LISTA DE ABREVIATURAS E SIGLAS

°C	Graus Celsius, unidade de medida de temperatura
$\Delta$ Ct	Diferença entre valores de Cq
A1-A6	Primeiro até sexto segmento abdominal
An	Antenula
A-P	Ântero-posterior
At	Antena
cDNA	DNA complementar a RNA (do inglês <i>Complementary DNA</i> )
CG-M	Gânglio cerebral e músculo
CG-HPT	Gânglio cerebral e hepatopâncreas
CRISPR/Cas9	do inglês <i>Clustered Regularly Interspaced Short Palindromic Repeats/Caspase 9</i>
Cq	Ciclo de quantificação
CV	Coeficiente de variação
DEPC	Dietilpirocarbonato
DNA	Ácido desoxirribonucleico (do inglês <i>Deoxyribonucleic Acid</i> )
dNTP	Desoxinucleotídeo tri-fosfatado
D-V	Dorso-ventral
E	Estágio embrionário
<i>E</i>	Eficiência de amplificação
EST	Etiqueta de sequência expressa (do inglês <i>expressed sequence tag</i> )
g	Grama, unidade de medida de massa
<i>g</i>	Força de gravidade
GC	Guanina - citosina
GO	Ontologia gênica
HPT-M	Hepatopâncreas e músculo
Int	Intercalar
JAK/STAT	Do inglês <i>Janus kinase /signal transducer and activator of transcription</i>
KEGG	Do inglês <i>Kyoto Encyclopedia of Genes and Genomes</i>
k-mer	Subsequência gerada por decomposição dos <i>reads</i> , o qual é utilizada para procurar sobreposição com outros <i>reads</i>
KV	Kilovolt
L	Litro, unidade de medida de volume
Lb	Lábio

M	Molar, unidade de medida de concentração
mDNA	DNA mitocondrial
min	Minutos, unidade de medida de tempo
mL	Mililitro, unidade de medida de volume
mm	Milímetro, unidade de medida de comprimento
mM	Milimolar, unidade de medida de concentração
Mn	Mandíbula
mRNA	RNA mensageiro
Mxl	Maxílula
Mx	Maxila
Mk	Marcador de peso molecular de DNA
NF	Fatores de normalização (do inglês <i>normalization factor</i> )
NR	Não redundantes
nt	Nucleotídeos
Oc	Ocular
ORF	Quadro aberto de leitura (do inglês <i>Open Reading Frame</i> )
Ova-E1	Ovário de fêmeas com embriões em E1
Ova-IM	Ovário imaturo
Ova-M	Ovário maduro
pb	Pares de bases
pmol	Picomol, unidade de massa
PCR	Reação em Cadeia da Polimerase (do inglês <i>Polymerase Chain Reaction</i> )
qPCR	PCR tempo real (do inglês <i>Quantitative Real Time Polymerase Chain Reaction</i> )
RNA	Ácido ribonucleico (do inglês <i>Ribonucleic Acid</i> )
RNA-Seq	Sequenciamento de RNA
RT-PCR	Transcriptase reversa - PCR (do inglês <i>Reverse transcriptase-Polymerase Chain Reaction</i> )
RT-qPCR	Transcriptase reversa - PCR tempo real (do inglês <i>Reverse transcriptase-Quantitative Real Time Polymerase Chain Reaction</i> )
R <sup>2</sup>	Coefficiente de correlação
rRNA	RNA ribossomal
SD	Desvio padrão
seg	Segundos, unidade de medida de tempo
SRA	Arquivo de sequências de <i>reads</i> (do inglês <i>sequence read archive</i> )
Ta	Temperatura de anelamento

TGF- $\beta$	Do inglês <i>Transforming growth factor-beta</i>
$T_m$	Temperatura de dissociação (do inglês <i>melting temperature</i> )
tRNA	RNA transportador
Te	Téilson
T1-T8	Primeiro até oitavo segmento torácico
U	Unidade enzimática
$\mu\text{g}$	Micrograma, unidade de medida de massa
$\mu\text{L}$	Microlitro, unidade de medida de volume
$\mu\text{M}$	Micromolar, unidade de medida de concentração
UTR	Regiões não traduzíveis (do inglês <i>untranslated region</i> )
UV	Ultravioleta
$V_n/V_{n+1}$	Varição média de pares, calculado entre $n$ e $n + 1$ genes de referência
Z	Larva zoea



## SUMÁRIO

<b>INTRODUÇÃO</b> .....	1
1.1. DIVERSIDADE MORFOLÓGICA EM ARTRÓPODES.....	1
1.2. GENES DE DESENVOLVIMENTO EMBRIONÁRIO EM ARTRÓPODES.....	3
1.2.1. Genes de segmentação.....	7
1.2.2. Genes <i>Hox</i> .....	10
1.3. <i>Macrobrachium olfersii</i> COMO MODELO DE ESTUDO DE DESENVOLVIMENTO DENTRO DA FAMÍLIA PALAEMONIDAE.....	12
1.4. DESENVOLVIMENTO EMBRIONÁRIO DE <i>M. olfersii</i> .....	14
1.5. JUSTIFICATIVA.....	18
<b>2. OBJETIVOS</b> .....	21
2.1. OBJETIVO GERAL.....	21
2.2. OBJETIVOS ESPECÍFICOS.....	21
<b>3. ORGANIZAÇÃO DA TESE</b> .....	23
<b>CAPITULO I. Montagem <i>de novo</i> e análise do transcriptoma de embriões de <i>Macrobrachium olfersii</i></b> .....	25
1. INTRODUÇÃO.....	25
2. MATERIAIS E MÉTODOS.....	27
2.1. OBTENÇÃO DE ANIMAIS E EMBRIÕES DE <i>M. olfersii</i> .....	27
2.2. EXTRAÇÃO DE RNA TOTAL.....	28
2.3. CONSTRUÇÃO DE BIBLIOTECA DE RNA-Seq E SEQUENCIAMENTO.....	29
2.4. MONTAGEM <i>de novo</i> DO TRANSCRIPTOMA DE EMBRIÕES DE <i>M. olfersii</i> .....	29
2.5. ANOTAÇÃO DE TRANSCRITOS DE GENES, DOMÍNIOS/FAMÍLIAS DE PROTEÍNAS E CLASSIFICAÇÃO GO/KEGG.....	30
2.6. IDENTIFICAÇÃO DE TRANSCRITOS DE GENES DO GENOMA MITOCONDRIAL.....	31
3. RESULTADOS.....	31
3.1. SEQUENCIAMENTO NA PLATAFORMA ILLUMINA E MONTAGEM DE SEQUÊNCIAS.....	31
3.2. ANOTAÇÃO DE UNIGENES DO TRANSCRIPTOMA.....	33
3.3. CLASSIFICAÇÃO DE UNIGENES POR ONTOLOGIA GÊNICA (GO).....	36
3.4. CLASSIFICAÇÃO DE UNIGENES EM VIAS BIOLÓGICAS DO KEGG.....	37
3.5. IDENTIFICAÇÃO DE TRANSCRITOS DE GENES DO GENOMA MITOCONDRIAL.....	38
4. DISCUSSÃO.....	41

<b>CAPÍTULO II. Identificação de transcritos de genes e vias relacionados ao desenvolvimento embrionário de <i>M. olfersii</i></b> .....	47
1. INTRODUÇÃO.....	47
2. MATERIAIS E MÉTODOS.....	49
2.1. IDENTIFICAÇÃO DE TRANSCRITOS DE GENES RELACIONADOS AO DESENVOLVIMENTO EMBRIONÁRIO POR ANÁLISE TRANSCRIPTÔMICA.....	49
2.2. IDENTIFICAÇÃO DE TRANSCRITOS DE GENES RELACIONADOS AO DESENVOLVIMENTO BASEADA EM RT-PCR.....	50
2.2.1. Embriões de <i>M. olfersii</i> e extração de RNA.....	50
2.2.2. Síntese de cDNA.....	51
2.2.3. Desenho de iniciadores degenerados.....	51
2.2.4. Amplificação pela PCR e purificação dos produtos.....	51
2.2.5. Clonagem dos produtos de RT-PCR e sequenciamento.....	52
2.3. ANÁLISE BIOINFORMÁTICA E FILOGENIA DE GENES RELACIONADOS AO DESENVOLVIMENTO EM <i>M. olfersii</i> .....	53
3. RESULTADOS.....	<b>Erro</b>
<b>! Indicador não definido.</b>	
3.1. IDENTIFICAÇÃO DOS TRANSCRITOS DE GENES RELACIONADOS AO DESENVOLVIMENTO EM <i>M. olfersii</i> ATRAVÉS DE ESTRATÉGIA BASEADO EM RT-PCR.....	<b>Erro! Indicador não definido.</b>
3.2. IDENTIFICAÇÃO DOS TRANSCRITOS DOS GENES RELACIONADOS AO DESENVOLVIMENTO E VIAS DE DESENVOLVIMENTO NO TRANSCRIPTOMA.....	<b>Erro! Indicador não definido.</b>
3.3. SIMILARIDADE DOS TRANSCRITOS DOS GENES RELACIONADOS AO DESENVOLVIMENTO DE <i>M. olfersii</i> COM OUTRAS ESPÉCIES.....	<b>Erro! Indicador não definido.</b>
3.4. ANÁLISE BIOINFORMÁTICA E FILOGENIA DOS GENES RELACIONADOS AO DESENVOLVIMENTO.....	<b>Erro! Indicador não definido.</b>
3.4.1. Genes <i>Hox</i> .....	<b>Erro!</b>
<b>Indicador não definido.</b>	
3.4.2. Genes de segmentação.....	<b>Erro! Indicador não definido.</b>
3.4.3. Genes da família <i>Wnt</i> .....	<b>Erro! Indicador não definido.</b>



3.5. IDENTIFICAÇÃO DE GENES *Hox* EM CRUSTÁCEOS.....**Erro! Indicador não definido.**

4. DISCUSSÃO.....**Erro! Indicador não definido.**

**CAPITULO III. Identificação e avaliação de genes de referência para análises por RT-qPCR no desenvolvimento embrionário de *M. olfersii*.....E rro! Indicador não definido.**

1. INTRODUÇÃO.....**Erro! Indicador não definido.**

2. MATERIAIS E MÉTODOS.....**Erro! Indicador não definido.**

2.1. ANIMAIS, TECIDOS E EMBRIÕES DE *M. olfersii*.....**Erro! Indicador não definido.**

2.2. EXTRAÇÃO DE RNA E SÍNTESE DE cDNA.....**Erro! Indicador não definido.**

2.3. IDENTIFICAÇÃO DE GENES DE REFERÊNCIA E DESENHO DE INICIADORES.....**E rro! Indicador não definido.**

2.4. EXTRAÇÃO DE DNA E DETERMINAÇÃO DE JUNÇÕES EXON-EXON.....**Er ro! Indicador não definido.**

2.5. PCR QUANTITATIVA (qPCR).....**Erro! Indicador não definido.**

2.6. EFICIÊNCIA DE AMPLIFICAÇÃO E ESPECIFICIDADE DOS INICIADORES DE GENES DE REFERÊNCIA.....**Erro! Indicador não definido.**

2.7. DETERMINAÇÃO DA ESTABILIDADE DE GENES DE REFERÊNCIA.....**Er ro! Indicador não definido.**

2.8. INFLUÊNCIA DA NORMALIZAÇÃO COM GENES DE REFERÊNCIA.....**Er ro! Indicador não definido.**

2.9. ANÁLISE ESTATÍSTICA.....**Erro! Indicador não definido.**

3. RESULTADOS.....**Erro! Indicador não definido.**

3.1. IDENTIFICAÇÃO DE CANDIDATOS A GENES DE REFERÊNCIA...**Erro! Indicador não definido.**

3.2. DETERMINAÇÃO DA JUNÇÃO EXON-EXON, ESPECIFICIDADE DOS INICIADORES E A EFICIÊNCIA DE AMPLIFICAÇÃO.....	<b>Erro! Indicador não definido.</b>
3.3. NÍVEIS DE TRANSCRITOS DE GENES DE REFERÊNCIA EM ESTÁGIOS EMBRIONÁRIOS E TECIDOS ADULTOS DE <i>M. olfersii</i> .....	<b>Erro! Indicador não definido.</b>
3.4. EXPRESSÃO DA ESTABILIDADE DOS CANDIDATOS A GENES DE REFERÊNCIA.....	<b>Erro! Indicador não definido.</b>
3.5. CLASSIFICAÇÃO GERAL DA EXPRESSÃO DE ESTABILIDADE.....	<b>Erro! Indicador não definido.</b>
3.6. NÚMERO IDEAL DE GENES DE REFERÊNCIA PARA A NORMALIZAÇÃO.....	<b>Erro! Indicador não definido.</b>
3.7. INFLUÊNCIA DE GENES DE REFERÊNCIA NA ANÁLISE DE DADOS DE RT-qPCR.....	<b>Erro! Indicador não definido.</b>
4. DISCUSSÃO.....	<b>Erro! Indicador não definido.</b>

#### **CAPÍTULO IV. Análise temporal dos transcritos dos genes *Hox* e de segmentação durante o desenvolvimento embrionário de *M. olfersii*.....**

1. INTRODUÇÃO.....	<b>Erro! Indicador não definido.</b>
2. MATERIAIS E MÉTODOS.....	<b>Erro! Indicador não definido.</b>
2.1. EMBRIÕES E EXTRAÇÃO DE RNA.....	<b>Erro! Indicador não definido.</b>
2.2. SÍNTESE DE cDNA E DESENHO DE INICIADORES.....	<b>Erro! Indicador não definido.</b>
2.3. RT-PCR SEMI-QUANTITATIVA EM ESTÁGIOS (E3-E10) DE <i>M. olfersii</i> .....	<b>Erro! Indicador não definido.</b>
2.4. DETECÇÃO DE TRANSCRITOS DE GENES RELACIONADOS AO DESENVOLVIMENTO EM OVÁRIO, ESTÁGIOS INICIAIS E FINAIS DO DESENVOLVIMENTO E LARVAS DE <i>M. olfersii</i> .....	<b>Erro! Indicador não definido.</b>

2.5. NÍVEIS DE TRANSCRITOS DE GENES RELACIONADOS AO DESENVOLVIMENTO POR RT-qPCR NOS ESTÁGIOS E3-E10.....	<b>Erro! Indicador não definido.</b>
2.6. ANÁLISE ESTATÍSTICA.....	<b>Erro! Indicador não definido.</b>
3. RESULTADOS.....	<b>Erro! Indicador não definido.</b>
3.1. PERFIL DE TRANSCRITOS DE GENES RELACIONADOS AO DESENVOLVIMENTO PELA RT-PCR SEMI-QUANTITATIVA.....	<b>Erro! Indicador não definido.</b>
3.2. NÍVEIS DE TRANSCRITOS DOS GENES RELACIONADOS AO DESENVOLVIMENTO POR RT-qPCR.....	<b>Erro! Indicador não definido.</b>
3.3. TRANSCRITOS DOS GENES RELACIONADOS AO DESENVOLVIMENTO EM OVÁRIO, ESTÁGIOS INICIAIS E FINAIS DO DESENVOLVIMENTO E LARVAS DE <i>M. olfersii</i> .....	<b>Erro! Indicador não definido.</b>
4. DISCUSSÃO.....	<b>Erro! Indicador não definido.</b>
<b>CAPÍTULO V: DISCUSSÃO GERAL</b>	
<b>CONCLUSÕES GERAIS</b>	
<b>GERAIS</b> .....	<b>Erro! Indicador não definido.</b>
<b>PERSPECTIVAS</b> .....	<b>Erro! Indicador não definido.</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS</b>	
<b>BIBLIOGRÁFICAS</b> .....	<b>Erro! Indicador não definido.</b>
<b>APÊNDICE</b>	
<b>APÊNDICE</b> .....	<b>Erro! Indicador não definido.</b>



## INTRODUÇÃO

---

### 1.1. DIVERSIDADE MORFOLÓGICA EM ARTRÓPODES

Os artrópodes são um grupo de invertebrados bem sucedido por apresentarem corpo segmentado, apêndices articulados e exoesqueleto quitinoso, características que permitem a este grupo uma maior flexibilidade funcional e, conseqüentemente, diversidade morfológica. Devido a essa diversidade morfológica, por mais de 150 anos, há um debate sobre quais são os animais mais próximos aos artrópodes e sobre as relações filogenéticas dentro este filo (BUDD; TELFORD, 2009). Tradicionalmente, tem sido considerado que há estreita relação entre os anelídeos e os artrópodes, os dois grandes filos que apresentam corpo segmentado (JOHNSON; RAVEN, 2002). No entanto, outros estudos sugerem que os artrópodes e anelídeos não compartilham um grupo irmão por isso, a segmentação corporal poderia ter surgido por convergência adaptativa (EERNISSE, 1998). Recentes estudos consideram o filo Onychophora (vermes-aveludados) como um grupo irmão dos artrópodes. O filo Onychophora apresenta um plano de corpo simples com apenas a cabeça e segmentos mais posteriores diferenciados e sem membros articulados (JANSSEN et al., 2014).

Apesar das relações filogenéticas complexas dos artrópodes ainda não estarem muito clara, a diversidade de artrópodes é surpreendente, com mais de 1,2 milhões de espécies existentes, incluindo pouco mais de 1 milhão de hexápodes, quase 112.000 de quelicerados, cerca de 67.000 crustáceos e 12.000 miriápodes (MINELLI; FUSCO; BOXSHALL, 2013). Os artrópodes apresentam tamanho bastante variável, desde indivíduos microscópicos, como alguns ácaros (menos de 1 mm) até enormes, como o caranguejo aranha japonês (3,8 m). Todos os artrópodes têm apêndices articulados, sendo observado número reduzido de apêndices em espécies mais derivadas evolutivamente. Nesse filo, os apêndices podem ser modificados em antenas para detectar o ambiente, em peças bucais para a alimentação e em apêndices locomotores para o deslocamento (JOHNSON; RAVEN, 2002).

Entre os artrópodes, o subfilo Crustacea é o que exhibe a maior diversidade morfológica (MARTIN; DAVIS, 2001), sendo esta diversidade, aliada à diversidade de planos corporais, o que torna o estudo de crustáceos tão interessante. No entanto, esta diversidade também dificulta a classificação e a filogenia das espécies. O subfilo

Crustacea é descrito como presente desde o início do Cambriano, apresentando um processo evolutivo divergente que se reflete na diversidade de formas corporais descritas (MARTIN; DAVIS, 2001). Atualmente, estima-se a existência de aproximadamente 67.000 espécies de crustáceos (MINELLI; FUSCO; BOXSHALL, 2013), colocando-os na quarta posição em termos de diversidade geral de espécies, atrás apenas dos insetos, moluscos e quelicerados (MARTIN; DAVIS, 2001).

A diversidade morfológica dos crustáceos ainda provoca discussões sobre sua origem filogenética, em especial sobre a possibilidade de serem um grupo monofilético (MARTIN; DAVIS, 2001). A introdução de ferramentas moleculares nessa discussão propiciou uma nova perspectiva, onde os insetos estão mais relacionados filogeneticamente aos crustáceos, formando o clado Pancrustacea (REGIER; SHULTZ; KAMBIC, 2005; REGIER et al., 2010; ROTA-STABELLI et al., 2011). Nesse contexto, os crustáceos podem ser utilizados como um grupo importante em análises filogenéticas dos insetos e dos diferentes padrões de desenvolvimento (REHM et al., 2009a).

Os crustáceos estão divididos em seis classes: Branchiopoda, Remipedia, Cephalocarida, Maxillopoda, Ostracoda e Malacostraca. Dentre eles, a classe Malacostraca possui o maior número de espécies conhecidas, constituindo um grupo heterogêneo que habita tanto ambientes marinhos, quanto dulcícolas (RÓLIER-LARA et al., 2013). A ordem Decapoda está dividida na subordem Dendrobranchiata (caracterizado por espécies cujas fêmeas não incubam os ovos) e na subordem Pleocyemata (caracterizado por espécies cujas fêmeas incubam seus ovos na região ventral do corpo). Dentre os Pleocyemata, destacam-se os camarões de água doce da família Palaemonidae, que estão distribuídos por todos os continentes, nas regiões tropicais e temperadas, com seus representantes habitando corpos de água doce ou salobra (HOLTHUIS, 1952).

Esta diversidade morfológica está relacionada aos diferentes hábitos de vida apresentados, assim como na variedade de modelos de desenvolvimento embrionário reconhecidos (MINELLI; FUSCO; BOXSHALL, 2013). Neste sentido, essa diversidade estaria relacionada com a expressão de genes envolvidos nos mecanismos moleculares, que coordenam os padrões de organização do plano corporal durante o desenvolvimento embrionário. Portanto, o estudo temporal e espacial dos transcritos de genes responsáveis pelas mudanças morfológicas (por exemplo, os genes *Hox*), durante o desenvolvimento embrionário das

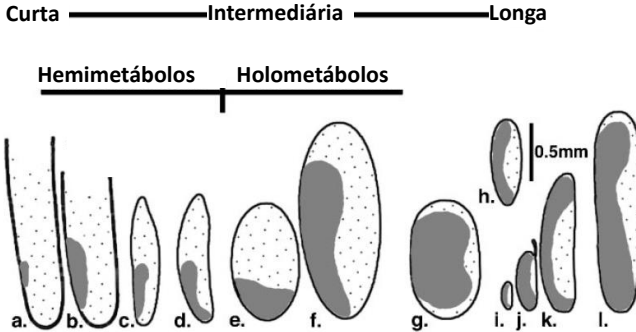
espécies, tem sido um tema de interesse dentre os estudos de artrópodes e, especialmente, em crustáceos (MARTIN; DAVIS, 2001).

## 1.2. GENES DE DESENVOLVIMENTO EMBRIONÁRIO EM ARTRÓPODES

Uma das características marcantes do desenvolvimento embrionário de artrópodes é a presença de um disco germinativo, uma área posicionada na superfície do ovo e formada por células do blastoderma, que constitui o início do processo de diferenciação das células do embrião. O crescimento longitudinal do disco germinativo resulta na formação da banda germinativa (Figura 1), formada por células das linhagens ecto e mesoteloblastos. Essa banda se estende ao longo do eixo longitudinal do embrião e delimita a futura superfície ventral do corpo do embrião (SCHOLTZ; WOLFF, 2013).

Entre os artrópodes, os insetos têm sido mais estudados quanto ao seu desenvolvimento embrionário. Isso se deve em parte à facilidade de reconhecimento da formação dos segmentos corporais durante o desenvolvimento. Esses segmentos são descritos na embriologia clássica como unidades corporais repetitivas ao longo do eixo ântero-posterior, possuindo uma correlação espaço-temporal específica para cada modelo de desenvolvimento (SCHOLTZ, 2002). São reconhecidos 3 modelos básicos para a embriogênese inicial, os quais são relacionados ao tipo de banda germinativa presente, sendo descritos como embriões de banda curta, banda intermediária ou banda longa (do inglês, *short, intermediate, and long germ embryogenesis*) (Figura 1). Essa terminologia foi originalmente estabelecida por Krause em 1939 (DAVIS; PATEL, 2002).

Em insetos com embriões de banda longa, como a *Drosophila melanogaster*, todos os segmentos do corpo são formados simultaneamente durante o estágio de blastoderma (LIU; KAUFMAN, 2005; LYNCH; EL-SHERIF; BROWN, 2012). Enquanto que, nos insetos com embriões de banda curta ou intermediária, apenas os segmentos anteriores são especificados antes da gastrulação, e os segmentos posteriores são modelados progressivamente, a partir de uma região indiferenciada posterior, chamada zona de crescimento (LYNCH; EL-SHERIF; BROWN, 2012). A zona de crescimento tem sido definida como uma região subterminal na parte posterior da banda germinativa (MARTIN; KIMELMAN, 2009).

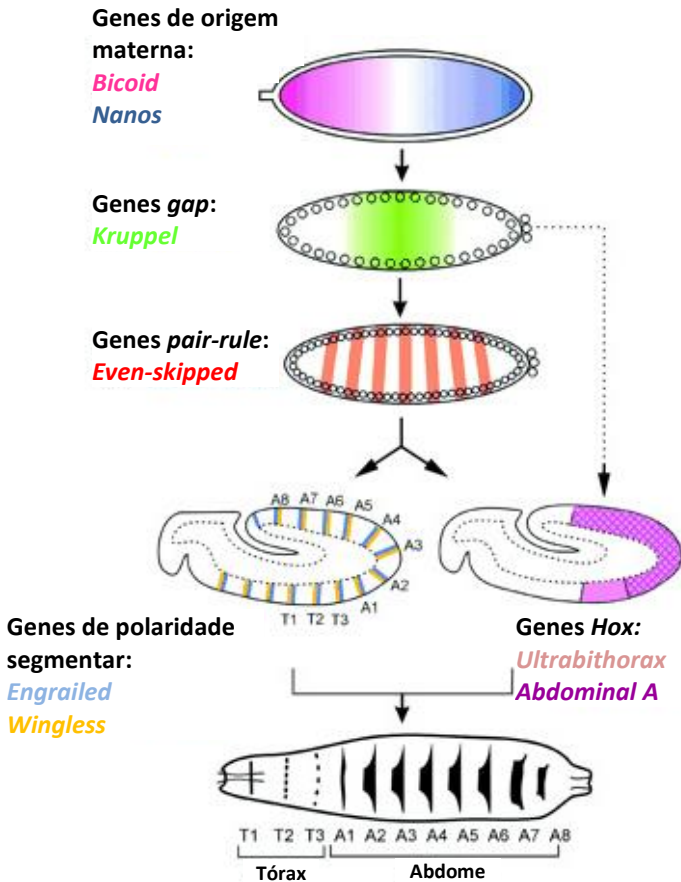


**Figura 1. Tipos de banda germinativa observadas em embriões de insetos.**

Os ovos são esquematizados em escala e em vista lateral, com a região anterior na parte superior. A área do embrião rudimentar é sombreada. Orthoptera: a, *Oecanthus pellucens*; b, *Acheta domesticus*. Odonata: c, *Platycnemis pennipes*. Hemiptera: d, *Euscelis plebejus*. Coleoptera: e, *Atrachya menetriesi*; f, *Leptinotarsa decemlineata*; h, *Bruchidius obtectus*. Lepidoptera: g, *Bombyx mori*. Diptera: i, *Smittia* sp.; j, *D. melanogaster*; k, *Calliphora erythrocephala*. Hymenoptera: l, *Apis mellifera*. Fonte: Adaptado de Davis e Patel (2002).

Os mecanismos moleculares envolvidos na padronização corporal durante o desenvolvimento embrionários dos artrópodes têm sido descritos a partir dos estudos de expressão, com embriões de *D. melanogaster*. Sendo possível verificar a presença de uma hierarquia em genes de contribuição materna, genes *gap*, genes *pair-rule* e genes de polaridade segmentar, ocorrendo em cascata durante o desenvolvimento embrionário (Figura 2). Nesse modelo, os RNA mensageiros maternos geram centros de sinalização ao longo do eixo Ântero-Posterior (A-P), estabelecendo os gradientes de morfógenos. Esses gradientes instruem os domínios de expressão dos genes precoces zigóticos, chamados genes *gap*. Isso é possível, em parte, pelo fato do blastoderma sincicial do embrião ser constituído por blastômeros não delimitados por membranas, permitindo a difusão de fatores de transcrição de morfógenos por um citoplasma compartilhado sem sinalização célula-célula. Nesse ambiente, a ampla ativação por fatores maternos, juntamente com as atividades de repressão pelos genes *gap* leva à expressão dos genes *pair-rule* em segmentos alternados, como faixas de expressão. Os genes *pair-rule* estabelecem a polaridade nos segmento através da ativação dos genes segmentares, cada um expresso em faixas ao longo dos segmentos (ROSENBERG et al., 2014).





**Figura 2. Padronização ao longo do eixo Ântero-Posterior (A/P) do embrião de *Drosophila melanogaster*.** Uma cascata de genes maternos e zigóticos é ativada no embrião sincicial para subdividir o ectoderma em domínios menores. Ocorre intensa proliferação celular após a ativação dos genes *pair-rule*. Os genes de polaridade de segmentos e os genes *Hox* são ativados pelos genes *pair-rule*, mas um subconjunto de genes *gap* também influencia diretamente aos genes *Hox*. Ambos os genes de polaridade de segmento e *Hox* atuam em conjunto para controlar a diferenciação de cada segmento da futura larva (SANSON, 2001). Fonte: Adaptado de Sanson (2001).

A identificação de homólogos desses genes, relacionados ao desenvolvimento de embriões de banda longa de *D. melanogaster*, permitiu comparações do perfil espacial dos transcritos desses genes em embriões de banda curta e longa de outros insetos (LIU; KAUFMAN,

2005). Homologias aos genes de segmentação, de desenvolvimento da região cefálica e de padronização dorso-ventral também foram identificados em embriões de banda curta do inseto *Tribolium castaneum* (BROWN; DENELL, 1996), mostrando que as redes de regulação do desenvolvimento parecem funcionar de modo similar entre os insetos (BROWN; DENELL, 1996; CHOE; MILLER; BROWN, 2006). Em geral, as análises moleculares sugerem que, embora os mecanismos genéticos gerais de segmentação sejam conservados, os detalhes deste processo apresentam variações entre as diferentes ordens de insetos (DAVIS; PATEL, 2002; LIU; KAUFMAN, 2005). Por outro lado, estudos em outros artrópodes demonstram que este mecanismo pode não ser único, uma vez que diferentes grupos de artrópodes apresentam distintos modelos de segmentação e também mecanismos moleculares distintos (PEEL; CHIPMAN; AKAM, 2005).

Nos crustáceos os modelos de bandas germinativas também estão presentes. A maioria dos crustáceos apresentam embriões de banda curta, semelhante ao apresentado pelo ancestral hipotético desse grupo, sendo que a segmentação nos embriões de banda longa derivou desse modelo ancestral (PEEL; CHIPMAN; AKAM, 2005; LIU; KAUFMAN, 2005). As lagostas de água doce (*Cherax destructor* e *Procambarus* sp.) e os isópodes (como *Porcellio scaber*), apresentam embriões de banda curta a intermediária (SCHOLTZ, 1992; ALWES; SCHOLTZ, 2006; WOLFF, 2009), enquanto que *Gammarus pulex* e outros anfípodes apresentam embriões de banda intermediária a longa (SCHOLTZ, 1990). O gênero *Daphnia* apresenta embriões de banda do tipo longa (SCHOLTZ; WOLFF, 2013). O camarão de água doce *Macrobrachium olfersii* apresenta o modelo de desenvolvimento de banda curta, característico dos crustáceos (JARAMILLO et al., 2016).

Nos crustáceos, os estudos relacionados aos genes de desenvolvimento têm sido realizados em poucas espécies, como *Artemia franciscana* (AVEROF; AKAM, 1993), *Procambarus clarkii*, *P. scaber* (ABZHANOV; KAUFMAN, 1999a, 2000b) e *Parhyale hawaiensis* (LIUBICICH et al., 2009). Esses estudos demonstram a existência de homologias entre os genes descritos para crustáceos com aqueles identificados no desenvolvimento de *D. melanogaster*. Recentemente, foram identificados os genes relacionados ao desenvolvimento embrionário de camarões de água salgada *L. vannamei* e *P. monodom*, e para os camarões de água doce *N. denticulata* (LI, C et al., 2012; SELLARS et al., 2015) (KENNY et al., 2014) e *M. olfersii* (JARAMILLO et al., 2016). Entre os genes de interesse descritos, destacam-se os genes *pair-rule* (BRENA; AKAM, 2013; GREEN;

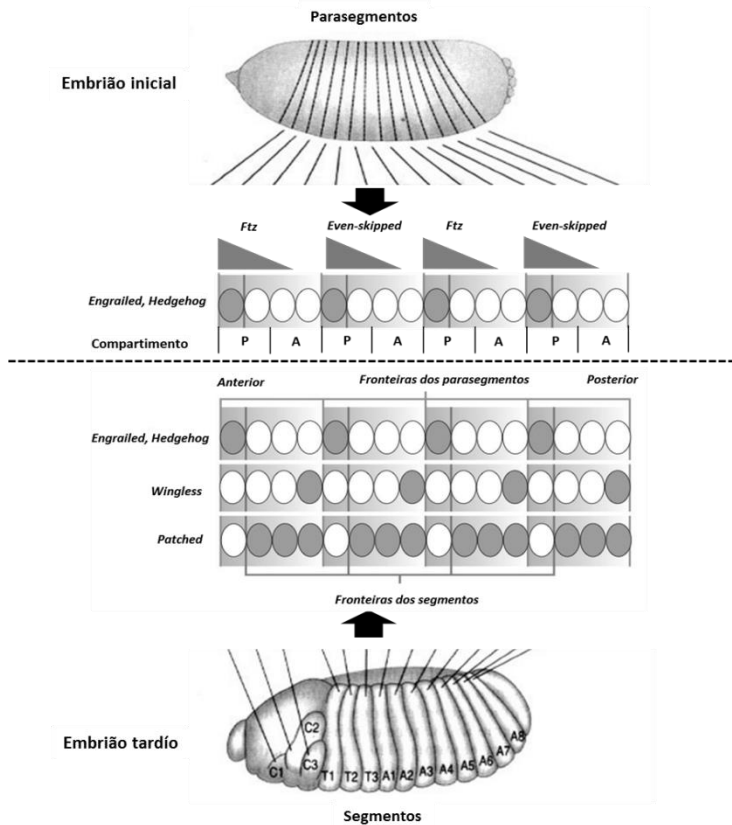
AKAM, 2013; JARAMILLO et al., 2016), os genes de polaridade segmentar, os genes *Hox* (KIM et al., 2016; SERANO et al., 2016; JARAMILLO et al., 2016) e os genes da família *Wnt* (BOLOGNESI et al., 2008; JARAMILLO et al., 2016).

### 1.2.1. Genes de segmentação

Os genes de segmentação são aqueles que especificam o número correto e a polaridade dos segmentos do corpo do embrião. Em *D. melanogaster*, eles podem ser sub-classificados em genes *gap*, genes *pair-rule* e os genes de polaridade de segmentos (VOET; VOET; PRATT, 2014). Um dos membros dos genes *pair-rule* é o gene *Even-skipped* (*Eve*) que codifica um fator de transcrição requerido para a ativação do gene *Engrailed* (*En*) durante a segmentação de *D. melanogaster* e para a correta organização dos parasegmentos ímpares (FUJIOKA et al., 2002). Os transcritos do gene *Eve* de *D. melanogaster* estão localizados em segmentos alternados (periodicidade de segmento duplo) (Figura 3), similar à maioria de genes *pair-rule*. No entanto, estudos de ortólogos do gene *Eve* em crustáceos, demonstram que eles são expressos em faixas (bandas) segmentares, mas sem um padrão claro de *pair-rule* (DAVIS; PATEL, 2003).

Os genes de polaridade de segmentos são ativados pelos produtos gênicos *pair-rule* que controlam sua expressão periódica, sendo que os mais estudados são os *En*, *Wingless* (*Wg*) e *Hedgehog* (*Hh*). Em *D. melanogaster*, a formação e manutenção das fronteiras dos parasegmentos dependem da interação entre as proteínas *wingless*, *engrailed* e *hedgehog* (SIMONNET; DEUTSCH; QUÉINNEC, 2004).

O gene *En* codifica um fator de transcrição requerido para a segmentação dos embriões e para estabelecer informação posicional dentro de cada segmento. O domínio de expressão na parte posterior de cada segmento (Figura 3) sugere uma função conservada na segmentação, como mostra os estudos com *D. melanogaster* e outros artrópodes. Dois genes parálogos *En* (*En1* e *En2*) têm sido encontrados no crustáceo *P. scaber* (Isopoda) e *P. clarkii* (Decapoda) (ABZHANOV; KAUFMAN, 2000c). A proteína *engrailed*, utilizando o anticorpo Mab4D9, em *P. scaber*, é identificada na parte posterior de cada segmento, incluindo os apêndices, o lado ventral e a parte lateral do tronco. Além disso, os estudos demonstram a participação das proteínas *even-skipped* e de *engrailed* na neurogênese de artrópodes (DUMAN-SCHEEL; PATEL, 1999). A expressão do gene *En* é considerada como um ótimo marcador molecular para determinar o processo de segmentação (PATEL; KORNBERG; GOODMAN, 1989).



**Figura 3. Expressão de genes *pair-rule* e de polaridade segmentar em *D. melanogaster*.** Embrião inicial (parte superior) indicando os parasegmentos e os transcritos de *Even-skipped*, *Ftz*, *Engrailed* e *Hedgehog*. Embrião tardio (parte inferior) indicando os segmentos e os transcritos de *Engrailed*, *Hedgehog*, *Wingless* e *Patched*. Durante o desenvolvimento de *D. melanogaster*, *Wg* é transcrita e secretada de uma fileira anterior de células e mantém os níveis de transcritos de *En* em células ectodermiais adjacentes, na região posterior. As células que expressam *En*, por sua vez, transcrevem e secretam o ligante *Hh*, que mantém mutuamente os níveis de transcritos de *Wg* nas células anteriores vizinhas (DINARDO et al., 1988; BEJSOVEC; MARTINEZ ARIAS, 1991). A interface entre estes dois domínios adjacentes define o limite do parasegmento, com transcritos de *En/Hh* na região anterior e *Wg* na região posterior final de cada parasegmento (BAKER, 1987; LEE et al., 1992; MOHLER; VANI, 1992). Os círculos e os triângulos cinzas indicam a presença de transcritos dos genes nas fileiras de células dos segmentos. Os círculos brancos indicam a ausência de transcritos de genes nas fileiras de células dos segmentos. Fonte: Adaptado de Wolpert et al. (2015).

Outro gene de segmentação é o gene *Wg*, que pertence aos genes da família *Wnt*. Os genes *Wnt* codificam glicoproteínas secretadas (ligantes) que atuam como moléculas de sinalização de curto alcance e morfógenos de longo alcance, dependendo do contexto de desenvolvimento (SWARUP; VERHEYEN, 2012). A sinalização *Wnt* regula muitos processos de desenvolvimento, incluindo manutenção de células tronco, proliferação, migração e morte celular (LOGAN; NUSSE, 2004), sendo portanto um importante componente do *toolkit* genômico que coordena o desenvolvimento embrionário nos animais (MURAT; HOPFEN; MCGREGOR, 2010). O gene *Wg* em *D. melanogaster* funciona como um gene de segmentação primária, envolvido na especificação e manutenção dos limites dos parasegmentos. Os transcritos de *Wg* são localizados no limite posterior de cada parasegmento, junto às células que expressam *En*, no limite anterior do parasegmento (Figura 3) (MURAT; HOPFEN; MCGREGOR, 2010). A expressão de *En* e de *Wg* nos segmentos é conservada também em *T. castaneum* (CHOE; BROWN, 2009) e em outros artrópodes (DAMEN, 2002). A expressão do *Wg* no crustáceo *Triops longicaudatus* cumpre uma função semiconservativa similar ao *Wg* de *D. melanogaster*, na segmentação e formação dos membros (NULSEN; NAGY, 1999).

*D. melanogaster* apresenta 7 genes da família *Wnt* (*Wg*, *Wnt5*, *Wnt6*, *Wnt7*, *WntD/8*, *Wnt9* e *Wnt10*); mas em outros artrópodes, especialmente nos crustáceos, miriápodes e quelicerados pouco se conhece sobre a expressão e função de genes da família *Wnt* (MURAT; HOPFEN; MCGREGOR, 2010). A expressão de *Wnt8* na aranha *Achaearanea tepidariorum* e no besouro *T. castaneum* é necessária para o estabelecimento da zona de crescimento e do desenvolvimento dos segmentos posteriores. A expressão de *Wnt5* em *Cupiennus salei* e *T. castaneum* indica sua participação na segmentação em artrópodes de banda curta (DAMEN, 2002; BOLOGNESI *et al.*, 2008). Recentemente, foram identificados 12 membros da família *Wnt* no genoma de *Daphnia pulex*, mas sua função durante o desenvolvimento ainda não é conhecida (JANSSEN *et al.*, 2010). De maneira geral, estudos sobre a função dos genes *Wnt* podem contribuir para a compreensão dos mecanismos que coordenam o desenvolvimento animal (MURAT; HOPFEN; MCGREGOR, 2010).

O gene *Hh* codifica uma proteína de sinalização que é secretada pelas células, estando envolvida com a polaridade segmentar durante o desenvolvimento embrionário de *D. melanogaster*. Os transcritos deste gene estão localizados em células epidérmicas, na região posterior dos

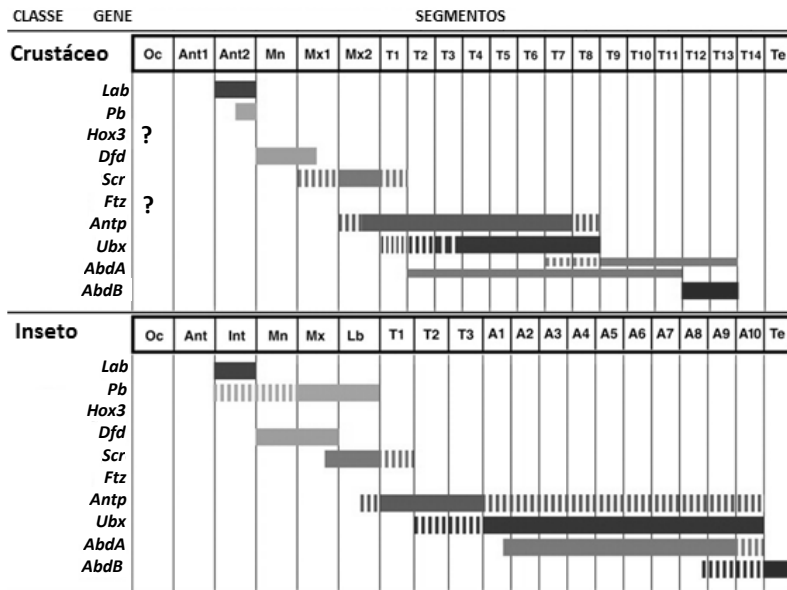
parasegmentos, junto com a expressão do gene *En* (Figura 3) (TABATA; EATON; KORNBERG, 1992; SWARUP; VERHEYEN, 2012). Durante a extensão da banda germinativa, a expressão de *Hh* é dependente de *En* (Figura 3) (TABATA; EATON; KORNBERG, 1992). Assim como observado para *Wg*, o gene *Hh* é ativado por genes *pair-rule* durante a embriogênese e, posteriormente, a expressão de *Wg* e *Hh* é regulada pela interação positiva entre seus produtos gênicos (DINARDO et al., 1994). Estudos de *Hh* no crustáceo *A. franciscana* e no quelicerado *Euscorpius flavicaudis* demonstram que a expressão *Hh* está localizada na parte posterior de cada segmento, atuando como um gene de polaridade segmentar nestas espécies. No entanto, outros sítios de expressão também foram identificados, como no rosto e nos apêndices de *E. flavicaudis* e no intestino de *A. franciscana* (SIMONNET; DEUTSCH; QUÉINNEC, 2004).

### 1.2.2. Genes *Hox*

Os genes *Hox* são necessários para a especificação da identidade dos segmentos do corpo no embrião ao longo do eixo ântero-posterior (LEWIS, 1978; MAEDA; KARCH, 2009; VOET; VOET; PRATT, 2014). Esses genes codificam fatores de transcrição e apresentam 3 características fundamentais: i) organização em complexos de genes, ii) expressão em regiões distintas na mesma sequência ao longo dos principais eixos do corpo, e iii) uma sequência de 180 pares de bases (homeodomínio), codificando um motivo de ligação ao DNA (CARROLL, 1995). Em *D. melanogaster* (Figura 2), assim como em vários artrópodes, a expressão da cascata de genes também regula a expressão dos genes *Hox* durante o desenvolvimento. Devido a esse caráter conservativo, os genes *Hox* têm sido utilizados para investigar o estabelecimento dos eixos corporais durante o desenvolvimento embrionário de espécies de artrópodes, como em *D. melanogaster*, *Schistocerca gregaria*, *T. castaneum*, *Manduca sexta*, *Bombyx mori* e *Oncopeltus fasciatus* (AKAM et al., 1994; HUGHES; KAUFMAN, 2002).

Semelhante aos insetos, os crustáceos também apresentam um único conjunto de genes *Hox* (Figura 4), composto por: *Labial (Lab)*, *Proboscipedia (Pb)*, *Zerknüllt (Zen* ou *Hox3)*, *Deformed (Dfd)*, *Sex combs reduced (Scr)*, *Fushi tarazu (Ftz)*, *Antennapedia (Antp)*, *Ultrabithorax (Ubx)*, *Abdominal A (AbdA)* e *Abdominal B (AbdB)* (HUGHES; KAUFMAN, 2002; KENNY et al., 2014; SUN et al., 2015). Portanto, as diferenças básicas na morfologia entre eles são decorrentes,

em parte, das diferenças nos domínios de expressão dos genes *Hox* e das suas funções (ABZHANOV; KAUFMAN, 2000a, 2000b).



**Figura 4. Genes *Hox* e sua expressão em crustáceos e insetos.** ocular (Oc); antena (Ant); mandibular (Mn); maxilar (Mx); tronco em crustáceo (T1-T14); telson (Te); lábio (Lb); intercalar (Int); primeiro ao terceiro torácico (T1-T3); abdominal 1 - abdominal 10 em insetos (A1-A10). Pontos de interrogação para *Hox3* e *Ftz* indicam que as funções destes genes em crustáceos ainda não são conhecidos. Nos insetos, *Hox3* e *Ftz* tem diferentes funções e não são apresentados na figura. Fonte: Adaptado de Hughes e Kaufman (2002).

Em *D. melanogaster*, os genes *Antp*, *Ubx* e *AbdA* especificam diferentes tipos de segmentos no tronco (KAUFMAN; SEEGER; OLSEN, 1990; LAWRENCE; MORATA, 1994). A expressão do gene *Antp* é necessária para a especificação de três segmentos locomotores do tórax em *D. melanogaster* (MARTINEZ-ARIAS, 1986; KAUFMAN; SEEGER; OLSEN, 1990). Ambos os genes *Ubx* e *AbdA* estão envolvidos no desenvolvimento da região abdominal (MORATA; KERRIDGE, 1981). O gene *Ubx* também se expressa na região inferior

do tórax e está envolvido no desenvolvimento dos halteres, as asas traseiras modificadas (LEWIS, 1978).

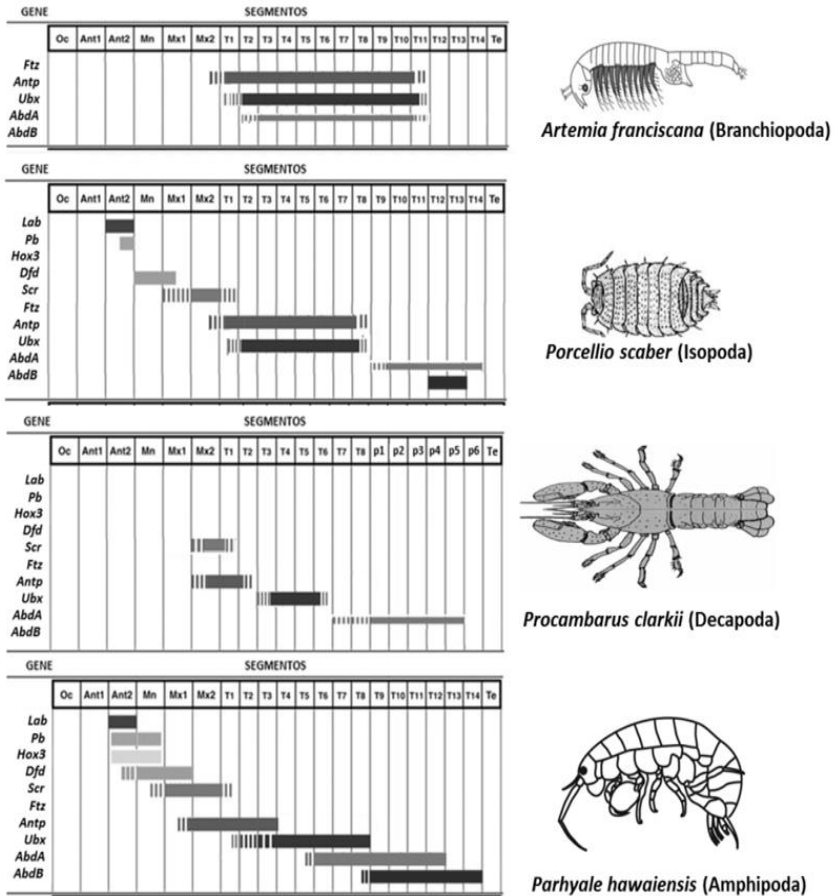
Nos crustáceos, os estudos dos padrões de expressão dos genes *Hox* durante o desenvolvimento são relativamente recentes e têm despertado especial interesse, devido ao fato dos crustáceos apresentarem a maior diversidade de formas corporais entre os artrópodes e, possivelmente, dentre os animais (FORTEY; THOMAS, 1998). Os poucos estudos da expressão dos genes *Hox* em crustáceos como em *A. franciscana* (AVEROF; AKAM, 1995), *P. scaber*, *P. clarkii* (ABZHANOV; KAUFMAN, 2000a,c) e *P. hawaiiensis* (SERANO et al., 2016) demonstram diferentes padrões de domínios de sua expressão (Figura 5). Uma análise comparativa dos padrões de expressão desses genes de *P. scaber* com outros artrópodes sugere que o último ancestral comum de insetos e crustáceos malacostracos provavelmente teve um tronco comum. Assim, os genes *Hox* podem ter facilitado o processo evolutivo de maneira diferente, por evolução de domínios de expressão distintos e com funções específicas ao longo do eixo ântero-posterior (ABZHANOV; KAUFMAN, 2000a).

### 1.3. *Macrobrachium olfersii* COMO MODELO DE ESTUDO DE DESENVOLVIMENTO DENTRO DA FAMÍLIA PALAEMONIDAE

Os palaemonídeos, popularmente conhecidos como pitus ou camarões de água doce, vivem abrigados junto às pedras ou entre a vegetação aquática (BOND-BUCKUP; BUCKUP, 1989; MÜLLER et al., 1999). As fêmeas dessa família carregam os ovos numa câmara incubadora, formada pela dilatação das pleuras abdominais e cerdas ovígeras dos pleópodos (MÜLLER et al., 1999), o que facilita estudos de reprodução, embriologia e biologia de desenvolvimento.

No Brasil, o gênero *Macrobrachium* é o mais representativo da família Palaemonidae (BOND-BUCKUP; BUCKUP, 1989), não só pelo número de espécies que o integram, mas também por sua biologia reprodutiva, distribuição geográfica, diversidade e importância econômica (BARROS; BRAUN, 1997).





**Figura 5. Padrões de expressão dos genes *Hox* em crustáceos.** Os segmentos do corpo são indicados no eixo ântero-posterior. Baseado nos estudos de alguns genes *Hox*, a expressão de cada gene *Hox* é mostrada nas diferentes espécies de crustáceos. Vista lateral de *A. franciscana* e *P. hawaiiensis*, e vista dorsal de *P. scaber* e *P. clarkii* são mostrados no lado direito. Fonte: Adaptado de Hughes e Kaufman (2002).

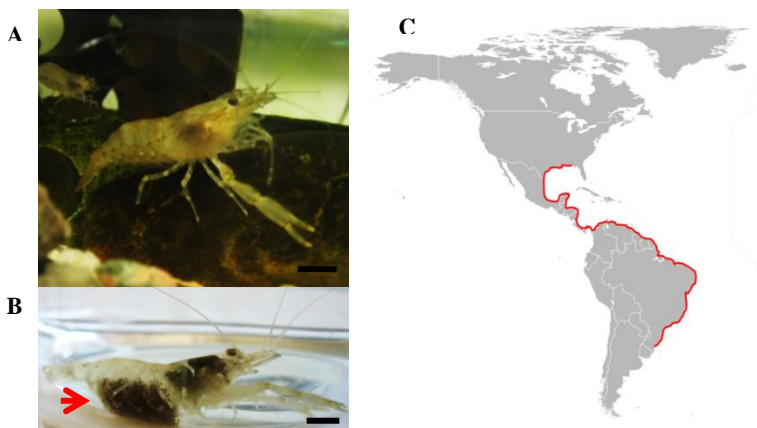
Na Ilha de Santa Catarina, estudos da biologia reprodutiva e do desenvolvimento embrionário de representantes dessa família, como *Macrobrachium acanthurus* (BRESSAN; MÜLLER, 1997, 1999), *Macrobrachium carcinus* (MÜLLER et al., 1999), *M. olfersii* (AMMAR; MÜLLER; NAZARI, 2001; MÜLLER; AMMAR; NAZARI, 2004; SIMÕES-COSTA et al., 2005), *Macrobrachium*

*potiuna* (MÜLLER; CARPES, 1991; MÜLLER; AMMAR; NAZARI, 2004), *Palaemon pandaliformis* (MÜLLER et al., 1996) e *Palaemonetes argentinus* (MÜLLER et al., 1996; MÜLLER; AMMAR; NAZARI, 2004) foram realizados pela equipe do Laboratório de Reprodução e Desenvolvimento Animal (BEG/CCB/UFSC).

Dos estudos realizados com diferentes espécies de Palaemonidae, *M. olfersii* é a espécie melhor caracterizada (Figura 6) e a que apresenta características favoráveis para ser usada como modelo de estudo para a biologia do desenvolvimento. Entre elas: (1) a adaptabilidade da espécie em aquário e possibilidade de se obter desovas em laboratório, controlando o ambiente de desenvolvimento durante todo o período de incubação; (2) o fato das fêmeas transportarem os ovos em uma câmara incubadora externa, que permite o fácil acesso aos embriões; (3) o elevado número de ovos e tempo de desenvolvimento embrionário conhecido, que possibilita a retirada parcelada de ovos e o acompanhamento sistemático dos estágios do desenvolvimento; (4) o fato dos embriões de uma mesma desova apresentarem desenvolvimento sincrônico, que contribui para a padronização (homogeneidade) das amostras; (5) a posição do embrião na superfície do ovo, o que facilita a visualização direta das mudanças morfológicas ao longo dos estágios do desenvolvimento; (6) a transparência das células embrionárias, que contribui para a visualização das mudanças internas no embrião; (7) o formato do embrião nos estágios de náuplios e pós-náuplios inicial (morfogênese e organogênese iniciais) e a sua posição em relação à massa de vitelo, que permite a confecção de preparados totais, possibilitando a realização de marcações celulares no embrião inteiro, sem a necessidade de cortes histológicos. Além disso, *M. olfersii* é uma espécie de camarão de água doce encontrada em abundância nas margens de águas rasas da Lagoa do Peri na Ilha de Santa Catarina (MÜLLER et al., 1999; AMMAR; MÜLLER; NAZARI, 2001), o que facilita sua captura para estudos de laboratório.

#### 1.4. DESENVOLVIMENTO EMBRIONÁRIO DE *M. olfersii*

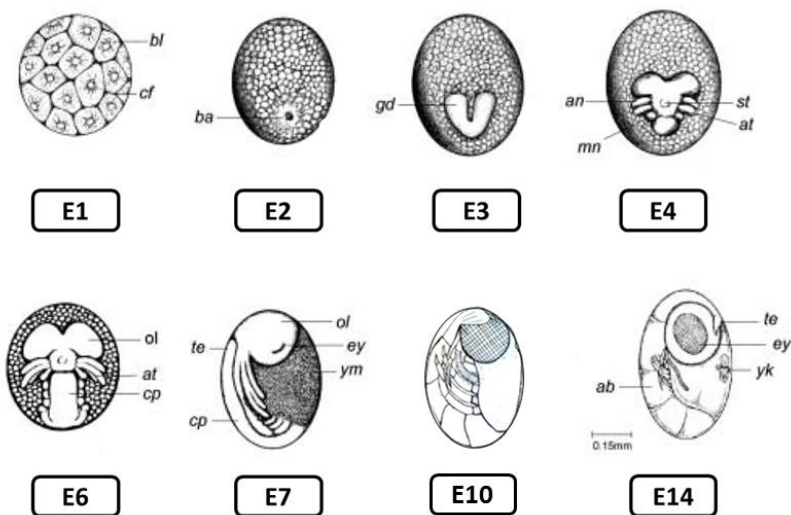
*M. olfersii* tem ampla distribuição geográfica e abrange quase toda a costa leste das Américas (Figura 6), desde o sul dos Estados Unidos, onde os camarões foram introduzidos, passando por México, Guatemala, Costa Rica, Panamá e Venezuela até o sul do Brasil (ROSSI; MANTELATTO, 2013). Na Ilha de Santa Catarina, é encontrado com frequência em rios e riachos de águas claras e com pouca profundidade (AMMAR; MÜLLER; NAZARI, 2001).



**Figura 6. Espécime de *M. olfersii* e sua distribuição geográfica.** A) Vista lateral do animal adulto. B) Vista lateral da fêmea transportando uma grande quantidade de ovos na câmara incubadora externa. Os ovos estão indicados pela seta vermelha. Barra de escala: 1cm. C) Distribuição geográfica de *M. olfersii* na costa leste das Américas. Fonte: Acervo do Laboratório de Reprodução e Desenvolvimento Animal-UFSC.

As fêmeas de *M. olfersii* transportam em média 1900 ovos por desova e o desenvolvimento embrionário tem duração de 14 dias na temperatura de 24 °C ( $\pm 2$ ) (NAZARI, et al., 2003; SIMÕES-COSTA et al., 2005). Nos estágios iniciais do desenvolvimento, o tamanho médio dos ovos é de 0,47 mm no maior eixo e 0,38 mm no menor eixo (SIMÕES-COSTA et al., 2005). Os ovos são do tipo centrolécito, cuja grande quantidade de vitelo, bem como a sua distribuição no citoplasma determinarão o tipo de clivagem e de desenvolvimento embrionário. Esses ovos, com formato levemente elíptico, apresentam o vitelo distribuído regularmente em grumos por todo o citoplasma, exceto em uma pequena área ao redor do núcleo e na região subjacente à membrana vitelínica ou membrana do ovo (MÜLLER; NAZARI; SIMÕES-COSTA, 2003; MÜLLER; AMMAR; NAZARI, 2004). Assim, a clivagem é meroblástica superficial (MÜLLER; NAZARI; SIMÕES-COSTA, 2003) com a formação da camada celular ou sincicial - a blastoderma - envolvendo toda a massa de vitelo (JIRIKOWSKI; RICHTER; WOLFF, 2013).

A caracterização dos principais eventos do desenvolvimento embrionário de *M. olfersii*, através do estagiamento por dia embrionário (E) (Figura 7), são descritos a seguir:



**Figura 7. Desenvolvimento embrionário de *M. olfersii*.** Dia embrionário (E). Vista ventral (E2-E6). Vista lateral (E7-E14). Antênula (an), antena (at), área blastoporal (a), blastômeros (bl), sulcos de clivagem (cf), papila caudal (cp), olho (ey), disco germinativo (gd), mandíbula (mn), lobo óptico (ol), apêndice pós-naupliar (pa), estomodeu (st), télson (te), vitelo (yk), massa de vitelo (ym). Fonte: adaptado de Müller; Nazari; Simões-Costa (2003).

**E1-E2:** Formação das enérgides no centro da massa de vitelo, posteriormente estas migram para a periferia para formar o blastoderma sincicial e depois o blastoderma celular (blastômeros).

**E3:** Inicia o processo de gastrulação a partir da formação da área blastoporal, resultante da migração dos blastômeros. Esses organizam o disco germinativo, que adquire a forma de “V”. As extremidades superiores do disco germinativo darão origem aos lobos ópticos e a extremidade inferior, à papila caudal. Na superfície do ovo ainda são observados alguns blastômeros dispersos, que originarão o ectoderma extraembrionário. Nesse estágio embrionário se reconhece o eixo ântero-posterior do embrião (SCHOLTZ, 1992; SIMÕES-COSTA et al., 2005).

**E4:** Caracterizado pela formação do náuplios embrionizado, constituído pelos lobos ópticos, por três pares de apêndices naupliares,

que são as antênulas, antenas e mandíbulas e ainda pela papila caudal. A formação do náuplios embrionizado dá início à morfogênese.

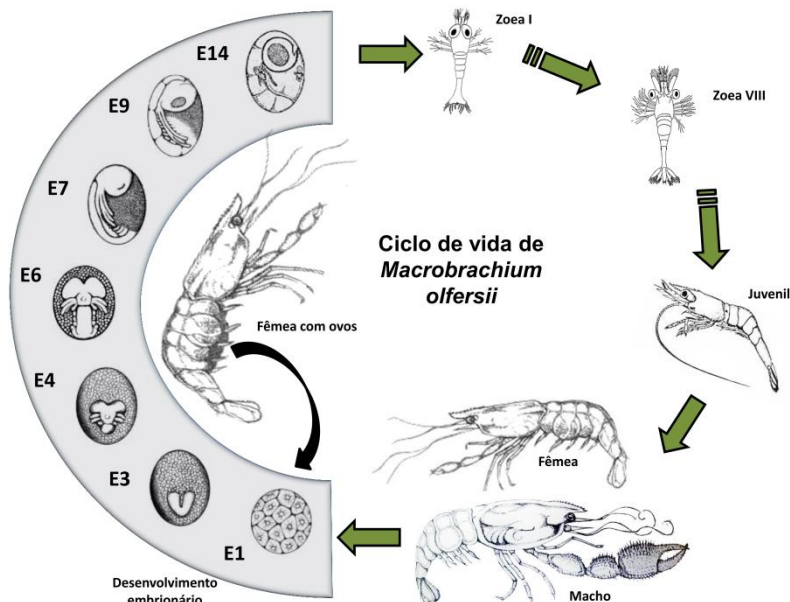
**E5 - E6:** A morfogênese prossegue e organiza-se o pós-náuplios embrionizado, caracterizado pelo alongamento do embrião no sentido ântero-posterior adquirindo formato semelhante à letra “C”. Nesses estágios ocorrem o crescimento e encurvamento da papila caudal, bifurcação dos apêndices naupliares, formação dos primeiros cromatóforos e pelo surgimento dos apêndices pós-naupliares, que são as maxílulas, as maxilas e os primórdios dos segmentos torácicos. Dessa forma, tem início a formação e organização do abdome (SIMÕES-COSTA et al., 2005).

**E7 - E9:** O processo de organogênese inicial em *M. olfersii* pode ser evidenciado pela pigmentação do olho na região distal do lobo óptico. A massa de vitelo, ainda presente na porção central do ovo, determina que o crescimento do embrião aconteça superficialmente no sentido céfalo-caudal. Observa-se a bifurcação dos apêndices naupliares e pós-naupliares. O sistema nervoso em formação é reconhecido pela organização dos gânglios cerebrais e abdominais.

**E10 - E14:** O olho apresenta um crescimento bastante expressivo, os apêndices naupliares e pós-naupliares apresentam cerdas e estão nitidamente bifurcados e a carapaça dorsal recobre todo o cefalotórax. A massa de vitelo diminui, gerando espaço para o crescimento do embrião. O olho aumenta e o télson atinge a porção mais céfalica do embrião. Formam-se a carapaça do cefalotórax e os segmentos abdominais. Nos estágios finais do desenvolvimento embrionário, o coração, o intestino primitivo, os gânglios cerebrais e os apêndices corporais estão preparados para a eclosão (SIMÕES-COSTA et al., 2005).

Após a eclosão tem início o desenvolvimento larval, que consiste em 12 estágios (zoeas) (DUGGER; DOBKIN, 1975), sendo que alguns estágios larvais precisam de salinidade durante seu desenvolvimento para completar o ciclo de vida da espécie (MCNAMARA; MOREIRA; SOUZA, 1986; DUGGER; DOBKIN, 1975; ROSSI; MANTELATTO, 2013). As larvas tornam-se juvenis, os quais migram para água doce e, ainda, apresentam porte semelhante entre machos e fêmeas até atingir a maturidade sexual (AMMAR; MÜLLER; NAZARI, 2001). Os adultos apresentam dimorfismo sexual externo acentuado, sendo que os machos são maiores, com os segundos pares de quelípodos robustos e com as quelas de tamanho desigual (BOND-BUCKUP; BUCKUP, 1989; JALIHAI; SANKOLLI; SHENOY, 1993; SAMPAIO et al., 2009) (Figura 8). Os quelípodos de

*M. olfersii*, além de diferenciar morfologicamente machos de fêmeas, desempenham função reprodutiva (AMMAR; MÜLLER; NAZARI, 2001).



**Figura 8. Ciclo de vida de *M. olfersii*.** Alguns estágios do desenvolvimento embrionário e larvais são mostrados. O tamanho dos embriões, larvas, juvenis e adultos não estão em proporção de tamanho. As ilustrações dos embriões foram de Müller et al. (2003) e dos adultos foram cortesia do Dr. Ammar. As ilustrações das larvas foram de Dugger e Dobkin (1975).

## 1.5. JUSTIFICATIVA

Embora os trabalhos atuais em embriologia sejam fortemente amparados por técnicas de biologia celular e molecular, há uma lacuna no estudo do desenvolvimento embrionário da maioria das espécies de crustáceos, em especial de *M. olfersii*, cujos mecanismos reguladores do desenvolvimento permanecem pouco esclarecidos. Um aspecto limitante é a falta de pesquisas sobre as seqüências de genes relacionadas ao desenvolvimento e o sequenciamento do genoma nesta espécie. A partir dessa informação seria possível gerar sondas específicas, validar genes

de referência e desenhar iniciadores para serem usados em RT-qPCR para determinar os níveis de transcritos de genes no desenvolvimento embrionário, como também outros estudos moleculares como transgeneses e *knockdown* de gene.

O camarão *M. olfersii* vem se tornando um novo modelo de estudo em biologia do desenvolvimento e de toxicidade no desenvolvimento. O desenvolvimento embrionário dessa espécie apresenta algumas características, mencionadas no *item* 1.4, que a tornam um excelente modelo dentro dos camarões palaemonídeos de água doce. Além disso, essa espécie tem sido utilizada em diferentes tipos de estudos que incluem estudos fisiológicos (SOUZA; MOREIRA, 1987; AUGUSTO et al., 2007; MENDONÇA et al., 2007; RIBEIRO; MCNAMARA, 2009), dinâmica populacional (MULLER; PRAZERES, 1992; ANGER; MOREIRA, 1998), com agentes físicos (NAZARI et al., 2010; ZENI et al., 2015) e químicos (MARTINS et al., 2006; BARBIERI et al., 2013), estudos imunológicos (ROSA; STOCO; BARRACCO, 2008), em sistemática molecular (MOSSOLIN; PILEGGI; MANTELATTO, 2010), entre outros. Portanto, existe uma necessidade de identificar transcritos de genes nesse camarão para entender os mecanismos moleculares nessa espécie, em especial durante o desenvolvimento embrionário.

Nesse sentido, considerando a escassez de: i) sequências de genes em *M. olfersii*, ii) descrição de genes, especialmente relacionados ao desenvolvimento embrionário de *M. olfersii* e de outros camarões de água doce, iii) estudos dos níveis de transcritos dos genes *Hox* e de segmentação durante o desenvolvimento de *M. olfersii*, e iv) estudos de identificação e validação de genes de referência para experimentos de RT-qPCR em *M. olfersii*, o presente estudo pretende contribuir com estas informações. Assim, os resultados do presente estudo poderão ser utilizados na orientação de futuras pesquisas para um melhor entendimento das bases moleculares dos eventos do desenvolvimento embrionário, bem como para estudos ecotoxicológicos na modulação desses genes frente a agentes físicos ou químicos. Além disso, estudos de estabelecimento dos eixos corporais em espécies de decápodes como *M. olfersii* são importantes para a compreensão do desenvolvimento embrionário e da filogenia deste grupo. Os conhecimentos gerados com essa espécie poderão ser utilizados também como modelo para compreensão dos mecanismos de ação desses genes em outras espécies de camarões de água doce.

Portanto, as hipóteses deste trabalho são:

- Nos estágios de E4-E8 de *M. olfersii*, transcritos de genes homólogos a outros artrópodes estão presentes, que participam de uma diversidade de vias biológicas, especialmente os genes relacionados ao desenvolvimento embrionário.
- Os genes *Hox* estão relacionados com a identidade segmentar dos embriões de *M. olfersii*.
- Os genes *Even-skipped*, *Engrailed*, *Hedgehog*, *Fushi tarazu*, *Hedgehog* e *Wingless* estão relacionados com a segmentação dos embriões de *M. olfersii*.
- Candidatos de genes de referência como fator de alongação  $1\alpha$  (*Ef-1 $\alpha$* ) e da proteína ribossomal L8 (*Rpl8*) são expressos nos estágios de E4-E8, sendo estáveis durante o desenvolvimento embrionário e podem ser utilizados nesta espécie como normalizadores para análises de RT-qPCR.



## 2. OBJETIVOS

### 2.1. OBJETIVO GERAL:

Identificar e caracterizar transcritos de genes de *M. olfersii*, especialmente dos genes relacionados ao desenvolvimento embrionário, e analisar a expressão temporal nos embriões de E3-E10.

### 2.2. OBJETIVOS ESPECÍFICOS:

- Identificar transcritos de *M. olfersii*, através de análise transcriptômica de embriões de E4-E8;
- Identificar transcritos de genes próprios do genoma mitocondrial de *M. olfersii*;
- Identificar transcritos de genes relacionados ao desenvolvimento embrionário e das vias de desenvolvimento de *M. olfersii*;
- Analisar a similaridade das sequências de genes relacionados com o desenvolvimento de *M. olfersii* e a filogenia dos genes dentro do filo artrópode, utilizando as sequências disponíveis no *GenBank*;
- Identificar genes de referência para análises de RT-qPCR e avaliar sua estabilidade de expressão durante o desenvolvimento embrionário de *M. olfersii* e em tecidos de animais adultos.
- Determinar o perfil temporal de transcritos dos genes *Hox* e de segmentação em embriões de *M. olfersii*, pela RT-PCR semi-quantitativa nas idades de E3 a E10;
- Quantificar os níveis de transcritos dos genes *Hox* e de segmentação em embriões de *M. olfersii*, pela RT-qPCR, nas idades de E3 a E10.
- Detectar os transcritos dos genes *Hox* e de segmentação no início e final do desenvolvimento embrionário de *M. olfersii*, pela RT-PCR;



### 3. ORGANIZAÇÃO DA TESE

O presente estudo realizou uma montagem *de novo* e a análise transcriptômica de um *pool* de embriões (E4-E8) para proporcionar informação geral dos transcritos de diferentes processos celulares, componentes celulares e função molecular, assim como transcritos de vias biológicas agindo durante o desenvolvimento embrionário de *M. olfersii* (Capítulo I). Além disso, o principal foco foram os genes relacionados ao desenvolvimento embrionário e vias de desenvolvimento, portanto, estes genes foram identificados a partir do transcriptoma de *M. olfersii* (Capítulo II). Em paralelo, genes relacionados ao desenvolvimento classicamente estudados em outros artrópodes, foram identificados por estratégia de RT-PCR com iniciadores degenerados e sequenciamento (Capítulo II). Considerando que a validação de genes de referência é necessária para a precisão dos níveis de transcritos de genes alvo, 6 genes de referência, classicamente estudados em animais, foram selecionados do transcriptoma e sua estabilidade de expressão avaliada nos diferentes estágios embrionários e em três tecidos de animais adultos (Capítulo III). Por último, RT-PCR, RT-PCR semi-quantitativa e RT-qPCR foram utilizadas para analisar os transcritos de genes relacionados ao desenvolvimento, classicamente estudados em outros artrópodes (Capítulo IV) e sugerir suas funções nos eventos morfológicos que acontecem no desenvolvimento embrionário de *M. olfersii*. Ao final é apresentada uma discussão geral do tema da tese, bem como conclusões e perspectivas do trabalho. Os materiais e métodos utilizados no presente estudo estão resumidos na Figura 9.



## CAPITULO I. Montagem *de novo* e análise do transcriptoma de embriões de *Macrobrachium olfersii*

---

### 1. INTRODUÇÃO

O camarão de água doce *M. olfersii* (Crustacea: Decápode: Palaemonidae), devido a sua ampla distribuição geográfica ao longo das costas orientais das Américas (MÜLLER; NAZARI; SIMÕES-COSTA, 2003; ROSSI; MANTELATTO, 2013), tem sido utilizado em diferentes estudos. Esses estudos compreendem o desenvolvimento embrionário (MÜLLER; NAZARI; SIMÕES-COSTA, 2003; SIMÕES-COSTA et al., 2005) e larval (DUGGER; DOBKIN, 1975), aspectos reprodutivos (AMMAR; MÜLLER; NAZARI, 2001; MOSSOLIN; BUENO, 2002; NAZARI et al., 2003; MAGALHÃES; MOSSOLIN; MANTELATTO, 2012), filogenia molecular (MURPHY; AUSTIN, 2005; PILEGGI; MANTELATTO, 2010; ROSSI; MANTELATTO, 2013), entre outros. Desse modo, essa espécie mostra um grande potencial como modelo emergente para estudos em biologia do desenvolvimento, de toxicologia no desenvolvimento e toxicologia ambiental. No entanto, pouco se conhece sobre os mecanismos moleculares em diferentes abordagens científicas, em especial durante o desenvolvimento embrionário.

Os mecanismos moleculares de uma espécie poderiam ser conhecidos através da identificação de genes ou transcritos utilizando estudos genômicos ou transcriptômicos. O termo genoma refere-se ao conteúdo total de material genético de um organismo, enquanto o termo transcriptoma refere-se ao conjunto de transcritos em um tempo e condição determinada em uma célula ou espécie. Nos últimos anos, a quantidade de sequências de DNA e cDNA em bancos de dados públicos tem crescido exponencialmente devido, principalmente, a redução dos custos e aumento da velocidade de obtenção de resultados (JUNG et al., 2013). No entanto, as bases de dados públicas possuem poucas sequências de crustáceos, em comparação a outros artrópodes. Estudos genômicos e obtenção de dados de EST (Etiqueta de sequência expressa, do inglês *expressed sequence tag*) foram especificamente focados em determinados grupos de crustáceos com interesses em áreas que abrangem ecologia e biologia evolutiva (*Daphnia*), biologia das espécies invasoras (*Eurytemora affinis*), balanço hídrico e de íons (*Carcinus maenas*), fisiologia térmica (gênero *Petrolisthes*) e biologia evolutiva (*P. hawaiiensis*) (STILLMAN et al., 2008). Recentemente, os métodos de sequenciamento de nova geração facilitaram a identificação

de um maior número de sequências dentre os crustáceos e alguns projetos de genomas e transcriptomas estão em andamento (caranguejos, lagostins, krill, lagostas, camarões e pulga da água) (JUNG et al., 2013).

Até maio de 2016, um total de 922.019 EST de crustáceos obtidos de bibliotecas de cDNA estão disponíveis na base de dados *GenBank* (<http://www.ncbi.nlm.nih.gov/nucest/?term=crustacea>), os quais incluem *L. vannamei* (161.377 EST), *D. pulex* (152.659 EST), *Lepeophtheirus salmonis* (129.250 EST), *Petrolisthes cinctipes* (97.806 EST), *Penaeus monodon* (39.877 EST), *A. franciscana* (37.702 EST), entre outros. No entanto, os projetos de sequenciamento de genomas de crustáceos são limitados em organismos considerados não-modelo ou devido ao tamanho grande do genoma de alguns crustáceos (LENZ et al., 2014), como em *Calanus finmarchicus* (valor C > 5 pg que corresponde a  $4,89 \times 10^9$  pb) e *M. acanturus* (valor C de 6,48 que corresponde a  $6,34 \times 10^9$  pb) em comparação ao genoma sequenciado de *D. pulex* (valor C de < 0,4 pg que corresponde a <  $0,39 \times 10^9$  pb) (DOLEZEL et al., 2003; [www.genomesize.com](http://www.genomesize.com)). O valor C se define como a quantidade de DNA genômico contido em um núcleo haplóide de um organismo. No entanto, um genoma grande não se refere, necessariamente, a uma maior complexidade morfológica.

Entre os crustáceos, os genomas de alguns decápodes têm sido sequenciados ou estão quase prontos para ser disponibilizados, tais como *Neocaridina denticulata* (KENNY et al., 2014), *L. vannamei* (ZHAO et al., 2012; YUAN et al., 2013; YU et al., 2015) e *D. pulex* (COLBOURNE et al., 2011). No entanto, em organismos com sequências genômicas não conhecidas, o sequenciamento do transcriptoma por RNA-Seq abre as portas para analisar a expressão global de genes. Assim, RNA-Seq não está limitado a detectar transcritos que correspondem a sequências genômicas conhecidas (SENGUPTA et al., 2011). Portanto, a montagem do transcriptoma na ausência de sequências genômicas existentes é conhecida como montagem *de novo* (SURGET-GROBA; MONTOYA-BURGOS, 2010). Uma abordagem para a montagem *de novo* de transcriptomas usando várias longitudes de k-mer é altamente desejável e tem mostrado ser muito útil (SURGET-GROBA; MONTOYA-BURGOS, 2010). Um comprimento de k-mer superior irá, teoricamente, gerar uma montagem mais contígua de transcritos altamente expressos, enquanto que transcritos com baixa expressão serão mais facilmente obtidos se um comprimento k-mer inferior for usado (ZERBINO; BIRNEY, 2008). As sequências obtidas do sequenciamento (*reads*) são tipicamente de 30-400 pb, dependendo da tecnologia de sequenciamento de DNA usada.

Em princípio, qualquer tecnologia de sequenciamento de nova geração pode ser utilizada para RNA-Seq, entre elas os sistemas Illumina, SOLiD de *Applied Biosystem* e *Roche 454 Life Science system* têm sido usados (WANG; GERSTEIN; SNYDER, 2009). O analisador de genoma Illumina realiza o sequenciamento de um grande número ( $> 10^7$ ) de *reads* de sequência curta ( $< 150$  pb). A abordagem *paired-end RNA-Seq*, em que um único *read* é sequenciado em ambas extremidades, permite rastrear junções de *splicing* alternativo, inserções e deleções, e tem sido usado para a montagem de transcriptoma *de novo* (SENGUPTA et al., 2011).

Dados transcriptômicos valiosos por RNA-seq têm sido obtidos para crustáceos, tais como *L. vannamei* (SOOKRUKSAWONG et al., 2013; LI et al., 2013; GHAFFARI et al., 2014), *Fenneropenaeus chinensis* (LI et al., 2013), *Eriocheir sinensis* (HE et al., 2012, 2013), *Macrobrachium rosenbergii* (RAO et al., 2015), *Macrobrachium nipponense* (MA et al., 2012; JIN et al., 2013), *Euphausia superba* (CLARK et al., 2011), *Portunus trituberculatus* (LV et al., 2013, 2014; MENG et al., 2015), *Calanus sinicus* (YANG et al., 2014), *C. finmarchicus* (LENZ et al., 2014), *Caligus rogercresseyi* (GALLARDO-ESCÁRATE; VALENZUELA-MUÑOZ; NUÑEZ-ACUÑA, 2014) entre outros. No entanto, estudos transcriptômicos de *M. olfersii* ainda não foram realizados e na base de dados *GenBank* estão disponíveis 107 sequências que correspondem a 8 genes desta espécie, e nenhum é relacionado ao desenvolvimento. Nesse sentido, o presente estudo realizou o sequenciamento na plataforma Illumina *paired-end*, montagem *de novo* com múltiplos k-mers e análise do transcriptoma de embriões de *M. olfersii* para obter informação de transcritos desta espécie durante o desenvolvimento embrionário.

## 2. MATERIAIS E MÉTODOS

### 2.1. OBTENÇÃO DE ANIMAIS E EMBRIÕES DE *M. olfersii*

O procedimento de coleta de *M. olfersii* foi aprovado pelo Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis, número 15294-1/IBAMA/2008. Fêmeas e machos adultos foram coletados na Lagoa de Peri (27°35' S, 48°35' W) transportados ao Laboratório de Reprodução e Desenvolvimento Animal- CCB/UFSC. Os camarões foram mantidos em aquário (60L), a 24°C ( $\pm 1$ ), com aeração constante, em condições de fotoperíodo (12 h claro: 12 h escuro) e alimentados diariamente com ração balanceada (Alcon Bottom

Fish). Uma proporção de 3:1 (fêmeas: machos) por aquário foi utilizada para favorecer a fecundação. O estagiamento dos embriões foi realizado segundo o método por dia embrionário (SIMÕES-COSTA et al., 2005). Embriões de E4 a E8 foram utilizados para a extração de RNA total, com a finalidade de realizar uma montagem do transcriptoma de embriões, para obter transcritos destas idades de desenvolvimento, especialmente os transcritos relacionados com o padrão do corpo e segmentação, assim como transcritos envolvidos nos processos de morfogênese e organogênese.

## 2.2. EXTRAÇÃO DE RNA TOTAL

A extração de RNA foi realizada para cada estágio, de E4- E8 (50-100 mg de ovos/estágio, aproximadamente 700-1500 ovos/estágio), usando 1 mL de Brazol por amostra (LGC Biotecnologia) de acordo com as instruções do fabricante. Em resumo, as amostras foram homogeneizadas em Brazol e mantidas a  $-20^{\circ}\text{C}$ . Após serem descongeladas, as amostras foram centrifugadas (12.000 g por 10 min a  $4^{\circ}\text{C}$ ) e o sobrenadante foi recuperado para a extração de RNA. Em seguida, foi adicionado 200  $\mu\text{L}$  de clorofórmio gelado (5:1 Brazol/Clorofórmio) e centrifugadas novamente (12.000 g por 15 min a  $4^{\circ}\text{C}$ ). O sobrenadante foi coletado e adicionado 500  $\mu\text{L}$  de isopropanol e novamente centrifugado (12.000 g por 15 min a  $4^{\circ}\text{C}$ ). Depois de eliminar o sobrenadante, o sedimento (*pellet*) foi lavado com etanol 75%, centrifugado (12.000 g por 10 min a  $4^{\circ}\text{C}$ ) e diluído em água tratada com 1% de dietilpirocarbonato (DEPC). As amostras foram tratadas com DNase I (1 U/mL, Thermo Scientific®) durante 30 min a  $37^{\circ}\text{C}$  e 10 min a  $65^{\circ}\text{C}$ . Em seguida, as amostras (100  $\mu\text{L}$ ) foram precipitadas utilizando 10  $\mu\text{L}$  acetato de sódio (3 M, pH 5,2) e 500  $\mu\text{L}$  de isopropanol. Depois da centrifugação a 12.000 g durante 10 min a  $4^{\circ}\text{C}$ , os sedimentos foram lavados com etanol 75%, secadas, e solubilizadas em água tratada com 1% de DEPC. A integridade de RNA foi verificada por eletroforese em géis de agarose a 1,5% corado com GelRed (1:500, Biotium). As eletroforeses foram realizadas em tampão TAE 0,5 X de corrida (Tris-hidroximetil-aminometano, ácido acético glacial, ácido etilenodiaminotetracético) a 100 V durante 30 min. A quantificação do RNA total foi realizada através do Nano espectrofotômetro BIO-5000-BI (KASUAKI), e as amostras com valores de 260/280 e 260/230  $> 1,8$  foram selecionados. Quantidades equimolares de RNA total, extraídos de cada idade embrionária - E4 a



E8 (4,1 µg de RNA/estágio), foram misturadas e usadas para a construção da biblioteca de RNA-Seq.

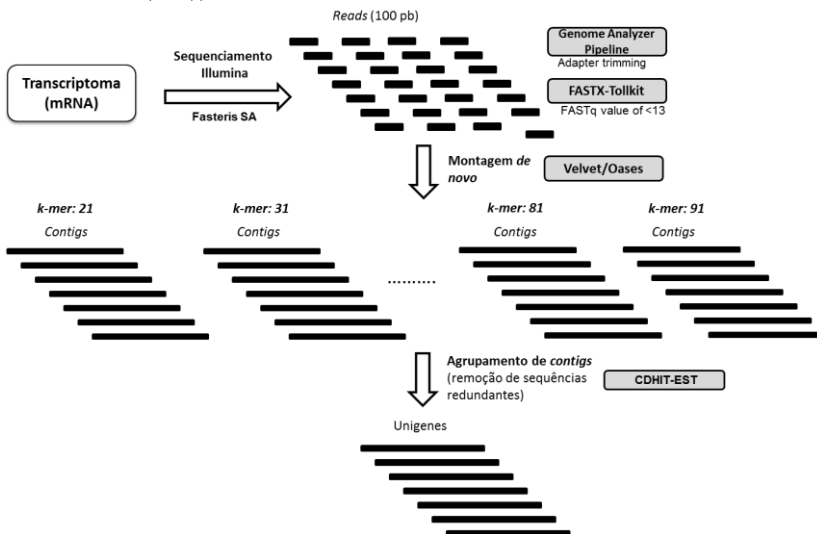
### 2.3. CONSTRUÇÃO DE BIBLIOTECA DE RNA-Seq E SEQUENCIAMENTO

O RNA total (20,5 µg) foi enviado à Fasteris SA (Plan-les-Ouates, Switzerland) para seu processamento. Uma biblioteca *paired-end* foi construída e sequenciada usando a plataforma Illumina® HiSeq2000. Em resumo, para a construção da biblioteca de RNA-seq, o sequenciamento de transcritos poliadenilados consistiu dos seguintes passos: Purificação de mRNA poli-A, síntese de cDNA utilizando iniciadores poli(T), método *shotgun* para gerar insertos de 250 nt, ligação dos adaptadores 3p e 5p, pré-amplificação, geração de colônia por amplificação pela PCR e sequenciamento *paired-end*. As sequências adaptadoras de 5' e 3' foram removidas (*trimming*) utilizando o *Genome Analyzer Pipeline* (Illumina) (Fasteris SA) e os dados de saída de Illumina incluíram *reads* “limpos” de 100 bases. A qualidade de todos os *reads* foram analisados e os *reads* com valores de qualidade FASTq < 13 foram removidos com o *software* FASTX-Tollkit. Os *reads* do presente estudo foram depositados no arquivo de sequências de *reads* (SRA) no NCBI sob o número de acesso SRS1112758.

### 2.4. MONTAGEM *de novo* DO TRANSCRIPTOMA DE EMBRIÕES DE *M. olfersii*

A montagem das sequências foi realizada no Laboratório de Genomas e Populações de Plantas - LGPP (Universidade Federal do Rio Grande do Sul, UFRGS) (Figura 10). Depois da remoção de *reads* de baixa qualidade e adaptadores das sequências, os dados de RNA-seq foram montados *de novo* dentro de *contigs* (sequências que são derivadas da coleção de *reads* sobrepostos) usando o *software* Velvet/Oases (SCHULZ et al., 2012) baseada na estratégia de multiples *k-mers* (21, 31, 41, 51, 61, 71, 81, 91 pb) e um mínimo de longitude de *contig* de 150 nt. A estratégia de múltiplos *k-mers* foi utilizada para melhorar a especificidade e sensibilidade da montagem e capturar transcritos de genes, especialmente transcritos menos abundantes. Para avaliar e remover sequências redundantes, ou com alta similaridade, todos os *contigs* obtidos com diferentes montagens de *k-mers* foram agrupados (*merged*) para produzir uma montagem combinada utilizando CDHIT-EST (LI; GODZIK, 2006) com um *threshold* final de identidade

de 95%. Após esta etapa, foram obtidos unigenes (sequências que são derivadas da montagem agrupada depois da remoção de *contigs* redundantes) com um comprimento mínimo de 150 nt. As métricas utilizadas para analisar a qualidade na montagem do transcriptoma incluíram o número total (cobertura) de *contigs*, a média da longitude de *contigs* e a diversidade de *contigs* (o número estimado de *contigs* não redundantes (NR)).



**Figura 10. Desenho esquemático do processo de montagem *de novo* dos unigenes.** Depois do sequenciamento na plataforma Illumina®, os *reads* (sequências geradas do sequenciamento) são utilizados para gerar os *contigs* (sequências geradas na montagem *de novo* derivadas da coleção de *reads* sobrepostos) com diferentes tamanhos ou valores de k-mer (k-mer é definido como subsequências geradas por decomposição, o qual é utilizado para procurar sobreposição com outros *reads*). Os *contigs* gerados com os diferentes valores de k-mer são agrupados em unigenes (sequências finais obtidas da montagem *de novo* que são derivadas da montagem agrupada depois da remoção de *contigs* redundantes). As caixas da cor cinza indicam os programas utilizados.

## 2.5. ANOTAÇÃO DE TRANSCRITOS DE GENES, DOMÍNIOS/FAMÍLIAS DE PROTEÍNAS E CLASSIFICAÇÃO GO/KEGG

Todos os unigenes foram comparados para identificação por similaridade de sequência, com aqueles armazenados no banco de dados de proteínas NR do NCBI utilizando Blastx ( $E\text{-value} < 1e^{-5}$ )

(ALTSCHUL et al., 1997); implementado no *software* Blast2GO v3.0 (CONESA; GÖTZ, 2008). O melhor resultado de alinhamento foi selecionado para anotar os unigenes. A anotação foi aprimorada por análises de domínios/famílias conservadas utilizando *InterProScan tool* de múltiplos bancos de dados, incluindo Gene3D, PANTHER, Pfam, PIR, PRINTS, ProDom, ProSITE, SMART, SUPERFAMILY e TIGERFAM. A classificação da ontologia gênica (*Gene Ontology*) foi determinada por *GOslim tool* a partir de *software* Blast2GO. A classificação de unigenes em vias biológicas foi realizada utilizando a base de dados de *Kyoto Encyclopedia of Genes and Genomes (KEGG)* (<http://www.genome.jp/kegg/>) e Blastall no *software* Blast2GO.

## 2.6. IDENTIFICAÇÃO DE TRANSCRITOS DE GENES DO GENOMA MITOCONDRIAL

Para identificar transcritos de genes (codificantes de proteínas, rRNA e tRNA) próprios do genoma mitocondrial de *M. olfersii*, o algoritmo Blastn ou tBlastn ( $E\text{-value} < 1e^{-5}$ ) foi utilizado. Para esta finalidade, sequências de aminoácidos ou de nucleotídeos de genes do genoma mitocondrial das espécies *Macrobrachium lanchesteri* (*GenBank ID*: NC\_012217.1); *M. rosenbergii*, (*GenBank ID*: NC\_006880.1); *M. nipponense* (*GenBank ID*: NC\_015073.1) foram comparadas com os unigenes do transcriptoma. A fase aberta de leitura (*ORF*, do inglês *open reading frame*) foi determinada para as sequências dos genes mitocondriais obtidas de *M. olfersii* utilizando o *ORF Finder* (<https://www.ncbi.nlm.nih.gov/orffinder/>). O algoritmo Blastp ou Blastn (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) foi realizado para determinar a similaridade com sequências NR do banco de dados *GenBank*.

## 3. RESULTADOS

### 3.1. SEQUENCIAMENTO NA PLATAFORMA ILLUMINA E MONTAGEM DE SEQUÊNCIAS

Um total de 25.636.097 *reads* de alta qualidade foram obtidos do sequenciamento na plataforma Illumina. Depois da montagem com múltiplos k-mers, foram obtidos diferentes quantidades de *contigs* para cada k-mer (Tabela 1). Um total de 457.896 *contigs* foi obtido e agrupado em 99.751 unigenes. A média do comprimento dos unigenes foi de 1.195 pb e o N50 foi de 2.440 pb. O N50 é definido como o comprimento do *contig* mais longo de tal modo que todos os *contigs*

com, pelo menos, este comprimento compõem pelo menos, 50% da montagem. O resumo dos *reads*, *contigs* e unigenes obtidos são indicados na Tabela 2.

**Tabela 1.** Parâmetros dos *contigs* obtidos utilizando Velvet/Oases com múltiplos k-mers para a montagem do transcriptoma de *M. olfersii*.

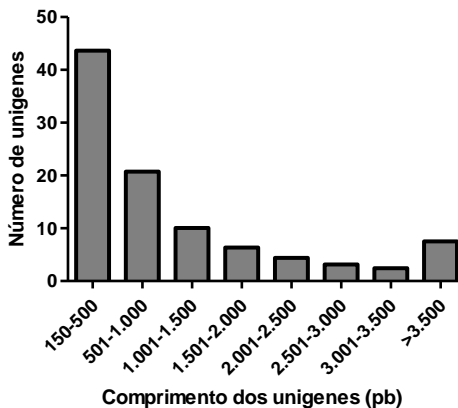
Descrição	k-mer	k-mer	k-mer	k-mer	k-mer	k-mer	k-mer	k-mer	Montagem combinada <sup>a</sup>
	21	31	41	51	61	71	81	91	
Número de <i>contigs</i>	108.765	84.123	67.848	60.120	53.697	42.914	29.425	11.004	99.751
Mediana do tamanho de <i>contig</i>	492	482	557	545	555	537	520	455	592
Média do tamanho de <i>contig</i>	1.146	1.171	1.262	1.227	1.220	1.133	999	783	1.195
Tamanho do maior <i>contig</i>	40.355	25.242	15.079	14.937	14.045	14.332	12.743	10.148	40.355
Número de <i>contig</i> > 1kpb	35.788	27.903	24.543	21.188	19.225	14.408	8.802	2.446	34.399
N50	2.616	2.755	2.789	2.691	2.584	2.267	1.782	1.129	2.440

(<sup>a</sup>) Os parâmetros de coluna "montagem combinada" correspondem aos unigenes e não aos *contigs*.

**Tabela 2.** Resumo de dados de RNA-Seq para a montagem *de novo* do transcriptoma de *M. olfersii* utilizando Velvet/Oases com múltiplos k-mers.

Sequência	Número	Média do tamanho (pb)	N50	Total de nucleotídeos (pb)
Read	25.636.097 x 2	100		5.127.219.400
Contig	457.896	1.168	2.557	534.915.933
Unigene	99.751	1.195	2.440	119.207.806

Diferentes comprimentos dos unigenes foram obtidos para os 99.751 unigenes do transcriptoma de *M. olfersii*, sendo que 44,4% contém 150-500 pb, 21,1% contém 500-1000 e 34,5% contém mais do que 1000 pb (Figura 11).



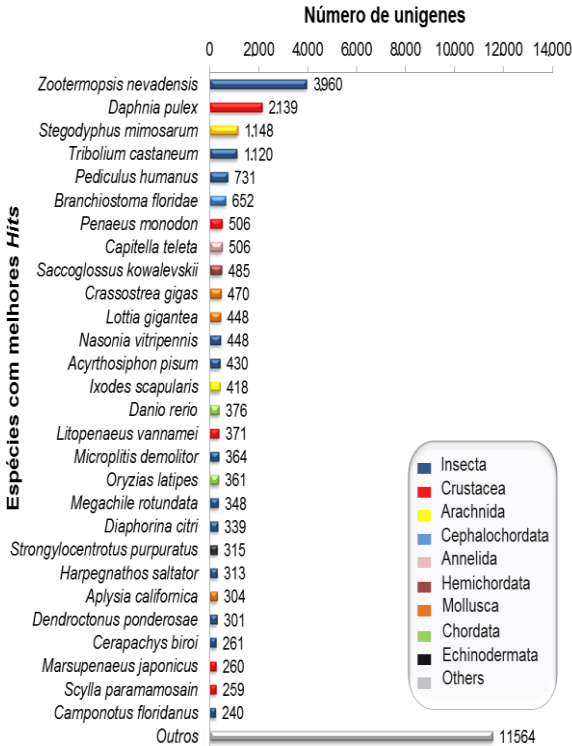
**Figura 11. Distribuição do comprimento dos unigenes montados de *M. olfersii* em função do número de unigenes.** Diferentes comprimentos dos unigenes foram obtidos para os 99.751 unigenes. Aproximadamente, 34,5% dos unigenes contém mais do que 1000 pb.

### 3.2. ANOTAÇÃO DE UNIGENES DO TRANSCRIPTOMA

Os 99.751 unigenes obtidos do transcriptoma foram submetidos a anotação por comparação de similaridade com o banco de dados não redundante (NR). Os resultados mostraram que 20.893 (20,95% do total) de 99.751 unigenes apresentaram similaridade com proteínas conhecidas no banco de dados. A distribuição dos melhores *hits* (*Top-Hit*) dos unigenes de *M. olfersii* comparadas às sequências de todas as espécies no banco de dados (Figura 12) indicou que o maior número de unigenes (13,5%) foi relacionado com sequências da termita *Zootermopsis nevadensis*. Seguido pelo crustáceo *D. pulex* (7,3%), a aranha *Stegodyphus mimosarum* (3,9%), o inseto *T. castaneum* (3,8%) entre outros. Das 28 espécies melhor classificadas, quatro são decápodes malacostracos de um total de cinco táxons (17,9%) de crustáceos (*D. pulex*, *P. monodon*, *L. vannamei*, *Marsopenaeus japonicus*, *Scylla paramamosain*). Além disso, outros artrópodes representam 46,4% (13 de 28) das espécies. *D. melanogaster* não foi relacionada dentro das 28 espécies com melhores *hits*.

Os domínios/famílias de proteínas conservadas foram identificados nos unigenes de *M. olfersii* usando ao banco de dados do InterPro como referência. Do total, 56.731 unigenes foram classificadas em 6.259 domínios/famílias. A maioria das categorias (4.594)

continham de 1 a 5 unigenes que apareceram com mais frequência. Os domínios de proteínas foram classificados de acordo ao número de unigenes, e os 60 domínios/famílias mais abundantes são mostrados na Tabela 3. A categoria de domínios de proteínas dedos de zinco tipo C2H2 foi a mais abundante (1.961 unigenes), outras representadas foram o domínio de proteína kinase (407) e domínio homeobox, entre outros.



**Figura 12.** Distribuição dos melhores hits (*Top Hits*) dos unigenes de *M. olfersii* contra as seqüências de todas as espécies do banco de dados de proteínas NR. O algoritmo Blastx ( $E\text{-value} < 1e^{-5}$ ) foi utilizado para busca de similaridade com seqüências de proteínas NR do NCBI. As espécies com melhores hits e o número de unigenes homólogos são mostradas. Das 28 espécies melhor classificadas, quatro são decápodes malacostracos de um total de cinco táxons de crustáceos (17,9%). Outros correspondem à soma de unigenes com hits a proteínas em outras espécies.

**Tabela 3.** Os 60 domínios/famílias de proteínas conservados do InterPro mais abundantes identificadas nos unigenes de *M. olearii*

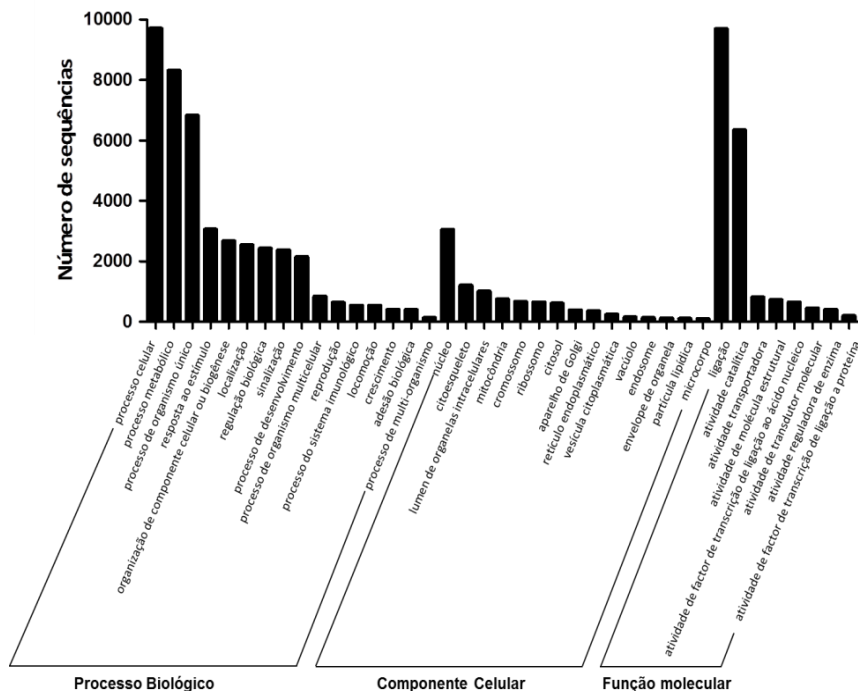
<b>Número de Acesso</b>	<b>Domínio/Família conservada</b>	<b>Número de unigenes</b>
IPR007087	Zinc finger, C2H2	1,961
IPR013087	Zinc finger C2H2-type/integrase DNA-binding domain	1,779
IPR015880	Zinc finger, C2H2-like	1,716
IPR027417	P-loop containing nucleoside triphosphate hydrolase	1,014
IPR011009	Protein kinase-like domain	471
IPR012677	Nucleotide-binding, alpha-beta plait	444
IPR013783	Immunoglobulin-like fold	429
IPR015943	WD40/YVTN repeat-like-containing domain	423
IPR000719	Protein kinase domain	407
IPR017986	WD40-repeat-containing domain	382
IPR016024	Armadillo-type fold	377
IPR000504	RNA recognition motif domain	375
IPR013083	Zinc finger, RING/FYVE/PHD-type	349
IPR001680	WD40 repeat	332
IPR011333	BTB/POZ fold	302
IPR016040	NAD(P)-binding domain	296
IPR013069	BTB/POZ	289
IPR000210	BTB/POZ-like	288
IPR009057	Homeodomain-like	283
IPR011989	Armadillo-like helical	279
IPR002290	Serine/threonine/dual specificity protein kinase, catalytic domain	269
IPR011990	Tetratricopeptide-like helical domain	261
IPR011993	Pleckstrin homology-like domain	252
IPR007110	Immunoglobulin-like domain	248
IPR008271	Serine/threonine-protein kinase, active site	246
IPR017441	Protein kinase, ATP binding site	246
IPR020683	Ankyrin repeat-containing domain	233
IPR002110	Ankyrin repeat	220
IPR011991	Winged helix-turn-helix DNA-binding domain	191
IPR001841	Zinc finger, RING-type	187
IPR029071	Ubiquitin-related domain	184
IPR012337	Ribonuclease H-like domain	181
IPR016196	Major facilitator superfamily domain, general substrate transporter	181
IPR000477	Reverse transcriptase domain	177
IPR029058	Alpha/Beta hydrolase fold	176
IPR001478	PDZ domain	174
IPR011992	EF-hand domain pair	173
IPR003599	Immunoglobulin subtype	169
IPR012336	Thioredoxin-like fold	169
IPR001650	Helicase, C-terminal	160
IPR013098	Immunoglobulin I-set	160
IPR019775	WD40 repeat, conserved site	158
IPR014001	Helicase, superfamily 1/2, ATP-binding domain	156
IPR000742	Epidermal growth factor-like domain	154
IPR001452	SH3 domain	153
IPR013026	Tetratricopeptide repeat-containing domain	152
IPR003598	Immunoglobulin subtype 2	150
IPR029063	S-adenosyl-L-methionine-dependent methyltransferase-like	147
IPR001849	Pleckstrin homology domain	146
IPR018247	EF-Hand 1, calcium-binding site	144
IPR003961	Fibronectin, type III	144
IPR002048	EF-hand domain	143
IPR013032	EGF-like, conserved site	142
IPR019734	Tetratricopeptide repeat	142
IPR011011	Zinc finger, FYVE/PHD-type	132
IPR017853	Glycoside hydrolase, superfamily	131
IPR008985	Concanavalin A-like lectin/glucanases superfamily	119
IPR001611	Leucine-rich repeat	116
IPR001356	Homeobox domain	112
IPR003593	AAA+ ATPase domain	111

### 3.3. CLASSIFICAÇÃO DE UNIGENES POR ONTOLOGIA GÊNICA (GO)

Do total de sequências anotadas, 21.845 unigenes (21,90% do total) foram classificados por GO utilizando o programa Blast2GO. Dentre desses, 13.841 unigenes (63,36%) foram classificados na categoria de processo biológico, 10.254 (46,94%) unigenes na categoria de componente celular e 19.604 (89,74%) unigenes na categoria de função molecular. Na categoria de processo biológico, as subcategorias mais representadas foram processos metabólicos (20,03%), processos celulares (13,94%) e processos de organismo único (10,85%). Outras subcategorias importantes incluíram processos de desenvolvimento (6,0%), respostas a estímulos (3,52%), localização (3,28%), organização de componente celular ou biogênese (2,99%), sinalização (2,11%), locomoção (1,74%) e processos do sistema imune (1,23%). Na categoria de componentes celulares, as subcategorias mais representativas foram núcleo (32,38%), citoesqueleto (12,76%), lúmen de organelas intracelulares (10,56%) e mitocôndria (7,9%). Na categoria de processos biológicos, as subcategorias mais representativas foram de ligação (50,37%) e atividade catalítica (32,97%) (Figura 13).

Também, foram identificados unigenes de processos celulares nos estágios embrionários de *M. olfersii* como: ciclo celular (908 unigenes), morte celular (548 unigenes), divisão celular (480 unigenes), proliferação celular (412 unigenes), entre outros.

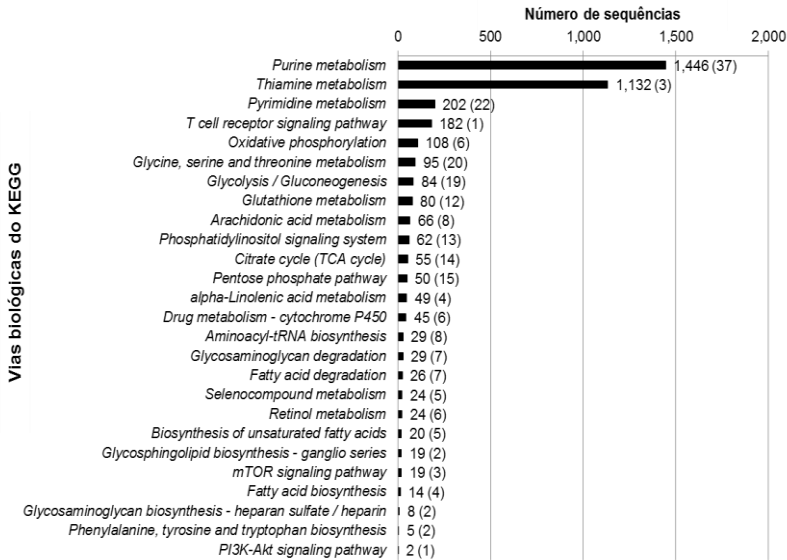




**Figura 13. Histograma da classificação GO de unigenes anotadas do transcriptoma de *M. olfersii*.** Um maior número de unigenes foi obtido para processo celular, núcleo e proteínas de ligação em cada categoria.

### 3.4. CLASSIFICAÇÃO DE UNIGENES EM VIAS BIOLÓGICAS DO KEGG

Para classificar os unigenes em vias biológicas, estes foram mapeados com as vias do banco de dados de KEGG, utilizando Blast2GO. Um total de 6.866 unigenes foi designado em 129 vias do KEGG. Entre essas vias, as mais representativas foram: metabolismo de purina (1.446 unigenes) e metabolismos de tiamina (1.132 unigenes). Outras vias classificadas foram fosforilação oxidativa (108 unigenes), glicólise/gliconeogênese (84 unigenes), ciclo do citrato (55 unigenes) e metabolismo de fármacos - citocromo P450 (45 unigenes) (Figura 14).



**Figura 14.** Unigenes de *M. ofersii* classificados em vias biológicas do KEGG. O número de enzimas em cada via é indicado entre parênteses.

### 3.5. IDENTIFICAÇÃO DE TRANSCRITOS DE GENES DO GENOMA MITOCONDRIAL

No presente estudo foram identificados os transcritos de 22 tRNA, 2 rRNA (12S rRNA e 16S rRNA) e de 13 genes do mtDNA codificantes de proteínas (*cytB*, *cox1*, *cox2*, *cox3*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *atp6* e *atp8*) do genoma mitocondrial de *M. ofersii*. A similaridade, utilizando o algoritmo Blastp ou tBlastn, destas seqüências com outros genes mitocondriais do banco de dados *GenBank* são mostrados na Tabela 4 e 5. Considerando como referência o genoma mitocondrial de *M. rosenbergii* (15.772 pb), foram identificados 68,9% do genoma mitocondrial de *M. ofersii* (Figura 15), aproximadamente 10.866 nt obtido a partir das seqüências do transcriptoma de *M. ofersii*.

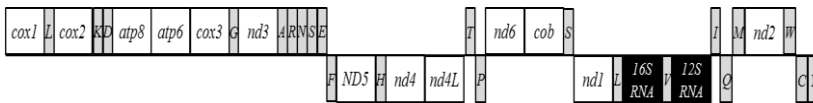
**Tabela 4.** Similaridade dos transcritos de tRNA do genoma mitocondrial de *M. olfersii* com as seqüências do banco de dados *GenBank*.

Gene	Tamanho (nt)	Identidade	E-value	Anotação dos genes	Número de acesso
<i>12S rRNA</i>	541	89%	0.0	Macrobrachium bullatum strain MAR41 mitochondrion, complete genome	KM978918.1
<i>16S rRNA</i>	101	100%	5.00E-43	Macrobrachium olfersii 16S ribosomal RNA gene, partial sequence; mitochondrial	GU929466.1
<i>tRNA-Leu</i>	64	95%	4.00E-22	tRNA-Leu gii61651658:1536-1599 Macrobrachium rosenbergii mitochondrion, complete genome	NC_006880.1
<i>tRNA-Lys</i>	68	98%	3.00E-29	tRNA-Lys gii61651658:2290-2357 Macrobrachium rosenbergii mitochondrion, complete genome	NC_006880.1
<i>tRNA-Asp</i>	66	86%	1.00E-06	tRNA-Asp gii61651658:2358-2423 Macrobrachium rosenbergii mitochondrion, complete genome	NC_006880.1
<i>tRNA-Gly</i>	65	96%	2.00E-23	tRNA-Gly gii61651658:4050-4114 Macrobrachium rosenbergii mitochondrion, complete genome	NC_006880.1
<i>tRNA-Ala</i>	63	93%	4.00E-19	tRNA-Ala gii61651658:4469-4531 Macrobrachium rosenbergii mitochondrion, complete genome	NC_006880.1
<i>tRNA-Arg</i>	62	96%	2.00E-08	tRNA-Arg gii61651658:4531-4592 Macrobrachium rosenbergii mitochondrion, complete genome	NC_006880.1
<i>tRNA-Asn</i>	65	100%	7.00E-30	tRNA-Asn gii61651658:4594-4658 Macrobrachium rosenbergii mitochondrion, complete genome	NC_006880.1
<i>tRNA-Ser</i>	67	92%	4.00E-19	tRNA-Ser gii61651658:4659-4725 Macrobrachium rosenbergii mitochondrion, complete genome	NC_006880.1
<i>tRNA-Glu</i>	68	97%	4.00E-25	tRNA-Glu gii61651658:4728-4796 Macrobrachium rosenbergii mitochondrion, complete genome	NC_006880.1
<i>tRNA-Phe</i>	66	85%	3.00E-07	tRNA-Phe gii61651658:4796-4861 Macrobrachium rosenbergii mitochondrion, complete genome	NC_006880.1
<i>tRNA-His</i>	64	94%	6.00E-18	tRNA-His gii61651658:6587-6650 Macrobrachium rosenbergii mitochondrion, complete genome	NC_006880.1
<i>tRNA-Thr</i>	64	87%	2.00E-08	tRNA-Thr gii61651658:8283-8347 Macrobrachium rosenbergii mitochondrion, complete genome	NC_006880.1
<i>tRNA-Pro</i>	66	90%	4.00E-16	tRNA-Pro gii61651658:8347-8412 Macrobrachium rosenbergii mitochondrion, complete genome	NC_006880.1
<i>tRNA-Ser</i>	69	92%	3.00E-20	tRNA-Ser gii61651658:10061-10129 Macrobrachium rosenbergii mitochondrion, complete genome	NC_006880.1
<i>tRNA-Leu</i>	64	100%	2.00E-24	tRNA-Leu gii61651658:11211-11184 Macrobrachium rosenbergii mitochondrion, complete genome	NC_006880.1
<i>tRNA-Val</i>	66	93%	4.00E-13	tRNA-Val gii61651658:12490-12556 Macrobrachium rosenbergii mitochondrion, complete genome	NC_006880.1
<i>tRNA-Ile</i>	67	95%	6.00E-24	tRNA-Ile gii61651658:14340-14406 Macrobrachium rosenbergii mitochondrion, complete genome	NC_006880.1
<i>tRNA-Gln</i>	68	95%	2.00E-24	tRNA-Gln gii61651658:14435-14502 Macrobrachium rosenbergii mitochondrion, complete genome	NC_006880.1
<i>tRNA-Met</i>	68	94%	4.00E-22	tRNA-Met gii61651658:14514-14581 Macrobrachium rosenbergii mitochondrion, complete genome	NC_006880.1
<i>tRNA-Trp</i>	69	94%	1.00E-22	tRNA-Trp gii61651658:15576-15644 Macrobrachium rosenbergii mitochondrion, complete genome	NC_006880.1
<i>tRNA-Cys</i>	63	86%	4E-09	tRNA-Cys gii61651658:15646-15709 Macrobrachium rosenbergii mitochondrion, complete genome	NC_006880.1
<i>tRNA-Tyr</i>	63	91%	5.00E-06	tRNA-Tyr gii61651658:15710-15772 Macrobrachium rosenbergii mitochondrion, complete genome	NC_006880.1

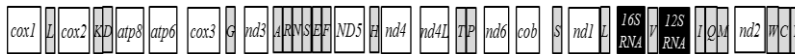
**Tabela 5.** Similaridade dos transcritos de genes do genoma mitocondrial de *M. olfersii* com as sequências do banco de dados *GenBank*.

Gene	Tamanho (aa)	Similaridade	E-value	Anotação dos genes	Número de acesso
<i>cox1</i>	70	100%	5.00E-52	cytochrome oxidase subunit I [Macrobrachium striatum]	AFH66770.1
<i>cox2</i>	94	98%	2.00E-44	cytochrome c oxidase subunit II [Macrobrachium nipponense]	YP_004221771.1
<i>atp8</i>	49	89%	2.00E-22	ATP synthase subunit 8 [Macrobrachium rosenbergii]	AFJ69913.1
<i>atp6</i>	72	98%	2.00E-35	ATP synthase subunit 6 [Macrobrachium rosenbergii]	AFJ69914.1
<i>cox3</i>	99	98%	2.00E-56	cytochrome c oxidase subunit III [Macrobrachium nipponense]	YP_004221774.1
<i>nd3</i>	117	95%	2.00E-58	NADH dehydrogenase subunit 3 [Macrobrachium nipponense]	YP_004221775.1
<i>nd5</i>	259	96%	1.00E-163	NADH dehydrogenase subunit 5 [Macrobrachium rosenbergii]	YP_214002.1
<i>nd4</i>	444	97%	0.0	NADH dehydrogenase subunit 4 [Macrobrachium nipponense]	YP_004221777.1
<i>nd4L</i>	136	96%	5.00E-60	NADH dehydrogenase subunit 4L [Macrobrachium rosenbergii]	YP_214004.1
<i>nd6</i>	162	91%	2.00E-82	NADH dehydrogenase subunit 6 [Macrobrachium lancesteri]	YP_002650702.1
<i>cytB</i>	237	100%	4.00E-147	cytochrome b [Macrobrachium nipponense]	YP_004221780.1
<i>nd1</i>	348	97%	0.0	NADH dehydrogenase subunit 1 [Macrobrachium nipponense]	YP_004221781.1
<i>nd2</i>	269	92%	1.00E-148	NADH dehydrogenase subunit 2 [Macrobrachium rosenbergii]	YP_214008.1

Pancrustacea (*N. denticulata*, *M. lancesteri*, *M. nipponense*, *M. rosenbergii*, *Rimicaris krei*)



*M. olfersii* (68,9% do genoma mitocondrial)



**Figura 15.** Transcritos do genoma mitocondrial de *M. olfersii* e a organização de genoma mitocondrial de Pancrustacea. As caixas de cor branca indicam os genes que codificam proteínas. As caixas de cor cinza indicam os genes de tRNA. As caixas da cor preta indicam os genes de rRNA. Os genes de tRNA são designados pelo código de uma letra para cada aminoácido específico. As caixas de acima e abaixo na representação do genoma mitocondrial de Pancrustacea indica a localização dos genes em uma ou outra fita do genoma. A localização dos genes no genoma de *M. olfersii* não é conhecida.

## 4. DISCUSSÃO

Nos últimos anos, uma grande quantidade de dados de seqüências de alguns crustáceos tem sido obtida por métodos de sequenciamento de nova geração, especialmente utilizando a plataforma Illumina, facilitado por seu baixo custo e alta eficiência (CLARK et al., 2011; MA et al., 2012; SOOKRUKSAWONG et al., 2013; ZENG et al., 2013; MENG et al., 2015). No entanto, poucas informações de seqüências estão disponíveis em crustáceos de água doce, como por exemplo do camarão *M. olfersii*, um novo modelo para estudos de desenvolvimento embrionário. A compreensão das bases moleculares dos eventos na morfogênese e organogênese que ocorrem nas idades embrionárias de E4 e E8 nessa espécie precisam ser investigadas. Neste contexto, no presente estudo, se realizou o sequenciamento na plataforma Illumina *paired-end*, a montagem de *novo* com múltiplos *k-mer* e a análise do transcriptoma de um *pool* de RNA de *M. olfersii*. A estratégia de uma mistura de quantidades equimolares de RNA extraídos dos estágios foi escolhida visto que a utilização de um mesmo número de ovos para estágios mais iniciais (E4-E5) poderiam apresentar transcritos com baixa representação no transcriptoma. Essa estratégia de mistura de RNA é a mais utilizada em estudos transcriptômicos (DAINES et al., 2011; LEGEI et al., 2014; JAYASWALL et al., 2016). No entanto, na análise transcriptômica de *P. hawaiiensis* foi utilizada como estratégia um maior número de ovos de estágios mais iniciais para ter mais chance de identificar transcritos nesses estágios (ZENG et al., 2011).

Neste estudo, 457.896 *contigs* e 99.751 unigenes de embriões de *M. olfersii* foram gerados a partir de 25,6 milhões de *reads* de alta qualidade. Do total de unigenes, 20.893 (20,95%) mostraram *hits* significativos no Blastx para proteínas conhecidas. Essa quantidade de unigenes, com similaridade a proteínas conhecidas no Blastx deve ser tomado com precaução, porque pode subestimar ou superestimar o número de transcritos verdadeiros expressos nesses estágios embrionários de *M. olfersii*. Isso porque, vários unigenes podem representar diferentes regiões do mesmo gene, mas apresentar diferentes *hits* no Blastx e, assim serem identificados como genes diferentes. Conclusões similares foram obtidas na análise transcriptômica de *P. hawaiiensis* (ZENG et al., 2011). Somado a isso, as seqüências, não relacionadas com proteínas no banco de dados, poderiam representar fragmentos de transcritos cuja similaridade aos genes conhecidos é baixa para atingir o valor de *cutoff* de *E-value*, ou ser seqüências não

codificantes. Portanto, 79% dos unigenes de *M. olfersii*, onde não foi encontrada similaridade utilizando o Blastx, poderiam ser genes específicos desta espécie, ou genes com baixa similaridade a genes conhecidos, ou ainda sequências não-codificantes (por exemplo, RNAs ou regiões não traduzidas (UTR, *untranslated region*)). Resultados similares também foram obtidos na análise transcriptômica de *P. hawaiiensis*, onde aproximadamente 70% das sequências de transcritos (*isotigs*) não teve similaridade com proteínas no banco de dados de sequências (ZENG et al., 2011). Neste contexto, o número total dos transcritos de *M. olfersii* não pode ser determinado a partir dos dados do presente estudo, pois outros transcritos podem estar presentes em estágios larvais, juvenis ou adultos. No entanto, o presente transcriptoma proporciona informação valiosa de genes de diferentes processos celulares essenciais da espécie, assim como dos estágios E4-E8.

Os transcritos de genes obtidos no presente estudo foram mais similares a genes de outros crustáceos e insetos. A distribuição das 28 espécies com melhores *hits* no Blastx para os unigenes do transcriptoma de *M. olfersii* indicou que os artrópodes representaram 67,8% (19 de 28) do total de espécies: 12 insetos, 2 quelicerados e 5 crustáceos. A maioria de unigenes (3.960) foi relacionada com genes da térmita *Z. nevadensis*, cujo genoma está disponível (TERRAPON et al., 2014). Ainda, 2.139 unigenes de *M. olfersii* foram relacionadas com genes de *D. pulex*, que também possui genoma disponível (COLBOURNE et al., 2011). Provavelmente, a alta similaridade com genes de *Z. nevadensis* seja devido à ausência de genomas ou de transcriptomas de outros crustáceos mais estreitamente relacionados a *M. olfersii*. Apesar das análises transcriptômicas de alguns crustáceos terem sido realizadas, como por exemplo, *M. rosenbergii* (RAO et al., 2015) e *M. nipponense* (MA et al., 2012), estas sequências não foram consideradas no presente estudo porque unicamente seus *reads* são disponíveis no banco de dados *GenBank* e não para uso no Blastx. Portanto, a baixa representação de sequências de crustáceos no banco de dados *GenBank* dificulta na obtenção de melhores *hits* no Blastx de alta confiança para este grupo (ZENG et al., 2011). No entanto, proporções similares de espécies de crustáceos e outros artrópodes foram obtidas no transcriptoma de *C. finmarchicus* (LENZ et al., 2014). No presente estudo, a sexta espécie com melhor *hit* foi o cefalocordado *Branchiostoma floridae*, espécie que também foi relacionada no transcriptoma do crustáceo anfípode *P. hawaiiensis* (ZENG et al., 2011) e de *C. finmarchicus* (LENZ et al., 2014). Embora *B. floridae* é filogeneticamente distante dos crustáceos, este organismo apresenta sequências similares, sendo que a maioria

delas poderia ser sequências de proteínas, contendo o domínio de dedo de zinco C2H2, como identificado em *P. hawaiiensis* (ZENG et al., 2011).

Nos estágios do desenvolvimento embrionário de *M. olfersii* estudados, uma grande diversidade de unigenes foi identificada e anotada funcionalmente. Foram identificados por GO os unigenes relacionados a processos celulares, metabólicos, de crescimento, de desenvolvimento e resposta a estímulos, assim como unigenes das diferentes organelas celulares e unigenes relacionados a proteínas com atividade catalítica ou fatores de transcrição. Resultados similares de grupos de GO funcionais foram encontrados no transcriptoma de embriões de *C. finmarchicus* (LENZ et al., 2014). Dentre as organelas, a mitocôndria é importante para estudos de estresse oxidativo e de fosforilação oxidativa. Estudos em embriões de *M. olfersii* expostos a radiação UVB demonstraram que a mitocôndria é um alvo importante, provocando mudanças ultra-estruturais e modulação de proteínas de fusão e fissão mitocondrial (DE QUADROS et al., 2016). Além disso, esses embriões expostos à UVB apresentaram diminuição da proliferação celular (ZENI et al., 2015) e aumento da apoptose (SCHRAMM, 2015). Neste estudo, foram identificados 743 unigenes da mitocôndria de *M. olfersii*, além de unigenes de processos celulares como sendo do ciclo celular (908 unigenes), morte celular (548 unigenes), divisão celular (480 unigenes), proliferação celular (412 unigenes), entre outros. Neste contexto, a identificação destes genes permitirá estudar as bases moleculares do desenvolvimento embrionário, assim como também os efeitos de outros estressores em embriões e em organismos adultos de *M. olfersii*.

Uma diversidade de vias metabólicas (129) do KEGG foi identificada para 6.866 unigenes de *M. olfersii*, o que demonstra uma grande representação de transcritos e vias metabólicas presentes nestes estágios embrionários. No entanto, a validação funcional destes unigenes deve ser realizada no futuro. No presente estudo, foram identificados alguns transcritos de genes relacionados com o metabolismo oxidativo primário (glicólise, ciclo de Krebs e fosforilação oxidativa). Além disso, foram identificados outros unigenes envolvidos na glicogênese, degradação de glicogênio, gluconeogênese, sínteses de ácidos graxos e  $\beta$ -oxidação de ácidos graxos. Essa informação será de utilidade para futuros estudos fisiológicos e moleculares nesta espécie, como por exemplo, caracterizar as respostas dessas vias frente a estressores físicos ou químicos. No crustáceo *E. superba*, a partir da análise transcriptômica, também foi possível identificar genes do

metabolismo oxidativo e outros relacionados ao metabolismo de carboidratos e ácidos graxos (MEYER et al., 2015). Os transcritos de genes de sistemas de detoxificação também foram reconhecidos em embriões de *M. olfersii*. Nesse estudo, foram identificados unigenes do sistema de detoxificação de contaminantes associados a citocromo P450 (39 unigenes), de peroxissomos (90 unigenes), de lisossomos (73 unigenes) e de respostas a estresse (1251 unigenes). Os peroxissomos são organelas essenciais que têm função chave na sinalização redox e na homeostase de lipídeos (NORDGREN; FRANSEN, 2014). Os transcritos destes genes poderia permitir a sobrevivência dos animais aquáticos em ambientes expostos a contaminantes como metais pesados, pesticidas e outros químicos. Genes de metabolismo de xenobiontes e genes do peroxissomo foram identificados no hepatopâncreas da lagosta *P. clarkii* (SHEN et al., 2014).

Em mamíferos, sabe-se que a transcrição mitocondrial se inicia no D-loop e é regulada por proteínas codificadas pelo núcleo. A partir de ambas as fitas do DNA mitocondrial são transcritos mRNA policistrônicos precursores longos e que, posteriormente, são cortados os inter-espacos de tRNA liberando rRNAs e mRNAs (OJALA; MONTOYA; ATTARDI, 1981). O RNA liberado sofre a poliadenilação da extremidade 3' do mRNA e rRNA. Em seguida, mudanças de nucleotídeos específicos e adição de trinucleotídeos CCA são realizados na extremidade 3' do tRNA (NAGAIKE et al., 2005). Embora o processo de transcrição mitocondrial ainda não é conhecido em crustáceos, no presente transcriptoma foram identificados os transcritos dos 22 tRNA, 2 rRNA e 13 genes codificados pelo DNA mitocondrial, similares a sequências de outras espécies de *Macrobrachium*. Anteriormente, genes próprios da mitocôndria têm sido recuperados a partir de transcriptomas na alga verde *Polytomella* spp. (TIAN; SMITH, 2016). Esta recuperação de mRNA do DNA mitocondrial a partir de transcriptoma, demonstra que estes mRNA também podem ser poliadenilados depois de serem transcritos. No gênero *Macrobrachium*, o tamanho do genoma mitocondrial foi determinado para *M. nipponense*, *M. rosenbergii*, *M. lanchesteri*, (15.806 pb, 15.772 pb e 15.694 pb, respectivamente) (MILLER et al., 2005) (MA et al., 2011). Os arranjos dos genes mitocondriais dessas espécies são idênticos ao padrão Pancrustacea (um clado composto por crustáceos e hexápodes) (SHEN et al., 2012). No entanto, ainda não eram conhecidos os genes do genoma mitocondrial de *M. olfersii*. A informação parcial do genoma mitocondrial dessa espécie obtida no presente estudo permitirá realizar estudos evolutivos e relações filogenéticas, bem como, no futuro, ajudar



no desenho de iniciadores específicos e sequenciamento do genoma mitocondrial de *M. olfersii*.

Com relação aos unigenes classificados por InterPRO, o maior número foi o domínio das proteínas dedo de zinco tipo C2H2. Uma alta proporção desses transcritos, contendo o domínio dedo de zinco tipo C2H2, também foi descrito no transcriptoma de *P. hawaiiensis* (ZENG et al., 2011) e no genoma do pulgão da ervilha *Acyrtosiphon pisum* (SHIGENOBU et al., 2010). Na maioria dos genomas de animais que tem sido sequenciado, o domínio de dedo de zinco tipo C2H2 está entre os mais abundantes. Além disso, os domínios de dedo de zinco são de interesse particular na biologia de desenvolvimento porque eles são predominantes em algumas proteínas reguladoras transcricionais (MATERNA et al., 2006). Outro domínio classificado foi o homeobox ou homeodomínio, que estão presentes em fatores de transcrição que regulam o desenvolvimento animal. Os genes homeobox codificam proteínas de ligação ao DNA regulando a expressão gênica e controlando vários aspectos da morfogênese e diferenciação animal (MARK; RIJLI; CHAMBON, 1997).

Em resumo, esta é a primeira descrição e análise transcriptômica descrita de *M. olfersii* e o primeiro trabalho que gera um banco de sequências desta espécie. Um total de 99.751 unigenes foram obtidos e anotados por Blastx, classificados nas categorias de GO e em vias metabólicas usando o banco de dados KEGG. Além disso, considerando as características biológicas e do desenvolvimento de *M. olfersii* e seu uso como um novo modelo para estudos de biologia do desenvolvimento, a disponibilidade dessas sequências permitirão delinear futuras pesquisas para um melhor entendimento das bases moleculares dos processos celulares dessa espécie, em especial os eventos durante o desenvolvimento.



## **CAPÍTULO II. Identificação de transcritos de genes e vias relacionados ao desenvolvimento embrionário de *M. olfersii***

---

### **1. INTRODUÇÃO**

O desenvolvimento de um organismo multicelular requer mudanças na expressão de genes de vários tipos celulares (SALAZAR-CIUDAD, 2003) e os diferentes eventos são coordenados por programas genéticos específicos. Entre estes eventos destacam-se a morfogênese e a organogênese, que são coordenados por processos celulares como a proliferação, diferenciação, migração, apoptose, adesão e mudanças na forma das células (BASSON, 2012). Uma consequência das mudanças ocorridas nas redes gênicas que coordenam esses eventos, incluindo sua expressão em diferentes grupos animais, pode estar relacionada com o processo de evolução do plano do corpo (PETER; DAVIDSON, 2011). Assim, a identificação de genes relacionados ao desenvolvimento embrionário, bem como a avaliação da função destes genes, pode ajudar a entender os mecanismos moleculares envolvidos no desenvolvimento embrionário das espécies.

Os artrópodes têm se tornado um grupo fascinante para os estudos de compreensão da função de diferentes genes relacionados ao desenvolvimento. Entre eles, *D. melanogaster*, que apresenta embriões de banda longa, tem servido como modelo para compreender a base molecular dos mecanismos de padronização do desenvolvimento (JENNINGS, 2011). No entanto, algumas espécies dentro dos artrópodes apresentam embriões de banda curta ou intermediária (DAVIS; PATEL, 2002). Estudos no inseto *T. castaneum*, que apresenta embriões de banda curta, demonstram que os mecanismos moleculares gerais do desenvolvimento são similares aos embriões de banda longa, no entanto, existem certas particularidades na função de alguns genes (BROWN; DENELL, 1996).

Desde a identificação dos genes relacionados ao desenvolvimento em *D. melanogaster* foram realizados diferentes estudos para identificar homólogos a esses genes em outros artrópodes. Esses estudos, em sua maioria, foram dirigidos para a identificação dos genes *Hox* em diversos grupos de artrópodes e outros animais (HUGHES; KAUFMAN, 2002), demonstrando que os animais com simetria bilateral compartilham um número de genes, que constituem um *kit* de ferramentas importantes no desenvolvimento embrionário (DEUTSCH; MOUCHEL-VIELH, 2003). Outros genes, *Engrailed* (*En*)

*Even skipped (Eve)*, *Wingless (Wg)* e *Hedgehog (Hh)*, relacionados com a segmentação do corpo, também têm sido identificados em outros artrópodes (PEEL; TELFORD; AKAM, 2006; COLBOURNE et al., 2011; LYNCH; EL-SHERIF; BROWN, 2012). Recentemente, os genes da família *Wnt* estão sendo estudados em relação a sua função no desenvolvimento (BOLOGNESI et al., 2008; JANSSEN et al., 2010)

Como mencionado anteriormente, homólogos desses genes foram identificados em alguns crustáceos: *Artemia* (AVEROF; AKAM, 1993), *P. clarkii*, *P. scaber* (ABZHANOV; KAUFMAN, 1999a, 2000b), *P. hawaiiensis* (LIUBICICH et al., 2009), *D. pulex* (JANSSEN et al., 2010) entre outros. No entanto, a informação dos processos moleculares da segmentação em crustáceos ainda é bastante fragmentada, pois há pouca informação sobre os genes deste grupo. Além disso, as funções desses genes durante o desenvolvimento embrionário ainda não são claras (ERIKSSON; UNGERER; STOLLEWERK, 2013). Destaca-se, entre os camarões de água doce, um dos potenciais modelos de estudo de desenvolvimento, *M. olfersii*, cuja informação de transcritos de genes foi recentemente descrita (Capítulo I).

*M. olfersii* apresenta embriões de banda curta, com desenvolvimento similar a alguns crustáceos também de banda curta. Nesses embriões, após a gastrulação ocorre a formação da região cefálica (lobos óticos), seguida por pares de apêndices naupliares e uma região de crescimento posterior indiferenciada (região pós-naupliar) que originará a papila caudal e os apêndices pós-naupliares. No crescimento da papila caudal, cada conjunto de quatro fileiras de células agrupadas transversalmente formará o esboço de um segmento corporal (SCHOLTZ, 1992). Portanto, genes envolvidos na formação dos segmentos corporais e de identidade segmentar em *M. olfersii*, devem estar presentes durante o desenvolvimento embrionário, sendo relevante reconhecê-los neste camarão. Além disso, as vias de sinalização relacionadas ao desenvolvimento embrionário têm sido identificadas em diferentes animais. No entanto, surpreendentemente, os estudos bioquímicos e genéticos demonstram que poucas vias de sinalização (vias Hedgehog, Wnt, *transforming growth factor-beta* (TGF- $\beta$ ), *receptor tyrosine kinase*, Notch, Janus kinase (JAK)/*signal transducer and activator of transcription* (STAT) e vias de hormônios nucleares) são suficientes para gerar uma ampla variedade de células e tecidos durante o desenvolvimento (PIRES-DA SILVA; SOMMER, 2003). Outros estudos demonstram que, os transcritos dos genes componentes dessas vias, durante o desenvolvimento, também estão presentes em crustáceos. No estudo transcriptômico de *P. hawaiiensis* e *L. vannamei*,

os transcritos de genes de vias relacionadas ao desenvolvimento (MAPK, TGF-beta, Wnt, Notch, VEGF, JAK/STAT, Hedgehog) foram identificados (ZENG et al., 2011; WEI et al., 2014).

Para a identificação de genes e seus transcritos em animais, dois métodos têm sido os mais utilizados: a) estratégia baseada em PCR ou RT-PCR e b) sequenciamento de genomas ou de transcriptomas. Em crustáceos, ambos os métodos têm sido aplicados na identificação de genes relacionados ao desenvolvimento. Por exemplo, utilizando a estratégia baseada em PCR foram identificadas as sequências dos genes *Hox* (*Antp*, *Ubx* e *AbdA*) (ABZHANOV; KAUFMAN, 2000a,c) e genes de segmentação (*Engrailed*) (ABZHANOV; KAUFMAN, 2000c; GIBERT et al., 2000) entre outros. Enquanto que, o sequenciamento do genoma de *N. denticulata* (KENNY et al., 2014) permitiu identificar as sequências dos genes *Hox* (*Lab*, *Dfd*, *Hox3/Zen1*, *Scr*, *Ftz*, *Antp*, *Ubx*, *AbdA* e *AbdB*). O sequenciamento do transcriptoma de *L. vannamei* (LI, C et al., 2012) permitiu identificar alguns genes relacionados ao seu desenvolvimento (*AbdA*, *AbdB*, *Antp*, *Wnt6*, *Wnt10*, entre outros).

Sendo assim, o objetivo deste trabalho foi identificar transcritos dos genes relacionados ao desenvolvimento embrionário e de vias de sinalização envolvidas no desenvolvimento no camarão *M. olfersii*. Neste contexto, foi utilizada a estratégia de RT-PCR com iniciadores degenerados e a busca de transcritos destes genes nas sequências obtidas da análise transcriptômica (Capítulo I).

## 2. MATERIAIS E MÉTODOS

### 2.1. IDENTIFICAÇÃO DE TRANSCRITOS DE GENES RELACIONADOS AO DESENVOLVIMENTO EMBRIONÁRIO POR ANÁLISE TRANSCRIPTÔMICA

Para a busca de transcritos de genes específicos relacionados ao desenvolvimento embrionário e vias de desenvolvimento, foram utilizadas sequências de genes ou de transcritos de outros artrópodes bem como o algoritmo tBlastn nos unigenes de *M. olfersii* (Capítulo I). Foram utilizados transcritos dos 13 genes de efeito materno, 40 genes *gap* e genes *pair-rule* que afetam o padrão de segmentação (<http://www.ncbi.nlm.nih.gov/books/NBK10039/>), 10 genes *Hox* de artrópodes, 12 genes da família *Wnt* de *D. pulex*, 6 genes do padrão dorso-ventral (DV) de artrópodes, vias de desenvolvimento (25 genes da via TGF- $\beta$ , 45 genes da via Wnt, 21 genes da via Notch, 17 genes da via MAPK, 15 genes da via Hedgehog, 17 genes da via Jak-STAT, 26 genes

da via VEGF) do banco de dados KEGG e 3 genes de receptores nucleares induzíveis por ecdisteróide de *D. melanogaster* e *D. pulex*. Depois desta etapa, a seleção da fase aberta de leitura (*ORF*) dos unigenes foi realizada com o programa *ORF Finder* (<https://www.ncbi.nlm.nih.gov/orffinder/>). Em seguida, uma busca de similaridade usando o algoritmo Blastp (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>), contra o banco de dados não redundante (NR) foi realizada para cada *ORF* selecionada. Para determinar isoformas e parálogos de genes *Hox* e de segmentação foi realizado um alinhamento múltiplo de sequências de aminoácidos destas *ORFs*. Para determinar a abundância de transcritos de genes relacionados ao desenvolvimento embrionário, os *reads* de alta qualidade foram mapeados aos unigenes utilizando o software Bowtie (v.0.12.7) (LANGMEAD et al., 2009). Os *reads* mapeados de cada unigene foram contados utilizando SAM tools v.0.1.16 (LI et al., 2009).

A similaridade dos transcritos de genes relacionados ao desenvolvimento em *M. olfersii* com outras sequências de outras espécies foi determinada usando as sequências da classificação de ontologia gênica da análise transcriptômica (Capítulo I). Os unigenes classificados nas categorias de processos de desenvolvimento (2.142 unigenes) e na subcategoria de desenvolvimento embrionário (553 unigenes) foram utilizadas para determinar as espécies com melhores *hits* (*Top-hit*) do Blastx através do programa BlastGO.

## 2.2. IDENTIFICAÇÃO DE TRANSCRITOS DE GENES RELACIONADOS AO DESENVOLVIMENTO BASEADA EM RT-PCR

Em paralelo à identificação pela análise transcriptômica, a estratégia baseada em RT-PCR foi utilizada para identificar transcritos dos genes mais conhecidos (*Lb*, *Pb*, *Hox3*, *Dfd*, *Antp*, *Ubx*, *AbdA*, *AbdB*, *Eve*, *En* e *Wg*) e que foram relacionados ao desenvolvimento em outros artrópodes.

### 2.2.1. Embriões de *M. olfersii* e extração de RNA

A coleta, manutenção de animais e o estagiamento embrionário foi realizado conforme descrito no Capítulo I (*item 2.1*). Onze diferentes estágios do desenvolvimento embrionário (E1, E3-E10, E12, E14) de *M. olfersii* foram utilizados na extração de RNA, o qual foi realizado como descrito no Capítulo I (*item 2.2*). Um *pool* de ovos (50-100 mg/estágio, aproximadamente um *pool* de 700-1500 ovos/estágio) para cada estágio

foi utilizado na extração de RNA e foi realizado em triplicata para cada estágio.

### 2.2.2. Síntese de cDNA

Para a síntese de cDNA foram utilizadas 1 µg de RNA total de embriões em E1 até E14 e o kit GoScript™ Reverse Transcriptase (Promega). A reação de transcrição reversa consistiu de 4 µL de *GoScript™ 5X Reaction Buffer*, *GoScript™ Reverse transcriptase* (160 U), 0,5 µM de cada dNTP, 0,5 µg de iniciador Oligo(dT)<sub>15</sub>, 2,5 mM de MgCl<sub>2</sub> e *ribonuclease inhibitor* (20 U). As reações foram incubadas a 25°C por 5 min, a 42°C por 60 min e a 70°C durante 15 min. A amplificação do gene  $\beta$ -actina foi utilizado como controle positivo de cada reação de síntese de cDNA.

### 2.2.3. Desenho de iniciadores degenerados

Iniciadores degenerados foram desenhados para identificar transcritos dos genes relacionados ao desenvolvimento de *M. olfersii*, tais como os genes *Hox* (*Lb*, *Pb*, *Hox3*, *Dfd*, *Antp*, *Ubx*, *AbdA* e *AbdB*), genes de segmentação (*Eve*, *En* e *Wg*) e genes da família *Wnt* (Tabela 6). Iniciadores adicionais para *Wnt2* e *Wnt8* também foram utilizados. Alguns iniciadores degenerados utilizados foram descritos na literatura, no entanto, algumas mudanças foram realizadas manualmente em alguns iniciadores para melhorar a especificidade em crustáceos, o qual foi realizado baseado no alinhamento múltiplo de sequências de crustáceos e artrópodes disponíveis no banco de dados *GenBank*. As regiões conservadas 5' e 3' do homeodomínio dos genes *Hox*, *Eve* e *En* foram escolhidas para o desenho dos iniciadores. Para os genes da família *Wnt*, iniciadores degenerados foram desenhados em regiões conservadas do alinhamento das sequências.

### 2.2.4. Amplificação pela PCR e purificação dos produtos

Para a PCR foi utilizado o kit *GoTaq® Green Master Mix* (Promega), 1 µM de iniciadores degenerados, cDNA diluído (1/10) de embriões e condições de PCR específicas para os diferentes genes. Assim, as condições da PCR para os genes *Antp*, *Ubx*, *AbdA* e *Eve* foram de 94°C durante 5 min, seguido por 35 ciclos (94°C por 1 min, 50°C por 1 min e 72°C por 1 min) e extensão final de 72°C durante 10 min. As mesmas condições de PCR foram utilizadas para os outros genes, com variação na temperatura de anelamento para 45°C (*Lb*, *Pb*, *Hox3*, *Dfd* e *AbdB*), 54°C (*Wnt2*, *Wnt8* e outros da família *Wnt*) e 56°C (*En*).

**Tabela 6.** Iniciadores degenerados utilizados na RT-PCR para a identificação dos transcritos dos genes relacionados ao desenvolvimento em *M. olfersii*.

Gene	Seqüência dos iniciadores Fw e Rv (5'-3')	Região conservada (aa)	Temperatura de anelamento	Referência
<i>Labial</i>	CCNACNTAYAARTGGATGCARGT	PTYKWMQ	50°C	Neste estudo
	CKNCKRRTTYTGRAACCADATYTT	KIWFQNR		Williams e Holland (2000)
<i>Antennapedia</i>	TAYCCITGGATGMGIWSICARTT	YPWMRSQF	50°C	Abzhanov e Kaufman (2000)
	CKICKRRTTYTGRAACCA	WFQNR		Abzhanov e Kaufman (2000)
<i>Ultrabithorax</i>	ACNTTYTAYCCITGGATG	TFYPWM	50°C	Neste estudo
	TGYTTYTCYTYGTCRTTIARYTC	ELNEQEK		Neste estudo
<i>Abdominal A</i>	TNCCIMGITAYCCITGGATG	PRYPWM	50°C	Neste estudo
	TCRTIDATYTCYTIACIGCIC	AVKEIN		Abzhanov e Kaufman (2000)
<i>Abdominal B</i>	CCTCTSGAATGGACNGGNMANRT	PLEWTGX	45°C	Neste estudo
	CKNCKRRTTYTGRAACCADATYTT	KIWFQNR		Williams e Holland (2000)
<i>Deformed</i>	ACRTTYTNGIRITNGGNA	PNTKN	45°C	Neste estudo
	CKNCKRRTTYTGRAACCADATYTT	KIWFQNR		Williams e Holland (2000)
<i>Hox3</i>	TAYATGTGGATGAARAAR	YPWMKK	45°C	Neste estudo
	CKNCKRRTTYTGRAACCADATYTT	KIWFQNR		Williams e Holland (2000)
<i>Hox3</i>	TACACVWSVGCNCARYTRG	YTSACL	45°C	Neste estudo
	CKNCKRRTTYTGRAACCADATYTT	KIWFQNR		Williams e Holland (2000)
<i>Proboscipedia</i>	TAYCCNTGGATGAARGARAAGA	YPWMKEKK	45°C	Neste estudo
	CKNCKRRTTYTGRAACCADATYTT	KIWFQNR		Williams e Holland (2000)
<i>Even-skipped</i>	ACNGCNTTYACNMNGARCA	TAFTXEF	50°C	Damen et al. (2000)
	TCNGTYTTRTANGGYTGRAANA	FQPYKT		Damen et al. (2000)
<i>Engrailed</i>	TGGCCNGCNTGGGNTAYTGY	WPAWVYC	56°C	Peterson et al. (1998)
	RTIRATANARNCCYTGNGCCAT	MAQGLYN		Peterson et al. (1998)
Família <i>Wnt</i>	GARTGYAARTGYCAYGG	ECKCHG	54°C	Damen (2002)
	CATNARRTCRCANCCRTC	DGCDLM		Damen (2002)
<i>Wingless</i>	GAHMGHTTYGACGGHGCSCTC	DRFDGA	54°C	Neste estudo
	CATNARRTCRCANCCRTC	DGCDLM		Damen (2002)
<i>Wnt2 e Wnt8</i>	TGGGGNGGNTGYWSNGA	WGCGX	45°C	McGregor et al. (2008)
	NAYNCCRTGRCAYTTRCA	CKCHGX		McGregor et al. (2008)

Para verificar a eficácia da amplificação foram realizadas eletroforeses em gel de agarose (1,5%), corado com brometo de etídeo (0,5 µg/ml) ou GelRed (1:500, Biotium) dos produtos da PCR. Posteriormente, os géis foram visualizados e fotodocumentados utilizando o sistema de fotodocumentação ChemiDoc MP. O tamanho aproximado dos produtos de amplificação dos transcritos dos genes foi determinado, a partir do alinhamento múltiplo, baseado em outras seqüências de crustáceos. As bandas referentes aos produtos da RT-PCR de cada gene foram cortadas do gel e purificadas utilizando o kit *Smarter Nucleic Acid Sample Preparation* (Strattec Molecular), de acordo com as instruções do fabricante.

### 2.2.5. Clonagem dos produtos de RT-PCR e sequenciamento

Os produtos da PCR foram ligados ao vetor pGEM-T ou pGEM-T *Easy* (Promega) e os plasmídeos foram utilizados para a transformação de células eletrocompetentes *E. coli* DH5α. Os clones positivos foram analisados através da amplificação por PCR, utilizando iniciadores universais M13-F (5'-



CGCCAGGGTTTTCCAGTCACGAC-3') e M13-R (5'-TCACACAGGAAACAGCTATGAC-3'). O protocolo de PCR utilizado foi o mesmo descrito no *item 2.2.4*, com a mudança na Ta para 58°C e a concentração dos iniciadores para 0,4 µM. Depois, os produtos de PCR foram purificados novamente do gel de agarose utilizando o kit *Smarter Nucleic Acid Sample Preparation* (Stratec Molecular) e posteriormente enviados para o sequenciamento. Os produtos de PCR purificados de mínimo 2 clones de *E. coli* para cada gene foi enviado para o sequenciamento.

O sequenciamento foi realizado no Laboratório de Protozoologia – MIP/CCB/UFSC utilizando o equipamento *Hitachi 3500 Genetic Analyzer* (AB Applied Biosystems®). As reações de sequenciamento foram realizadas com o kit *BigDye® Terminator* (AB Applied Biosystems®), de acordo com as instruções do fabricante. As reações de sequenciamento foram realizadas com 5 pmol dos iniciadores universais M13-F e/ou M13-R e aproximadamente 30 ng - 120 ng de fragmento da PCR purificado, nas seguintes condições térmicas: 96°C por 1 min, seguidos de 35 ciclos de desnaturação a 96°C por 15 seg, ligação dos iniciadores a 50°C por 15 seg e extensão a 60°C por 4 min. Posteriormente, os produtos da reação de sequenciamento foram precipitados com etanol/EDTA para retirar nucleotídeos e iniciadores não incorporados. Os produtos foram eluídos em formamida e eletroinjetados no equipamento *Hitachi 3500 Genetic Analyzer* com programa padrão. A qualidade do sequenciamento foi analisada utilizando pacote *Phred/Phrap/Consed* (<http://www.phrap.org>) no Laboratório de Bioinformática (MIP/CCB/UFSC). As bases das sequências com um *score* de qualidade de Phred  $\geq 20$  foram consideradas aceitáveis no resultado do sequenciamento. Para confirmar a identidade das sequências obtidas, cada sequência de nucleotídeos foi traduzida a aminoácidos com o programa *ORF Finder* (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>), e comparadas utilizando o algoritmo Blastp com o banco de dados de sequências *GenBank*.

### 2.3. ANÁLISE BIOINFORMÁTICA E FILOGENIA DE GENES RELACIONADOS AO DESENVOLVIMENTO EM *M. olfersii*

As sequências de nucleotídeos e de aminoácidos dos genes relacionados ao desenvolvimento (genes *Hox*, genes de segmentação e da família *Wnt*) do filo Artropoda foram obtidas do banco de dados *Genbank*. Além disso, as sequências dos transcritos dos genes obtidos

no presente trabalho foram traduzidas em aminoácidos utilizando o programa ORF *Finder* (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>). O alinhamento dessas sequências de aminoácidos foi realizado com ClustalW e editadas com o programa Bioedit (HALL, 1999). As árvores filogenéticas foram construídas a partir do alinhamento das sequências de nucleotídeos ou de aminoácidos, dependendo da saturação de dados das sequências dos genes, utilizando o programa ClustalX2 v.2 (LARKIN et al., 2007). Para as árvores filogenéticas foram utilizadas 8 sequências de *Eve*, 12 sequências para *Ftz* e *Hh*, 14 sequências de *Lab* e *Scr*, 16 sequências de *AbdA* e *AbdB*, 17 sequências de *Dfd*, 19 sequências de *Antp*, 20 sequências de *En*, 21 sequências de *Ubx*. As construções das árvores filogenéticas foram realizadas com o programa MEGA v.5.05 (TAMURA et al., 2011). Os *gaps* foram excluídos da análise e foi utilizado o método de Neighbor-Joining. A consistência da árvore foi avaliada pela análise de *bootstrap* de 1.000 replicatas. As sequências dos genes de espécies do filo Onycophora foram utilizadas como grupo externo (*outgroup*) das árvores filogenéticas.

#### 2.4. IDENTIFICAÇÃO DE GENES *Hox* OU SEUS TRANSCRITOS EM CRUSTÁCEOS NA BASE DE DADOS *GenBank*

O algoritmo tBlastn (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>), que compara sequências de aminoácidos com sequências de nucleotídeos traduzidas do banco de dados, foi utilizado para identificar as sequências dos genes *Hox* ou seus transcritos em outros crustáceos decápodes. A busca foi realizada em *Transcriptome shotgun assembly* (TSA) de crustáceos no banco de dados *GenBank*. Além disso, foi realizada uma busca em *Sequence reads archive* (SRA) de *M. rosenbergii* e *M. nipponense*. As sequências de transcritos destes genes, identificados em crustáceos, foram utilizadas como sequência *query* (sequência alvo utilizada como base para encontrar similaridade com as sequências do banco de dados). Para os genes *Hox*, um especial interesse foi dirigido para a identificação da sequência de nucleotídeos que codifica a região da proteína compreendida entre o motivo YPWM e o homeodomínio.