

Daniella Helena Hock

**RNA-SEQ E EXPRESSÃO GÊNICA DIFERENCIAL DURANTE
REGENERAÇÃO TECIDUAL INDUZIDA EM *Calliactis polyopus*
(CNIDARIA: HORMATHIIDAE)**

Trabalho de Conclusão de Curso
apresentado como requisito para
cumprimento da disciplina TCC II
(BIO7016) do currículo do Curso de
Graduação em Ciências Biológicas da
Universidade Federal de Santa
Catarina.

Orientadora: Prof^ª Dr^ª Andrea Rita
Marrero

Co-orientador: Prof. Dr. Peter Prentis

Florianópolis

2016

Daniella Helena Hock

**RNA-SEQ E EXPRESSÃO GÊNICA DIFERENCIAL DURANTE
REGENERAÇÃO TECIDUAL INDUZIDA EM *Calliactis polypus*
(CNIDARIA: HORMATHIIDAE)**

Este Trabalho de Conclusão de Curso foi julgado adequado para o cumprimento da disciplina TCC II (BIO7016) e aprovado em sua forma final pelo Banca Examinadora.

Florianópolis, 20 de Julho de 2016

Prof^ª Dr^ª Maria Risoleta F. Marques
Coordenadora do Curso de Ciências Biológicas

Banca Examinadora:

Prof^ª Dr^ª Andrea Rita Marrero
Orientadora
UFSC

Dr.^a Sara Emelie Löfgren
UFSC

Dr. Renato Hajenius Aché de Freitas
UFSC

Aos meus pais, amigos e professores.

AGRADECIMENTOS

À Universidade Federal de Santa Catarina (UFSC) por ser sede de realização desta graduação. À todos os professores que dedicadamente lecionaram suas disciplinas ao longo destes anos.

À *Queensland University of Technology* (QUT) e aos queridos professores Peter Prentis e Ana Pavasovic por terem acreditado no meu potencial e por todo o apoio pessoal e acadêmico prestado para que eu pudesse realizar esta pesquisa na Austrália. Agradeço também aos colegas do e-PGL, Joki, Zach, Amanda, Chloe, Mathilde e Shorash pelo suporte e conversas nas bancadas do laboratório.

Ao CNPq pelo fomento concedido através das bolsas PIBIC – ITI/A. À UFSC pelas bolsa de extensão e monitoria recebidas durante a minha graduação. Ao programa Ciência sem Fronteiras pelo intercâmbio realizado no exterior.

À banca examinadora Sara Emelie Löfgren e Renato Hajenius Aché de Freitas pela cuidadosa correção e valiosa contribuição para este trabalho.

À minha querida orientadora Andrea Rita Marrero por ter aceitado me orientar neste TCC às escuras enquanto ainda estava na Austrália. Agradeço imensamente pelas conversas que tivemos ao longo deste semestre e por ser minha guia acadêmica, profissional e pessoal. À todos os integrantes do LAPOGE pelo acolhimento e apoio.

Ao professor Boris Stambuk e colegas de pesquisa do LBMBL por proporcionar meu primeiro contato com o mundo da pesquisa. À professora Malva Hernández pela participação no projeto de extensão realizado no Parque do Córrego Grande.

Aos meus amigos e colegas que de alguma forma fizeram-se presente nesta caminhada até hoje. À Joane Prata por toda a amizade e suporte que compartilhamos nestes últimos meses. Ao Nestor Wendt por toda ajuda e companhia para os almoços no RU. Ao Gabriel Ferrari pela amizade e contribuições para este trabalho.

Aos meus pais e irmãos por todo o apoio e suporte para que eu pudesse concluir esta graduação além de realizar meus sonhos pessoais e profissionais. Vocês são meu incentivo diário e meu orgulho. À todos os meus familiares, em especial à minha Tia Rose, ao meu Tio Ralf e à minha vózinha querida por serem responsáveis de tudo o que hoje sou.

“Science never solves a problem without creating ten more.”

George Bernard Shaw

RESUMO

A cicatrização é um processo regenerativo complexo que visa restaurar a integridade e função dos tecidos danificados. Em cnidários, o filo irmão de Bilateria (vertebrados), a habilidade de algumas espécies de reparar injúrias e regenerar estruturas corporais perdidas se mantém pouco elucidada. A compreensão dos mecanismos moleculares que sustentam a resposta à cicatrização em cnidários pode revelar alguns processos moleculares que contribuem para o reparo dos tecidos. A anêmona do mar *Calliactis polypus* tem sido observada pela sua excepcional capacidade de cicatrização e regeneração ainda não descrita na literatura. RNA-seq e expressão gênica diferencial foram realizados em dois indivíduos de *C. polypus* sendo um em regeneração e outro controle afim de determinar quais genes estavam sendo expressos diferentemente após dano epidérmico. Verificou-se que genes sobre-expressos durante a cicatrização pertencem a classes biológicas como diferenciação celular, morfogênese, ciclo celular, apoptose, estresse celular e detoxificação. Além disso, tanto a ativação quanto a inativação de proteínas contendo o domínio proteico transcriptase reversa foram identificadas como domínios mais sobre-expressos durante o reparo tecidual na anêmona do mar *C. polypus*. Portanto, uma resposta sistêmica e complexa aparenta estar associada com a cicatrização em *C. polypus*, sendo que a chave para o reparo em cnidários pode estar envolvida com novas proteínas que não puderam ser identificadas pelos bancos de dados atuais. Futuros estudos em genômica funcional e proteômica sobre a regeneração em *C. polypus* poderão oferecer um melhor entendimento da função biológica destes genes desconhecidos na regulação do reparo de lesão e regeneração em cnidários e, possivelmente, em outros filos.

Palavras-chave: Cnidaria; Bioinformática; Reparo tecidual; cicatrização.

ABSTRACT

Wound healing is a complex regenerative process that aims to restore integrity and function to damaged tissues. In cnidarians, the sister phylum to Bilateria (vertebrates), the outstanding ability of some species to repair wounds and regenerate lost body structures remains poorly understood. Understanding the molecular mechanisms that underpin the wound healing response in cnidarian species can unravel some molecular processes that contribute to tissue repair. The sea anemone *Calliactis polypus* has been observed to have remarkable abilities of wound healing and regeneration which has not been previously characterized in the literature. RNA-seq and differential gene expression analysis were performed on two individuals of *C. polypus*, under regeneration and control in order to determine which genes were differentially expressed following epidermal injury. As a result, genes over-expressed during wound healing belonged to biological classes such as cellular differentiation, morphogenesis, cell cycle, apoptosis, cellular stress and detoxification. Furthermore, both activation and inactivation of proteins containing the reverse transcriptase domain were identified as the topmost over-expressed protein domain during tissue repair in the sea anemone *C. polypus*. Therefore, a systemic and complex response seems to be associated with wound healing in *C. polypus* and the key to the successful repair in cnidarians may involve novel proteins that remain unidentified due to current databases. Further genomic functional and proteomic studies of the wound healing response in *C. polypus* can offer insight into the role of these unknown genes in regulating wound healing and regeneration in cnidarians and possibly in other phyla.

Keywords: Cnidaria; Bioinformatics; Tissue repair; Wound healing.

LISTA DE ABREVIATURAS E SIGLAS

BLAST	<i>Basic Local Alignment Search Tool</i>
BUSCO	<i>Benchmarking Universal Single-Copy Orthologs</i>
CD-HIT	<i>Cluster Database at High Identity with Tolerance</i>
cDNA	DNA complementar
DNA	Ácido desoxirribonucleico
dNTP	Desoxirribonucleotídeos fosfatados
EdgeR	<i>Differential Expression Analysis of Digital Gene Expression Data</i>
ePGL	<i>Evolutionary and Physiological Genomics Laboratory</i>
HMMER	<i>Hidden Markov Model-based sequence alignment tool</i>
mRNA	RNA mensageiro
NCBI	<i>National Center for Biotechnology Information</i>
ORFs	Janela de leitura aberta
OrthoDB	<i>Database of Orthologous Group</i>
PCR	Reação em cadeia da polimerase
Pfam	<i>Protein Family Database</i>
QUT	<i>Queensland University of Technology</i>
RNA	Ácido ribonucleico
SMART	<i>Simple Modular Architecture Research Tool</i>
Swiss-Prot	<i>Nucleotide Sequence Database</i>
TrEMBL	<i>Translated European Molecular Biology Laboratory</i>
UniProtKB	<i>Universal Protein Resource Knowledgebase</i>

SUMÁRIO

1. INTRODUÇÃO	19
1.1 DO REPARO DA LESÃO À REGENERAÇÃO	19
1.2 EVOLUÇÃO DA REGENERAÇÃO.....	20
1.3 A REGENERAÇÃO NO FILO CNIDARIA	21
1.4 A ANÊMOMA DO MAR <i>CALLIACTISPOLYPUS</i>	24
1.5 SEQUENCIAMENTO DO TRANSCRIPTOMA.....	23
1.6 OBJETIVOS.....	25
1.5.1 <i>Objetivo Geral</i>	25
1.5.2 <i>Objetivos Específicos</i>	25
2. DESENVOLVIMENTO	26
2.1 MATERIAIS E MÉTODOS	26
2.1.1 <i>Coleta e manutenção dos animais</i>	26
2.1.2 <i>Ensaio de dano tecidual</i>	26
2.1.3 <i>Extração de RNA total</i>	27
2.1.4 <i>Preparação da biblioteca de cDNA e RNA-seq</i>	27
2.1.5 <i>Análises de bioinformática</i>	28
2.1.5.1 <i>Montagem do transcriptoma</i>	29
2.1.5.2 <i>Expressão gênica diferencial</i>	30
2.1.5.3 <i>Anotação dos contigs</i>	30
2.2. RESULTADOS	31
2.2.1 <i>Sequenciamento do cDNA</i>	31
2.2.2 <i>Montagem de novo do transcriptoma</i>	32
2.2.3 <i>Expressão gênica diferencial</i>	33
2.1.5.3 <i>Anotação dos contigs</i>	34
2.3. DISCUSSÃO	35
2.3.1 <i>Resposta molecular na regeneração tecidual induzida</i>	35
2.3.2 <i>Ativação de proteínas contendo o domínio “Transcriptase reversa” durante o processo de regeneração</i>	38
3. CONCLUSÃO	39
REFERÊNCIAS	41
APÊNDICE A – DESCRIÇÃO	49

1. INTRODUÇÃO

1.1 DO REPARO DA LESÃO À REGENERAÇÃO

Regeneração é um termo abrangente que envolve desde eventos de reparo em nível celular e tecidual até a recuperação de membros perdidos e recomposição de indivíduos inteiros a partir de fragmentos do corpo, sendo vital para organismos multicelulares na restauração da função e arquitetura do tecido lesionado.

O processo regenerativo pode ser classificado em duas etapas teóricas: cicatrização e regeneração. Estas etapas são compostas de eventos contínuos e dinâmicos passíveis de sobreposição e podem ocorrer em diferentes tecidos e intervalos de tempo em um mesmo organismo. A cicatrização é o primeiro estágio do processo de regeneração e envolve complexos mecanismos moleculares, celulares e químicos para assegurar um rápido fechamento da área danificada (MENDONÇA; COUTINHO-NETTO, 2009). A cicatrização é classicamente dividida em três fases: inflamação, proliferação e remodelação (GURTNER et al., 2008). A fase inflamatória acontece imediatamente após a injúria do tecido e promove a ativação de: [1] vias do sistema imune para prevenção de infecções pela área afetada [2] cascatas de coagulação a fim de evitar a perda de fluidos e componentes celulares e [3] vias inflamatórias para a remoção do tecido danificado. A fase proliferativa da cicatrização é caracterizada pela formação e migração de diferentes células e compostos químicos, como fibroblastos e colágenos e possui como função principal a recomposição da matriz extracelular, a qual desempenha papel fundamental como substrato para a posterior formação do tecido novo (MENDONÇA; COUTINHO-NETTO, 2009). A terceira fase da cicatrização é chamada de remodelação ou maturação onde as células recrutadas para iniciar o reparo da injúria sofrem apoptose e componentes da matriz extracelular como o colágeno e fibroblasto são remodelados (GURTNER et al., 2008, MENDONÇA; COUTINHO-NETTO, 2009).

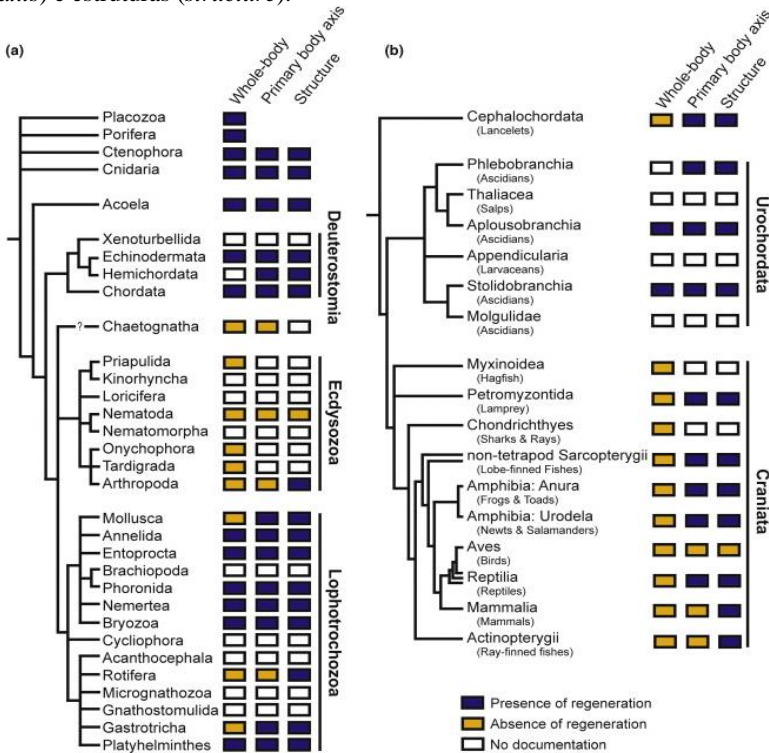
O segundo estágio do processo regenerativo é a regeneração propriamente dita, a qual é classificada em morfolática e epimórfica (SÁNCHEZ-ALVARADO, 2000). Na regeneração morfolática a recuperação do tecido perdido é feita a partir do conjunto de células remanescentes no local da injúria, através do rearranjo de tecidos pré-existentes, ocorrendo assim pouca ou nenhuma proliferação celular. A morfalaxia é observada por exemplo no cnidário *Hydra* (CUMMINGS; BODE, 1984). Já na epimorfia, processos como desdiferenciação, transdiferenciação, reprogramação celular, proliferação e uso de células

tronco podem estar presentes na restauração do tecido após a injúria. Este tipo de regeneração ocorre em diversos táxons como em platelmintos, moluscos, urocordados, anfíbios e participa da regeneração da cauda e membros de vertebrados e do fígado de humanos (SÁNCHEZ-ALVARADO, 2000, GARCIA-ARRARAS et al., 2006)

1.2 EVOLUÇÃO DA REGENERAÇÃO

Embora todos os animais apresentem a habilidade de cicatrizar e regenerar tecidos danificados, a capacidade de regeneração completa foi, de alguma forma, perdida ou inativada ao longo da evolução das espécies (Figura 1).

Figura 1. Poder regenerativo de diferentes táxons quanto a capacidade de regenerar o corpo todo (*whole-body*), eixo primário do corpo (*primary body axis*) e estruturas (*structure*).



Fonte: Bely e Nyberg (2010).

Os cnidários (Cnidaria), planárias (Platyhelminthes) e anfíbios (Amphibia) pertencem a táxons com excepcional poder regenerativo quando comparados a seus grupos próximos (SÁNCHEZ-ALVARADO, 2000; GALLIOT; GHILA, 2010, BRADSHAW; THOMPSON; FRANK, 2015).

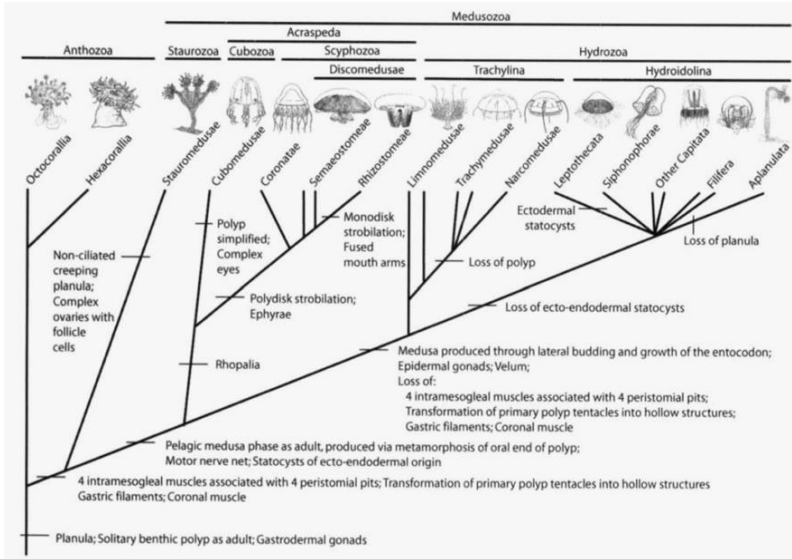
Whitehead et. al (2005) realizaram experimentos com mutantes de *Danio rerio* (Actinopterygii) desprovidos de mecanismos moleculares desencadeadores da regeneração e demonstraram que apesar de os mutantes não serem capazes de regenerar as nadadeiras perdidas, a cicatrização ocorria normalmente. Isto indica que o processo de regeneração é independente da cicatrização, sendo esta uma possível explicação da razão pela qual a cicatrização se mantém conservada em todas as espécies ao passo que a regeneração é conservada para alguns táxons.

Estudos recentes com sequenciamento do genoma e transcriptoma de *Hydra* tem revelado que o filo Cnidaria compartilha quase completamente os componentes moleculares (genes) com Bilateria, táxon que inclui desde platelmintos até vertebrados. (CHAPMAN et al., 2010; HEMMRICH et al., 2012; KRISHNA et al., 2013; WENGER; GALLIOT, 2013). Uma possível explicação para estas observações pressupõe que o componente genético necessário para a regeneração completa possa ter origem primitiva e se mantém presente em todos os animais, mas que a inativação ou perda de determinados genes tem limitado a capacidade de regeneração em determinados táxons. (BRADSHAW; THOMPSON; FRANK, 2015).

1.3 A REGENERAÇÃO NO FILO CNIDARIA

A capacidade regenerativa é altamente variável dentro do filo Cnidaria, sendo uma das linhagens evolutivas menos derivadas conhecidas pela presença de células tronco (HEMMRICH et al., 2007). Os cnidários são organismos diploblásticos, radialmente simétricos e organizados em tecidos (KRISHNA et al., 2013). Eles pertencem ao clado irmão de Bilateria, tendo divergido há mais de 500 milhões de anos atrás (DUNN et al., 2008; PHILIPPE et al., 2009). Este filo é classificado em quatro classes distintas: Anthozoa, Hydrozoa, Cubozoa e Scyphozoa (Figura 2) (BRIDGE et al., 1992, 1995, CARTWRIGHT et al., 2007).

Figura 2. Árvore filogenética hipotética para o filo Cnidaria considerando caracteres ancestrais.



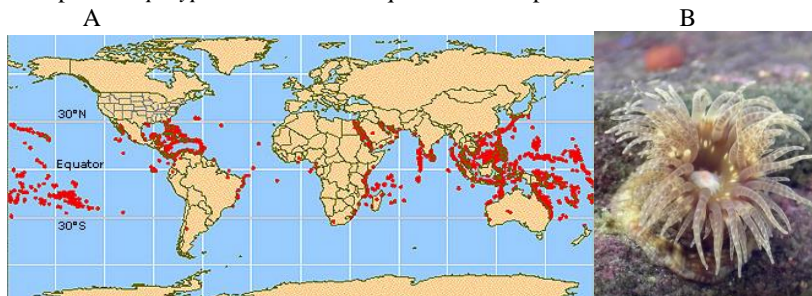
Fonte: Collins et al. (2006).

O gênero *Hydra* (Medusozoa: Hydrozoa) possui uma incrível habilidade de regenerar o pólipo por completo a partir de pequenos fragmentos ou até de células únicas dissociadas (LI; YANG; ZHONG, 2015). Já pólipos da anêmona do mar *Nematostella vectensis* (Anthozoa: Hexacorallia) possuem uma capacidade de regeneração limitada, não sendo capazes de regenerar um organismo adulto a partir de células únicas dissociadas (PETERSEN et al., 2015). Apesar de alguns estudos terem contribuído para o entendimento da regeneração nos animais, os mecanismos genéticos que sustentam esta variação na capacidade de regeneração entre cnidários permanecem ainda desconhecidos (PETERSEN et al., 2015).

1.4 A ANÊMONA DO MAR *CALLIACTIS POLYPUS*

A espécie de anêmona do mar *Calliactis polypus* (Forsskål, 1775) apresenta distribuição ao sul do Oceano Pacífico, Golfo do México, Moçambique, Havaí e Mar Vermelho (FAUTIN, 2015) (Figura 3). Esta espécie pode chegar a 8 cm de tamanho e é conhecida pela associação simbiótica com o caranguejo eremita (ROSS, 1970).

Figura 3. (A) Distribuição global da espécie *C. polypus*. (B) Indivíduo da espécie *C. polypus* em um dos tanques mantidos pelo ePGL.



Fonte: (A) NOAA (B) Daniella Helena Hock.

Pouco se tem estudado sobre a biologia e fisiologia do cnidário *C. polypus*. Esta espécie possui uma rápida e distinta capacidade de regeneração (dados não publicados pelo ePGL) cuja resposta a injúrias não tem sido documentada na literatura. O genoma da anêmona *C. polypus* ainda não foi sequenciado, fazendo com o que o alinhamento e identificação de genes seja feito a partir de espécies próximas que tenham seu genoma sequenciado ou através da alta similaridade das seqüências com os genes de espécies distantes.

1.5 SEQUENCIAMENTO DO TRANSCRIPTOMA

O transcriptoma pode representar desde o perfil completo dos genes que estão sendo expressos em uma única célula até em tecidos e organismos inteiros (WANG; GERSTEIN; SNYDER, 2009). Além disso, o sequenciamento do transcriptoma (ou RNA-seq) tem sido amplamente utilizado na quantificação da expressão de genes, identificação de novos transcritos e isoformas gênicas, mesmo para organismos cujo genoma de referência ainda não foi finalizado (MORTAZAVI et al., 2008, FINSETH; HARRISON, 2014).

O sequenciamento do transcriptoma inicia-se com a extração do mRNA, podendo este ser a partir de uma célula, tecido ou indivíduo e para uma determinada condição ou estágio do desenvolvimento. O mRNA é então transformado em cDNA pela enzima transcriptase reversa e adicionado à célula de fluxo sequenciador onde se hibridiza com oligonucleotídeos complementares pré-fixados na célula de fluxo. A técnica de sequenciamento do Illumina baseia-se na detecção simultânea da fluorescência liberada pelo dNTP quando incorporado à

sequência de cDNA que está sendo sintetizada na célula de fluxo (FULLER et al, 2009).

Os dados produzidos ao final do sequenciamento podem ser analisados através do alinhamento contra um genoma de referência ou, caso a espécie não possua seu genoma sequenciado, pela montagem *de novo*[†] dos *reads* (sequências brutas extraídas do sequenciador) (HAAS et al, 2013). A montagem *de novo* do transcriptoma requer estratégias de bioinformática, matemáticas e grande poder computacional para que os *reads* possam ser remontados em sequências maiores, chamadas de *contigs*. Um dos métodos utilizados para a montagem *de novo* do transcriptoma envolve a decomposição dos *reads* em fragmentos menores, chamados de grafos de Bruijn, com o intuito de montar as sequências considerando possíveis *splicing* alternativos e isoformas ocorridos na transcrição do DNA para o RNA (HAAS et al, 2013). Os *contigs* agora montados podem ser alinhados para anotação contra um ou mais banco de dados.

A fim de extrair informação biológica das proteínas correspondentes aos *contigs* montados, o banco de dados do NCBI (www.ncbi.nlm.nih.gov) e UniProtKB promovem uma anotação de proteínas com alto grau de informação, acurácia, estabilidade e informação (UNIPROT CONSORTIUM, 2007, 2011). Este banco de dados conta com uma rica gama de sequências de proteínas que podem ter sido manualmente depositadas e curadas no banco interno do UniProtKB, o Swiss-Prot, ou eletronicamente depositadas e anotadas no banco de dados do TrEMBL, também pertencente ao UniProtKB. O banco de dados do UniProtKB/Swiss-Prot contém informação sobre a função, estrutura e interações da sequência da proteína, domínios biologicamente relevantes, especificidade tecidual, modificações pós-translacionais, isoformas, dentre outras (UNIPROT CONSORTIUM, 2009). Já o banco de dados do UniProtKB/TrEMBL conta com informações sobre as traduções da grande maioria das regiões codantes presentes nos bancos de dados EMBL, GenBank, Ensembl *Homo sapiens*, dentre outros (UNIPROT CONSORTIUM, 2009). O banco de dados do Pfam contém mais de 13.000 famílias de proteínas anotadas

[†] *de novo*: expressão em latim que significa “do começo” e amplamente utilizada para descrever a montagem *de novo* de transcriptomas e genomas que não possuem referência.

manualmente e que podem ser buscadas através da similaridade com regiões altamente conservadas chamadas de domínios proteicos (FINN, 2009, PUNTA et al., 2011). Infelizmente, transcriptomas sequenciados *de novo* são insuficientemente anotados por serem alinhados contra banco de dados que contém sequências de proteínas identificadas para outras espécies (HAAS et al, 2013), fazendo com que a identificação de proteínas seja um desafio na anotação de sequências montadas *de novo*.

Uma das técnicas utilizadas para identificar quais genes estão sendo ativados ou inativados durante uma condição ambiental é a expressão diferencial de genes. Esta técnica utiliza os *reads* brutos após RNA-seq e um alto poder computacional para identificar os *reads* que estão sendo expressos em uma condição, mas não em outra (HAAS et al, 2013). A grande vantagem da expressão diferencial dos genes é de poder ser realizada mesmo na ausência de réplicas biológicas, através da utilização de modelos estatísticos, como a distribuição binomial negativa, que infere uma variação esperada em amostras biológicas para as análises (GRABHERR et al., 2011, HAAS et al, 2013).

O estudo do reparo de feridas em vários filos pode melhorar o entendimento sobre a cicatrização em organismos mais derivados, contribuindo assim para a identificação de moléculas ou vias metabólicas que permitam restaurar capacidades regenerativas (GURTNER et al., 2008) e contribuir com o desenvolvimento da engenharia tecidual e da medicina regenerativa.

1.6 OBJETIVOS

1.6.1 Objetivo Geral

Investigar presença ou ausência de expressão diferencial dos genes entre um indivíduo de anêmona do mar *C. polypus* em processo de regeneração tecidual e outro indivíduo da mesma espécie sem regeneração tecidual aparente a partir de transcriptomas previamente sequenciados pelo ePGL.

1.6.2 Objetivos Específicos

- Identificação dos genes expressos diferencialmente no processo de regeneração induzida entre dois indivíduos de anêmona do mar *C. polypus*, sendo um em regeneração tecidual e outro controle de mesma espécie.
- Anotação dos *contigs* em diferentes bancos de dados visando aumentar o grau de anotação e, conseqüentemente, função biológica dos

contigs expressos diferencialmente no processo de regeneração tecidual induzida.

- Analisar as classes das proteínas expressas diferencialmente, inferindo as possíveis funções e vias metabólicas ativadas durante o processo de regeneração tecidual na anêmona do mar através da anotação dos domínios de proteínas.

- Investigar a sobre-expressão (ou ativação) de domínios proteicos altamente conservados dentre as proteínas expressas diferencialmente no processo de regeneração tecidual da anêmona do mar *C. polypus*.

2. DESENVOLVIMENTO

2.1 MATERIAIS E MÉTODOS

2.1.1 Coleta e manutenção dos animais

Dois espécimes de anêmona do mar da espécie *C. polypus* foram coletados na região de Sunshine Coast, no estado de Queensland, Austrália e mantidos em tanques pelo *Evolutionary and Physiological Genomics Laboratory* (ePGL) vinculado ao grupo *Molecular Genetics Research* na *Queensland University of Technology* (QUT), Brisbane, Austrália. A coleta foi realizada de acordo com *Fisheries Act 1994 (General Fisheries Permit)* da Austrália, que dispensa aprovação do conselho de ética. As condições físico-químicas dos tanques foram mantidas estáveis pelos técnicos do laboratório respeitando as condições particulares para cada espécie de invertebrado marinho contida nos tanques.

2.1.2 Ensaio de dano tecidual

Para estimular o processo regenerativo na anêmona do mar, uma perfuração foi realizada na região do disco basal, danificando as camadas da epiderme e endoderme. A coleta do animal foi seguida de congelamento instantâneo com nitrogênio líquido três dias após o ensaio de dano tecidual. Os subsequentes ensaios e análises foram aplicados a dois espécimes de anêmonas, uma sem regeneração tecidual aparente (controle) e outra coletada após três dias do ensaio de dano tecidual (em regeneração).

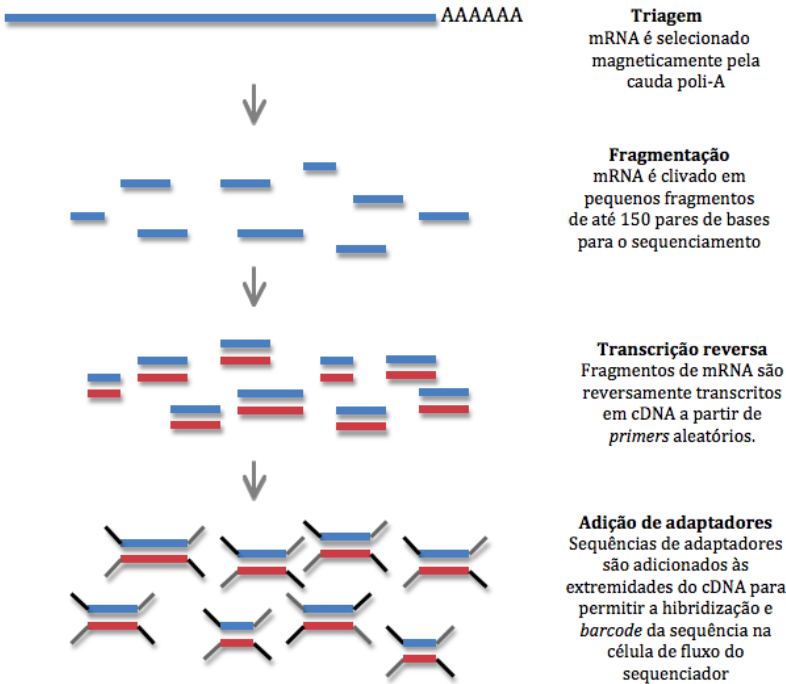
2.1.3 Extração de RNA total

Para a extração do RNA total, ambos animais inteiros foram macerados com nitrogênio líquido em pilões distintos e estéreis até que se obtivesse a textura fina de pó. As amostras foram homogeneizadas com 1 ml de TRIzol[®] até completa dissolução e incubadas por 5 minutos em temperatura ambiente. Foi adicionado 0,2 ml de Clorofórmio às amostras, homogeneizadas por 30 segundos e incubado em temperatura ambiente por 5 minutos, realizando estes últimos dois passos duas vezes. As amostras foram centrifugadas a 14.000 x g por 15 minutos e os sobrenadantes coletados para novos tubos e homogeneizados com 0,5 ml de isopropanol. O passo seguinte foi adicionar 100 ml de cloreto de sódio 1.2 M às amostras e incubá-las por 10 minutos em temperatura ambiente para então serem centrifugadas a 14.000 x g por 15 minutos. O sobrenadante foi coletado cuidadosamente e descartado, evitando o *pellet* de RNA. Para remover os resíduos não coletados pela pipeta, 200 ml de etanol 70% foram adicionados e homogeneizados por 10 segundos. As amostras foram então centrifugadas por 10 minutos a 14.000 x g para remoção do sobrenadante aguardando até a completa evaporação do etanol. O *pellet* de RNA foi ressuscitado em 50 µl de água livre de RNA e mantido em freezer -80 graus até a preparação da biblioteca de cDNA.

2.1.4 Preparação da biblioteca de cDNA e sequenciamento do transcriptoma

As bibliotecas de cDNA foram preparadas de acordo com as instruções do kit *Illumina TruSeq Stranded mRNA Library Preparation* em fragmentos de 150 pares de base (pb). O princípio de seleção do mRNA é por oligo(dT) com *beads* magnéticos que se ligam à cauda poli-A do mRNA e também a um suporte magnético que é posicionado na base da placa de ELISA com 96 poços. Este processo permite purificar a amostra com apenas mRNA contendo cauda poli-A dos demais RNAs extraídos previamente. Os mRNAs são então clivados quimicamente em fragmentos de até 150 pb e convertidos em cDNA a partir de *primers* aleatórios pela enzima transcriptase reversa. As extremidades dos cDNAs são reparadas e adeniladas para receber os adaptadores que possibilitam a hibridização na placa de fluxo do sequenciador além de um *barcode* único para a identificação pós-sequenciamento (Figura 4). As amostras foram amplificadas por PCR e validadas quanto a qualidade e concentração por um chip de DNA de alta sensibilidade no equipamento BioAnalyzer 2100 (Agilent).

Figura 4. Preparação da biblioteca de cDNA. A cauda Poli-A dos mRNAs é selecionada pelo magnetismo com oligo(dT) *beads*. A preparação é seguida da clivagem do mRNA em fragmentos de até 150 pb e de transcrição reversa para obtenção do cDNA. As terminações 5' e 3' do cDNA recebem adaptadores com *barcodes* únicos e sítios de ligação (*primers*) para hibridização na célula de fluxo do sequenciador.



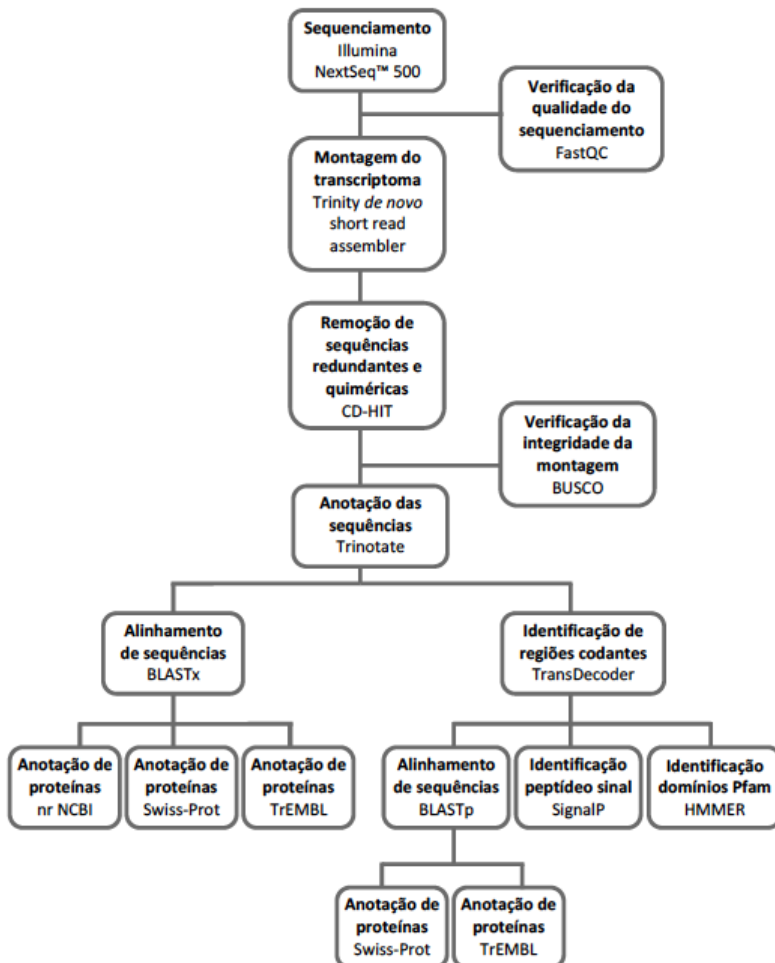
Fonte: Daniella Helena Hock

As amostras foram sequenciadas pelo Illumina NextSeq™ 500 para ambas fitas complementares (*paired-end reads*). Ao término do sequenciamento, os *reads* brutos gerados são filtrados para cada amostra através do *barcode* correspondente.

2.1.5 Análises de bioinformática

A série de análises de bioinformática realizadas a partir dos *reads* brutos do RNA-seq está representada na Figura 5. Relatórios FastQC foram gerados a fim de verificar o resultado e a qualidade do sequenciamento antes da montagem do transcriptoma.

Figura 5. Fluxograma representando as análises de bioinformática realizadas a partir dos *reads* brutos do RNA-seq, incluindo os programas, extensões e banco de dados utilizados.



Fonte: Daniella Helena Hock

2.1.5.1 Montagem do transcriptoma

O transcriptoma de referência foi montado *de novo* a partir dos *reads* brutos do sequenciamento do mRNA dos indivíduos controle e em regeneração pelo *Trinity short-read de novo assembler* (v. 2.0.6)

(GRABHERR et al., 2011). Os fragmentos brutos do RNA-seq de ambos indivíduos foram organizados independentemente em diversos grafos de Bruijn (COMPEAU; PEVZNER; TESLER, 2011), conceito matemático que utiliza a sobreposição de nucleotídeos das extremidades dos fragmentos sequenciados para unir e reconstruir a sequência original do mRNA (HAAS et al., 2013). As sequências quiméricas e redundantes foram removidas pelo CD-HIT (v. 4.6.1) (LI; JAROSZEWSKI; GODZIK, 2001; LI; GODZICK, 2006) com similaridade de 95% para a seleção da sequência mais representativa e remoção das demais. A totalidade e qualidade da montagem do transcriptoma foi estimada pelo programa BUSCO (v. 1.1b1) com *e-value* igual a 1×10^{-3} através da comparação das sequências montadas (*contigs*) a sequências ortólogas conservadas de cópia única dentre Metazoa contidas no banco de dados do (v. 8) DB (SIMAO et al., 2015). A estimativa da abundância dos *reads* foi realizada pelo método RSEM com alinhamento pelo Bowtie2 (LI et al., 2011; LANGMEAD; SALZBERG, 2012).

2.1.5.2 Expressão gênica diferencial

O nível de expressão gênica diferencial foi quantificado pelo pacote EdgeR (v. 2.2.1) (Bioconductor) no *Trinity short-read* de novo *assembler* através do alinhamento e contagem de cada um dos *reads* brutos contra o transcriptoma referência montado previamente (ROBINSON; MCCARTHY; SMYTH, 2009; ANDERS; HUBER, 2010). Testes estatísticos foram realizados pelo EdgeR com base na distribuição binomial negativa, a qual considera o aumento de variação observado entre amostras biológicas para calcular a expressão gênica diferencial, empregando *fold-change* (\log_2) igual a 4 e valor estatístico significativo $p = 0.001$. A normalização da contagem dos *reads* foi realizada com base no método TMM (*Trimmed Mean of M-values*) para possibilitar uma correta análise da expressão gênica diferencial visto que o número inicial de *reads* para o indivíduo controle e em regeneração apresentavam valores muito distantes. Ao final, foi gerado o gráfico *heatmap* para visualização geral dos resultados da expressão gênica diferencial (ROBINSON; OSHLACK, 2010; DILLIES et al., 2012; HAAS et al., 2013).

2.1.5.3 Anotação dos *contigs*

Os *contigs* foram primeiramente anotados pela *pipeline* Trinotate (HAAS et al., 2013) e então submetidos ao BLASTx contra o banco de dados não redundante (nr) do NCBI. Os *contigs* foram também anotados pelo banco de dados de proteínas do UniProtKB/Swiss-Prot e

UniProtKB/TrEMBL com *e-value* de 1×10^{-5} . Visando a ampliação da anotação dos *contigs*, a segunda estratégia envolveu a precedente identificação das potenciais regiões codantes (ORFs) dos *contigs* pelo programa TransDecoder (v. 2.0.1) para então serem submetidas ao BLASTp com anotação da sequência pelo UniProtKB/Swiss-Prot e UniProtKB/TrEMBL (*e-value* de 1×10^{-5}). Os *contigs* foram também identificados quanto ao peptídeo sinal (SignalP, v. 3.0) e domínios conservados de proteínas Pfam (HMMER, v. 3.1b2). As anotações provenientes do Swiss-Prot foram utilizadas como referência primária para a identificação da provável proteína e função por este ser um banco de dados anotado manualmente e revisado, valendo-se das demais anotações geradas como dados complementares na identificação da proteína. Os domínios Pfam identificados foram checados com o banco de dados do SMART (www.smart.embl-heidelberg.de), considerando como apropriada a anotação quando os domínios se apresentavam iguais. Um *script* em Python desenvolvido pelo próprio ePGL foi utilizado para a identificação dos domínios Pfam mais abundantes dentre as sequências expressas diferencialmente. O *script* contabilizou cada domínio Pfam apenas uma vez para cada *contig*, não contando múltiplas vezes quando os domínios eram repetidos em um mesmo *contig*, caracterizando assim quais foram os domínios Pfam sobre-expressos no processo de regeneração.

2.2. RESULTADOS

2.2.1 Sequenciamento do cDNA

De acordo com os relatórios FastQC, foram obtidos 209.875.116 *reads* brutos a partir do RNA-seq do indivíduo controle e 79.931.258 *reads* do indivíduo em regeneração, com o comprimento destas *reads* entre 35 e 151 pb para ambos RNA-seq (Tabela 1). A porcentagem de bases GC foi de 42% para o indivíduo controle e 43% para o indivíduo em regeneração. O PHRED *score* foi maior do que 21 para ambos transcriptomas, correspondendo a uma confiabilidade maior do que 99% de que cada fluorescência associada a uma base nitrogenada sequenciada foi corretamente identificada pelo sequenciador.

Tabela 1. Resumo das características gerais do relatório FastQC a partir dos dados brutos do RNA-seq (Illumina) dos indivíduos controle e em regeneração da anêmona do mar *C. polypus*.

Dados do RNA-seq	Controle	Em regeneração
<i>Reads</i> totais	209.875.116	79.931.258
Comprimento das sequências (pb)	35-151	35-151
Porcentagem GC	42%	43%
PHRED <i>score</i>	> 26	> 21

2.2.2 Montagem *de novo* do transcriptoma

Após a montagem *de novo* do transcriptoma referência e remoção das sequências redundantes e quiméricas, o número total de bases nitrogenadas montadas foi de 165.718.619 as quais geraram um total de 214.675 *contigs* (Tabela 2). O valor do *contig* N50 foi de 1.577 caracterizando que 50% ou mais das bases montadas estão em *contigs* com comprimento igual ou maior que 1.577 pb. Este valor é uma das estatísticas geradas para verificação da montagem do transcriptoma uma vez que a montagem dos *reads* em *contigs* mais longos tende a proporcionar uma maior qualidade para subseqüentes análises.

Tabela 2. Resultados obtidos da montagem do transcriptoma *de novo* da anêmona do mar *C. polypus*.

Dados do transcriptoma referência	
Número total de bases montadas (pb)	165.718.619
Número total de <i>contigs</i>	214.675
<i>Contig</i> N50	1.577
Porcentagem GC	37.63%
Média do <i>contig</i> (pb)	771.95

A segunda avaliação da qualidade do transcriptoma referência foi realizada pelo do programa BUSCO através da comparação com 843 sequências ortólogas de cópia única do banco de dados de Metazoa. O nível de integridade atingido pela montagem *de novo* do transcriptoma foi de 92% com 37% de duplicação, sendo 3,9% das sequências montadas fragmentadas e 3,6% faltantes (Tabela 3).

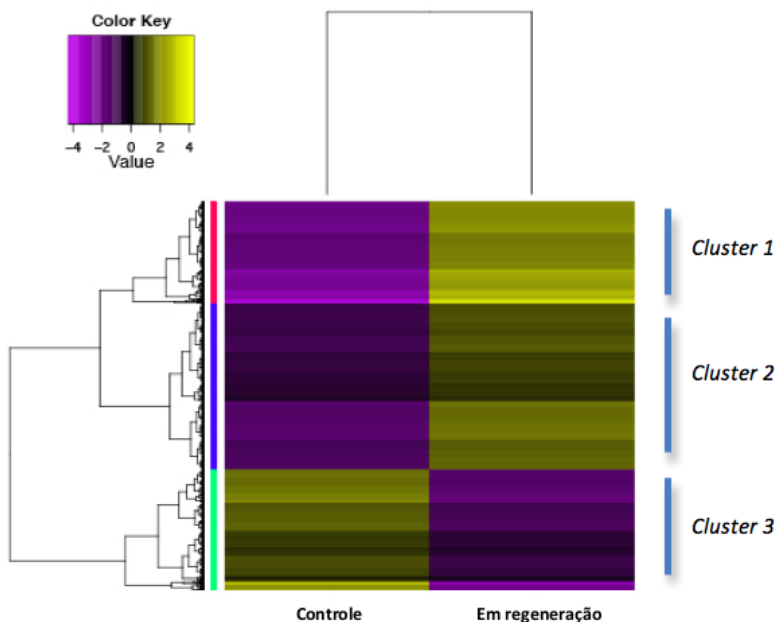
Tabela 3. Dados da qualidade e nível de integridade da montagem *de novo* do transcriptoma referência gerados pelo programa BUSCO.

Classificação das sequências pelo BUSCO	Avaliação
Completas	92%
Duplicadas	37%
Fragmentadas	3.9%
Faltantes	3.6%
Número total de genes analisados (Metazoa)	843

2.2.3 Expressão gênica diferencial

Foram identificados 3.218 *contigs* expressos diferencialmente entre os indivíduos controle e em regeneração, sendo que 2.445 *contigs* (*cluster 1* e *cluster 2*) foram sobre-expressos durante o processo de regeneração tecidual na anêmona do mar conforme apresentado na Figura 6, contrastando com os 773 *contigs* indicados no *cluster 3* que apresentaram repressão da transcrição gênica.

Figura 6. Heatmap dos genes expressos diferencialmente ($p = 0.001$ e 4-*fold*) durante a regeneração tecidual da anêmona do mar *C. polypus*. As regiões em roxo, preto e amarelo indicam supressão da expressão, expressão média padrão e sobre-expressão do mRNA, respectivamente.



Fonte: Daniella Helena Hock

2.2.4 Anotação dos *contigs*

Foram anotadas 1.297 proteínas com o banco de dados do Swiss-Prot a partir do *top* BLASTx *hit* dos *contigs* expressos diferencialmente durante a regeneração tecidual. A segunda estratégia de anotação de proteínas utilizada - com prévia identificação das ORFs - resultou em um total de 815 proteínas anotadas com o Swiss-Prot a partir do *top*

BLASTp *hit*, sendo que 33 destas proteínas não haviam sido previamente anotadas pelo BLASTx e Swiss-Prot. A Figura 7 (ver apêndice A para *e-values*) sumariza algumas das proteínas anotadas e suas prováveis funções no processo de regeneração tecidual da espécie *C. polypus*.

Figura 7. Prováveis proteínas identificadas e suas potenciais funções na regeneração tecidual da anêmona do mar *C. polypus*.



Fonte: Daniella Helena Hock

O domínio proteico Pfam identificado como mais abundante entre os 3.218 *contigs* expressos diferencialmente foi o da transcriptase reversa (PF00078), apresentando-se em 23 *contigs* dentre os 2.445 sobre-expressos e em 12 *contigs* dos 773 suprimidos durante a regeneração tecidual induzida na *C. polypus*.

2.3. DISCUSSÃO

2.3.1 Resposta molecular à regeneração tecidual induzida

O processo regenerativo desencadeia uma resposta sistêmica no organismo visando a restauração morfofuncional da área lesionada. Este processo parece promover um aumento na transcrição do DNA na anêmona do mar *C. polypus*, como indicado pelos 2.445 *contigs* ativados no indivíduo em regeneração (*Clusters 1 e 2*) em relação ao controle (*Cluster 3*) quando realizada a expressão gênica diferencial (Figura 6). O aumento da transcrição gênica na *C. polypus* está relacionado à ativação de diferentes sistemas e vias metabólicas, identificadas pela anotação das proteínas expressas diferencialmente (Figura 7). Quando a barreira epitelial é danificada por uma injúria, a troca livre de solutos e solventes da água para o interior do organismo é capaz de provocar estresse hídrico e iônico nas células. As chaperonas Hsp70 e Hsp90 são proteínas classicamente conhecidas pelo papel fundamental na manutenção da conformação correta de proteínas afetadas pelo estresse celular e garantia do enovelamento adequado de proteínas recém sintetizadas (HENDRICK; HARTL, 1995). A sobre-expressão de chaperonas Hsp70 e Hsp90 durante regeneração tecidual na espécie *C. polypus* pode ter função direta na manutenção da conformação correta das proteínas durante o estresse hídrico e químico causado pelo ferimento. Além disto, a sobre-expressão de genes identificada pela expressão diferencial de genes durante o reparo tecidual requer que as chaperonas realizem o correto enovelamento das proteínas recém traduzidas.

No processo de cicatrização e regeneração, a apoptose é responsável pela remoção de células comprometidas, para que estas sejam substituídas por células novas. A apoptose é um processo necessário para iniciar a regeneração da região anterior em *Hydra* (CHERA et al., 2009) e também no reparo de lesão ectodermal em *N. vectensis*, sendo que a presença de sinais apoptóticos nas células é encontrada em até seis horas após a injúria (DUBUC; TRAYLOR-KNOWLES; MARTINDALE, 2014). Em *C. polypus* a expressão de genes que traduzem para proteínas que desencadeiam a cascata apoptótica, como a caspase-2, ainda se apresentam sobre-expressos três dias após a injúria, diferentemente do que ocorre a nível celular na anêmona do mar *N. vectensis*. Estudos histoquímicos são necessários para investigar se a presença de sinais apoptóticos também ocorre a nível celular em três dias após injúria em *C. polypus*.

A restauração do epitélio danificado pode ser favorecida na espécie *C. polypus* pela caspase-14, um tipo de caspase não apoptótica

que está envolvida com diversos processos celulares como promoção da diferenciação e maturação de células epidérmicas, inflamação e apoptose (DENECKER et al., 2008).

O processo de regeneração tecidual requer a ativação de uma gama de genes e vias metabólicas, inevitavelmente desencadeando uma maior produção de produtos do metabolismo que podem ser tóxicos à célula. Para isto, o aumento de expressão de proteínas como a catalase e a glutathiona peroxidase auxilia na decomposição de produtos tóxicos do metabolismo em formas menos prejudiciais à célula, prevenindo assim do dano oxidativo.

Um dos mecanismos essenciais na recuperação da funcionalidade e estrutura do tecido lesionado é a proliferação celular a fim de restituir as células perdidas pela lesão tecidual. A sobre-expressão de proteínas como ciclinas G1/S, ciclinas G2 e treslina são responsáveis pela regulação do ciclo celular, garantindo assim a correta proliferação das células, mecanismo que pode estar envolvido no processo regenerativo na anêmona *C. polypus*. Proteínas como *homeobox* e *hedgehog* desempenham papéis fundamentais na regulação da expressão de genes que controlam a formação dos eixos e estruturas corporais bem como a de órgãos e tecidos (CHAPMAN et al., 2010). A orientação do eixo corporal é essencial para a diferenciação das novas células em tipos celulares corretos, garantindo assim o retorno da funcionalidade e arquitetura do tecido lesionado. Algumas proteínas da família *homeobox* estão envolvidas na manutenção da pluripotencialidade em células tronco embrionárias e necessárias à regeneração (GARDINER; BRYANT, 1996, ZAEHRES et al., 2005), indicando que a anêmona *C. polypus* pode ter a presença de populações de células-tronco favorecendo o processo regenerativo.

A presença de células tronco em organismos adultos ocorre em diversos táxons, sendo que alguns estudos tem elucidado a participação de células-tronco na cicatrização e regeneração. O cnidário *Hydra* apresenta populações de células tronco nas duas camadas do corpo (ectoderme e endoderme) (HOLSTEIN; HOBMAYER; TECHNAN, 2003). As células tronco presentes na *Hydra* se dividem passivamente, sendo necessária uma contínua ativação de proteínas responsáveis pela morfogênese para o direcionamento e diferenciação destas células tronco em seus tipos celulares corretos (HOLSTEIN; HOBMAYER; TECHNAN, 2003). Cummings e Bode (1984) utilizaram um agente químico que bloqueava a fase S da divisão celular antes de induzir regeneração em *Hydra*. Os autores concluíram que a divisão celular não é necessária para a formação de uma nova cabeça em *Hydra*,

caracterizando a regeneração em *Hydra* como morfolática. Além disso, o mecanismo de transdiferenciação celular parece não ser necessário para a regeneração da cabeça em *Hydra*, já que as células tronco presentes no epitélio são responsivas aos sinais de início da regeneração (HOLSTEIN; HOBMAYER; TECHNAU, 2003).

Estudos pioneiros com a água viva (Cnidaria) *P. carnea* demonstraram que o tecido muscular de um adulto sofre transdiferenciação em diversos outros tipos celulares durante a regeneração (SCHMID, 1992). A transdiferenciação é um mecanismo presente na regeneração epimórfica que, apesar de já ter sido descrita para cnidários, é predominantemente encontrada na regeneração de vertebrados. A regeneração da cauda em anfíbios urodelos e de cardiomiócitos em peixes é caracterizada pela desdiferenciação do tecido local da injúria a fim de formar uma massa de células não diferenciadas (blastema). As células do blastema então proliferam e rediferenciam para formar as estruturas perdidas (GARDINER; BRYANT, 1996, JOPLING; BOUE; BELMONTE, 2011). Dois mecanismos fundamentais descritos no processo de desdiferenciação celular na regeneração de membros em anfíbios são a presença da proteína supressora de tumor retinoblastoma (JOPLING; BOUE; BELMONTE, 2011) e o recrutamento de células satélites Pax7+ (TANAKA et al., 2016). Em *C. polypus*, proteínas ligantes em células satélites Pax7 (ver apêndice A para *e-value*) e proteínas retinoblastoma (ver apêndice A para *e-value*) foram identificadas no transcriptoma do indivíduo em regeneração, indicando que a anêmona do mar *C. polypus* pode utilizar a desdiferenciação celular no reparo da lesão.

A caracterização do tipo de regeneração (morfolaxia ou epimorfia) e dos processos envolvidos no reparo do tecido (desdiferenciação, transdiferenciação e reprogramação) na anêmona do mar *C. polypus* não puderam ser determinados apenas com base na anotação do transcriptoma. Estudos em nível celular, como histoquímica, e genômicos funcionais, como silenciamento dos genes de interesse, são estratégias fundamentais para suportar as ideias aqui discutidas. Apesar disso, sugere-se que o cnidário *C. polypus* apresenta uma complexa interação de mecanismos necessários para uma eficiente regeneração tecidual.

2.3.2 Ativação e inativação de proteínas contendo o domínio transcriptase reversa durante o processo de regeneração

O processo de regeneração na anêmona do mar *C. polypus* desencadeou mais expressivamente a ativação e também inativação de

proteínas contendo o domínio proteico biologicamente funcional da transcriptase reversa. Grande parte destas proteínas ainda não possui sua função completamente documentada na literatura, sendo o conjunto dos domínios um indicativo de prováveis funções.

Três explicações para a regulação de proteínas contendo o domínio transcriptase reversa serão aqui discutidas, ressaltando que estas três podem não corresponder à realidade ou ainda ocorrer simultaneamente. A primeira possibilidade pode ser relacionada a uma maior infecção das células da anêmona do mar por retrovírus – um tipo vírus que possui a enzima transcriptase reversa com a função de converter RNA em DNA afim de inserir o material genético viral na célula hospedeira e replicar – facilitado pelo ferimento em aberto e transportado através da água.

A segunda possibilidade para discutir a ativação e inativação de proteínas contendo o domínio transcriptase reversa é pela associação direta com retrotransposons, que são elementos genéticos que uma vez transcritos em mRNA são capazes de se reinserir em regiões aleatórias – ou não – do genoma. Caso esta inserção seja próxima ou dentro de regiões gênicas, pode-se induzir mutações, regular a expressão de genes através da inativação ou ativação de genes além de promover cópias de genes em diferentes lugares dentro do genoma. Isso pode, então, ter papel central na regulação da expressão gênica em *C. polypus* e na evolução da regeneração nas espécies.

A terceira explicação envolve a manutenção de células tronco e transformação celular pelos processos de desdiferenciação ou transdiferenciação. A conservação dos telômeros é processo essencial na manutenção de linhagens de células tronco. A presença da proteína telomerase transcriptase reversa foi identificada nos *contigs* sobre-expressos em *C. polypus*, indicando a possível existência e manutenção de células tronco em organismos adultos. A presença de populações de células tronco já tem sido previamente descrita em outros cnidários como a *Hydra* (PETERSEN et al., 2015) e aparentemente ausente em outros, como na anêmona do mar *N. vectensis* (DUBUC; TRAYLOR-KNOWLES; MARTINDALE, 2014). Estudos proteômicos e com silenciamento de genes ainda precisam ser realizados com a anêmona do mar *C. polypus* a fim de se obter mais evidências sobre a função biológica da ativação de proteínas contendo o domínio transcriptase reversa e da proteína telomerase transcriptase reversa. Por enquanto, os resultados obtidos com o sequenciamento do transcriptoma da *C. polypus* criaram diversos caminhos para pesquisas futuras, fornecendo um vasto campo de estudos sobre novos mecanismos responsáveis pela

cicatrização e regeneração em espécies evolutivamente basais.

3. CONCLUSÃO

A regeneração tecidual na anêmona do mar *C. polypus* desencadeia uma resposta complexa com expressão diferencial de mais de 3 mil genes em relação ao controle que não apresentava regeneração tecidual aparente. A anotação destes genes contra diferentes bancos de dados indicou que eles controlam processos biológicos como a diferenciação celular, morfogênese, ciclo celular, apoptose, estresse e detoxificação celular são sobre-expressos durante o processo de regeneração tecidual em *C. polypus*. Além disso, o domínio proteico transcriptase reversa foi o mais presente dentre todas as proteínas sobre-expressas, sugerindo que este domínio pode ter papel fundamental no processo regenerativo em *C. polypus*.

Mais estudos a nível celular, proteômico e genômico-funcional são necessários para que hipóteses mais acuradas sobre os componentes biológicos que governam a regeneração tecidual em *C. polypus* sejam elucidadas. Este trabalho aponta diversos caminhos para futuras investigações assim como contribui para o conhecimento biológico de uma espécie de anêmona do mar com poucas informações disponíveis na literatura.

A regeneração completa está presente em táxons evolutivamente basais, sendo que a anêmona do mar *C. polypus* ocupa uma dessas posições na evolução dos metazoários, indicando que a consolidação dos genes necessários para a regeneração tenha origem primitiva. Portanto, o estudo da capacidade regenerativa pode identificar novos genes e vias metabólicas essenciais à regeneração, gerando conhecimento que pode ser potencialmente aplicado ao desenvolvimento de novas terapias na medicina regenerativa.

REFERÊNCIAS

ANDERS, S.; HUBER, W. Differential expression analysis for sequence count data. **Genome Biol.** v. 11, n. 10, R106 p.1-12, out. 2010.

BELY, A. E.; NYBERG, K. G. Evolution of animal regeneration: re-emergence of a field. **Trends Ecol. Evol.**, v. 25, n. 3, p.161-170, mar. 2010.

BRADSHAW, B.; THOMPSON, K.; FRANK, U. Distinct mechanisms underlie oral vs aboral regeneration in the cnidarian *Hydractinia echinata*. **Elife**, v. 4, p.1-19, abr. 2015.

BRIDGE, D. et al. Class-level relationships in the phylum Cnidaria: evidence from mitochondrial genome structure. 1992. **Proc. Natl. Acad. Sci.**, n. 18 v. 89, p. 8750-8753.

BRIDGE, D. et al. Class-level relationships in the phylum Cnidaria: molecular and morphological evidence. **Mol. Biol. Evol.**, v. 12, n. 5, p.679-689, fev. 1995.

CARTWRIGHT, P. et al. Exceptionally Preserved Jellyfishes from the Middle Cambrian. **Plos One**, v. 2, n. 10, p.e1121, 31 out. 2007.

CHAPMAN, J. A. et al. The dynamic genome of *Hydra*. **Nature**, v. 464, n. 7288, p.592-596, mar. 2010.

CHERA, S. et al. Apoptotic Cells Provide an Unexpected Source of Wnt3 Signaling to Drive *Hydra* Head Regeneration. **Dev. Cell**, v. 17, n. 2, p.279-289, ago. 2009.

COLLINS, A. et al. Medusozoan Phylogeny and Character Evolution Clarified by New Large and Small Subunit rDNA Data and an

Assessment of the Utility of Phylogenetic Mixture Models. **Syst. Biol.**, v. 55, n. 1, p.97-115, fev. 2006.

COMPEAU, P.; PEVZNER, P.; TESLER, G. How to apply de Bruijn graphs to genome assembly. **Nat Biotechnol.** v. 29, n. 11, p.987-991, nov. 2011.

CUMMINGS, Susan G.; BODE, Hans R.. Head regeneration and polarity reversal in *Hydra attenuata* can occur in the absence of DNA synthesis. **Wilhelm Roux' Archiv**, v. 194, n. 2, p.79-86, 1984.

DENECKER, G. et al. Caspase-14 reveals its secrets. **J Cell Biol**, v. 180, n. 3, p.451-458, 4 fev. 2008.

DILLIES, M. et al. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. **Brief. Bioinform.** v. 14, n. 6, p.671-683, set. 2012.

DUBUC, T. Q.; TRAYLOR-KNOWLES, N.; MARTINDALE, M. Q. Initiating a regenerative response; cellular and molecular features of wound healing in the cnidarian *Nematostella vectensis*. **Bmc Biology**, v. 12, n. 1, p.24-54, 2014.

DUNN, C. W. et al. Broad phylogenomic sampling improves resolution of the animal tree of life. **Nature**, v. 452, n. 7188, p.745-749, 5 mar. 2008.

FAUTIN, D. *Calliactis polypus* (Forsskål, 1775). 2015. Disponível em: <<http://www.marinespecies.org/aphia.php?p=taxdetails&id=1360>>. Acesso em: 25 jun. 2016.

FINN, R. D. et al. The Pfam protein families database. **Nucleic Acids Res**, v. 38, p.D211-D222, 17 nov. 2009.

FINSETH, F. R.; HARRISON, R. G.. A Comparison of Next Generation Sequencing Technologies for Transcriptome Assembly and Utility for RNA-Seq in a Non-Model Bird. **Plos One**, v. 9, n. 10, p. e108550, out. 2014.

FULLER, C. et al. The challenges of sequencing by synthesis. **Nat Biotechnol.** v. 27, n. 11, p.1013-1023, nov. 2009.

GALLIOT, B.; GHILA, L. Cell plasticity in homeostasis and regeneration. **Mol. Reprod. Dev.**, v. 77, n. 10, p.837-855, jul. 2010.

GARCÍA-ARRARÁS, J. E. et al. Spherulocytes in the echinoderm *Holothuria glaberrima* and their involvement in intestinal regeneration. **Dev. Dyn.**, v. 235, n. 12, p.3259-3267, dez. 2006.

GARDINER, D. M.; BRYANT, S. V. Molecular mechanisms in the control of limb regeneration: the role of homeobox genes.. **Int J Dev Biol.**, v. 4, n. 40, p.797-805, ago. 1996.

GRABHERR, M. et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. **Nat Biotechnol.** v. 29, n. 7, p.644-652, maio 2011.

GURTNER, G. C. et al. Wound repair and regeneration. **Nature**, v. 453, n. 7193, p.314-321, maio 2008.

HAAS, B. et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. **Nat Protoc.** v. 8, n. 8, p.1494-1512, jul. 2013.

HEMMRICH, G. et al. Molecular signatures of the three stem cell lineages in *Hydra* and the emergence of stem cell function at the base of multicellularity. **Mol Biol Evo**, v. 29, n. 11, p.3267-3280, maio 2012.

HENDRICK, Jp; HARTL, Fu. The role of molecular chaperones in protein folding. **Faseb J**, v. 15, n. 9, p.1559-1569, dez. 1995.

HOLSTEIN, T. W.; HOBMAYER, E.; TECHNAU, U. Cnidarians: An evolutionarily conserved model system for regeneration?. **Dev. Dyn.**, v. 226, n. 2, p.257-267, jan. 2003.

KRISHNA, S. et al. Deep sequencing reveals unique small RNA repertoire that is regulated during head regeneration in *Hydra magnipapillata*. **Nucleic Acids Res**, v. 41, n. 1, p.599-616, 19 nov. 2012.

JOPLING, C.; BOUE, S.; BELMONTE, J. C. I. Dedifferentiation, transdifferentiation and reprogramming: three routes to regeneration. **Nat Rev Mol Cell Bio**, v. 12, n. 2, p.79-89, fev. 2011.

LANGMEAD, B.; SALZBERG, S. Fast gapped-read alignment with Bowtie 2. **Nat Methods**. v. 9, n. 4, p.357-359, mar. 2012.

LI, B. et al. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. **Bmc Bioinformatics**. v. 12, n. 1, p.323-323, ago. 2011.

LI, W.; GODZIK, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. **Bioinformatics**. v. 22, n. 13, p.1658-1659, maio 2006.

LI, W.; JAROSZEWSKI, L.; GODZIK, A. Clustering of highly homologous sequences to reduce the size of large protein databases. **Bioinformatics**. v. 17, n. 3, p.282-283, mar. 2001.

LI, Q.; YANG, H.; ZHONG, T. P. Regeneration across Metazoan Phylogeny: Lessons from Model Organisms. **J Genet Genomics**, v. 42, n. 2, p.57-70, fev. 2015.

MENDONÇA, R. J.; COUTINHO-NETTO, J. Aspectos celulares da cicatrização. **An. Bras. Dermatol.**, v. 84, n. 3, jul. 2009.

MORTAZAVI, A. et al. Mapping and quantifying mammalian transcriptomes by RNA-Seq. **Nat Methods**, v. 5, n. 7, p.621-628, 30 maio 2008.

NOAA. **National Oceanic and Atmospheric Administration**. Disponível em: <<http://oceanservice.noaa.gov>>. Acesso em: 21 jul. 2016.

PETERSEN, H. O. et al. A Comprehensive Transcriptomic and Proteomic Analysis of *Hydra* Head Regeneration. **Mol Biol Evol**, v. 32, n. 8, p.1928-1947, abr. 2015.

PHILIPPE, H. et al. Phylogenomics Revives Traditional Views on Deep Animal Relationships. **Current Biology**, v. 19, n. 8, p.706-712, abr. 2009.

PUNTA, M. et al. The Pfam protein families database. **Nuc Acids Res**, v. 40, n. 1, p.D290-D301, 29 nov. 2011.

ROBINSON, M.; MCCARTHY, D.; SMYTH, G. EdgeR: a Bioconductor package for differential expression analysis of digital gene expression data. **Bioinformatics**. v. 26, n. 1, p.139-140, nov. 2009.

ROBINSON, M.; OSHLACK, A. A scaling normalization method for differential expression analysis of RNA-seq data. **Genome Biol.** v. 11, n. 3, p.R25, mar. 2010.

ROSS, D. M. The commensal association of *Calliactis polypus* and the hermit crab *Dardanus gemmatus* in Hawaii. **Can. J. Zool.**, v. 48, n. 2, p.351-357, mar. 1970.

SÁNCHEZ-ALVARADO, A. Regeneration in the metazoans: why does it happen?. **Bioessays**, v. 22, n. 6, p.578-590, maio 2000.

SCHMID, V. Transdifferentiation in Medusae. **Int. Rev. Cytol.** p.213-261, 1992.

SIMÃO, F. et al. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. **Bioinformatics.** v. 31, n. 19, p.3210-3212, jun. 2015.

TANAKA, Hibiki Vincent et al. A developmentally regulated switch from stem cells to dedifferentiation for limb muscle regeneration in newts. **Nat. Comm.**, v. 7, p.11069-11077, mar. 2016

UNIPROT CONSORTIUM. The Universal Protein Resource (UniProt). **Nucleic Acids Res**, v. 36, p. D190-D195, dez. 2007.

_____. The Universal Protein Resource (UniProt) in 2010. **Nucleic Acids Res**, v. 38, p.D142-D148, out. 2009.

_____. Reorganizing the protein space at the Universal Protein Resource (UniProt). **Nucleic Acids Res**, v. 40, n. 1, p. D71-D75, nov. 2011.

WANG, Z.; GERSTEIN, M.; SNYDER, M. RNA-Seq: a revolutionary tool for transcriptomics. **Nat Rev Genet.** v. 10, n. 1, p.57-63, jan. 2009.

WENGER, Y.; GALLIOT, B. RNAseq versus genome-predicted transcriptomes: a large population of novel transcripts identified in an Illumina-454 *Hydra* transcriptome. **Bmc Genomics**, v. 14, n. 1, p.204, 2013.

WHITEHEAD, G. G. et al. Fgf20 Is Essential for Initiating Zebrafish Fin Regeneration. **Science**, v. 310, n. 5756, p.1957-1960, dez. 2005.

ZAEHRES, H. et al. High-Efficiency RNA Interference in Human Embryonic Stem Cells. **Stem Cells**, v. 23, n. 3, p.299-305, mar. 2005.

APÊNDICE A – Lista de proteínas e códigos dos domínios Pfam.

Tabela 4. Lista de proteínas anotadas com o banco de dados do Swiss-Prot e seus respectivos *e-values* e códigos dos domínios proteicos Pfam identificados.

Anotação	<i>e-value</i>	Pfam
<i>PAX3- and PAX7-binding protein 1</i>	2×10^{-161}	PF15458 PF07842
<i>Retinoblastoma-binding protein 5</i>	0	PF00400
<i>Catalase</i>	5×10^{-156}	PF00199
<i>Glutathione peroxidase</i>	1×10^{-25}	PF00255
<i>Caspase-14</i>	1×10^{-28}	PF00656
<i>Homeobox protein Hmx</i>	2×10^{-18}	PF00046 PF11569
<i>Desert hedgehog protein A</i>	1×10^{-22}	PF01079
<i>G1/S-specific cyclin-E</i>	3×10^{-105}	PF00134 PF02984
<i>G2/mitotic-specific cyclin-B2</i>	6×10^{-49}	PF00134 PF02984
<i>G2/mitotic-specific cyclin-B3</i>	2×10^{-91}	PF00134 PF02984
<i>Treslin</i>	7×10^{-26}	PF15292
<i>Caspase-2</i>	2×10^{-29}	PF00656
<i>Heat shock protein HSP 90-alpha 1</i>	4×10^{-134}	PF00183
<i>Heat shock 70 kDa protein C</i>	1×10^{-78}	PF00012

Fonte: Daniella Helena Hock

