



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE CIÊNCIAS BIOLÓGICAS
CURSO DE CIÊNCIAS BIOLÓGICAS

Beatriz Mira Lahoz

Montagem e anotação do transcriptoma tegumentar da Baleia Jubarte, *Megaptera novaeangliae* (Borowski, 1781).

Florianópolis

2024

Beatriz Mira Lahoz

Montagem e anotação do transcriptoma tegumentar da Baleia Jubarte, *Megaptera novaeangliae* (Borowski, 1781)

Trabalho de Conclusão de Curso submetido ao curso de Ciências Biológicas do Centro de Ciências Biológicas da Universidade Federal de Santa Catarina como requisito parcial para a obtenção do título de Bacharela em Ciências Biológicas.

Orientador: Dr. Clei Endrigo Piazza
Coorientador: Prof. Dr. Afonso Celso Dias Bainy

Florianópolis

2024

Lahoz, Beatriz Mira

Montagem e anotação do transcriptoma tegumentar da Baleia Jubarte, *Megaptera novaeangliae* (Borowski, 1781). / Beatriz Mira Lahoz ; orientador, Clei Endrigo Piazza, coorientador, Afonso Celso Dias Bainy, 2024.

42 p.

Trabalho de Conclusão de Curso (graduação) - Universidade Federal de Santa Catarina, Centro de Ciências Biológicas, Graduação em Ciências Biológicas, Florianópolis, 2024.

Inclui referências.

1. Ciências Biológicas. 2. Bioinformática. 3. Transcriptoma. 4. Mamíferos marinhos. 5. Biomarcadores. I. Piazza, Clei Endrigo . II. Bainy, Afonso Celso Dias. III. Universidade Federal de Santa Catarina. Graduação em Ciências Biológicas. IV. Título.

Beatriz Mira Lahoz

**Montagem e anotação do transcriptoma tegumentar da Baleia Jubarte, *Megaptera novaeangliae*
(Borowski, 1781)**

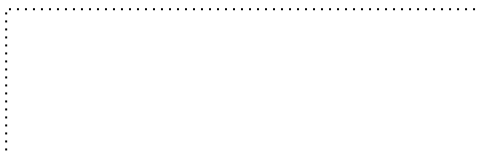
Este Trabalho de Conclusão de Curso foi julgado adequado para obtenção do título de Bacharela e
aprovado em sua forma final pelo Curso de Ciências Biológicas.

Florianópolis, 12 de dezembro de 2024.



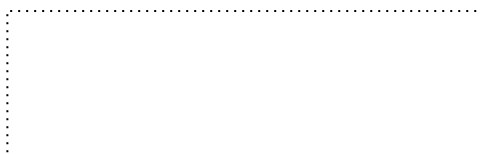
Coordenação do Curso

Banca examinadora



Dr. Clei Endrigo Piazza

Orientador(a)



Prof.(a) Dr.(a) Maria Risoleta Freire Marques

Universidade Federal de Santa Catarina



Prof.(a) Dr.(a) Guilherme de Toledo e Silva

Universidade Federal de Santa Catarina

Florianópolis, 2024.

AGRADECIMENTOS

A conclusão deste trabalho só foi possível graças ao apoio, orientação e incentivo de muitas pessoas e instituições, às quais dedico minha mais profunda gratidão.

À SOCIOAMBIENTAL, pelo compromisso exemplar na execução do Projeto de Monitoramento de Cetáceos (PMC) e pela coleta das amostras essenciais para este estudo, deixo meu sincero agradecimento. À Petrobras, por financiar este projeto e torná-lo viável, minha imensa gratidão pela confiança e investimento. Ao Laboratório Central de Tecnologias de Alto Desempenho em Ciências da Vida (LACTAD), que, com o sequenciamento do transcriptoma, forneceu os dados fundamentais para as análises aqui desenvolvidas, meu muito obrigado por contribuir tão diretamente com o avanço desta pesquisa. À UFSC e ao LABCAI pela estrutura e pelo curso de Ciências Biológicas.

Ao meu orientador, Dr. Clei Piazza, que acreditou em mim desde o início, compartilhando paciência, conhecimento e ensinamentos preciosos, tanto na ciência quanto na vida. Sua dedicação e orientação foram fundamentais para minha formação. Sou grata de coração. Ao meu coorientador, Prof. Dr. Afonso Bainy, pelo acolhimento em seu laboratório e por me permitir integrar o importante Projeto de Padronização de Técnicas e Execução de Análises de Biomarcadores Bioquímicos e Moleculares em Amostras Obtidas em Espécies de Aves, Quelônios e Mamíferos Marinhos. Sua contribuição foi essencial para meu crescimento acadêmico e profissional.

Agradeço também a todos que fazem ou fizeram parte deste projeto: Dai, Vanessa, Bárbara, Jacó, Ju, Luiza, Thiago e Gui. Cada um de vocês me ensinou algo que ajudou a moldar a pequena cientista que sou hoje. A todos os colegas do LABCAI, que se tornaram uma verdadeira família, agradeço por fazer do laboratório um lugar onde estar nunca foi uma obrigação, mas um prazer.

Ao Bruno, da Superintendência de Governança Eletrônica e Tecnologia da Informação e Comunicação (SETIC), pela paciência inestimável ao me ajudar com comandos que insistiam em não funcionar, mas que, no fim, eram apenas *typos*. Sua boa vontade e apoio técnico fizeram toda a diferença para a continuidade deste trabalho.

Agradeço ao meu pai, Carlos, que como cientista me inspira e apoia em cada momento da minha vida, me encorajando a seguir nesta jornada. À minha mãe, que, mesmo sem entender todos os detalhes da ciência, se orgulha profundamente de mim. À minha madrastra, Daniela, e ao meu padrasto, Mark, pelo suporte constante e incondicional ao longo da minha caminhada acadêmica.

Ao meu namorado, Pedro, que sempre acreditou em mim e me apoiou, e à minha melhor amiga e colega de graduação, Bianca, que compartilhou comigo as ansiedades e desafios deste período, o meu carinho e gratidão por estarem sempre ao meu lado.

Agradeço à Banca examinadora por se disponibilizar a avaliar esse trabalho.

A todos vocês, que direta ou indiretamente contribuíram para a realização deste trabalho, meu mais sincero e profundo agradecimento.

RESUMO

Os poluentes orgânicos persistentes (POPs) são xenobióticos frequentemente produzidos pelos seres humanos, que geralmente apresentam grande toxicidade, quando biodisponíveis. Esses contaminantes se acumulam principalmente nos ambientes aquáticos, que atuam como sumidouros devido à sua utilização para obtenção de alimento, matéria-prima, despejo de esgoto e recreação. Nesse contexto, torna-se necessário o monitoramento da qualidade da água, o qual pode ser realizado por meio de uma combinação de técnicas, como o uso de biomarcadores de contaminação aquática. Dentre os principais biomarcadores bioquímicos, destacam-se as enzimas de biotransformação, que tornam os contaminantes mais facilmente excretáveis pelos organismos e as enzimas antioxidantes, que protegem o organismo contra espécies reativas de oxigênio. Adicionalmente, os biomarcadores moleculares atuam sinergicamente com os bioquímicos, permitindo detectar o mRNA de genes envolvidos no metabolismo de xenobióticos, mesmo em níveis muito baixos. Alguns POPs podem agir como desreguladores endócrinos, imitando hormônios esteroides sexuais ao se ligarem a receptores nucleares, perturbando a homeostase. Ademais, o sistema imunológico atua como uma barreira essencial na defesa contra esses contaminantes, contribuindo para a resposta imunológica e proteção dos organismos. A ecotoxigenômica, através da montagem de transcriptomas, possibilita a caracterização de padrões de expressão gênica, que podem ser usados para identificar contaminação aquática, servindo como uma ferramenta valiosa no contexto do monitoramento ambiental. Nesse cenário, a baleia-jubarte (*Megaptera novaeangliae*) é uma ótima candidata para tais estudos, por ser um animal longo vivo, filtrador e com grande quantidade de gordura corporal, características que favorecem o acúmulo de toxinas. Portanto, o presente estudo buscou realizar a montagem e anotação do transcriptoma tegumentar de *Megaptera novaeangliae*, visando analisar a transcrição de genes envolvidos no metabolismo de xenobióticos. Para isso, foram obtidas biópsias de pele de baleias-jubarte por meio do Projeto de Monitoramento de Cetáceos (PMC). O RNA das amostras foi extraído, purificado e utilizado para a construção de uma biblioteca de cDNA, que foi sequenciada na plataforma Illumina NextSeq2000 com leituras pareadas (*paired-end*). A montagem *de novo* foi realizada com o software Trinity, enquanto a anotação funcional foi feita com o Trinotate, integrando análises como BLAST+, HMMER e Infernal, além de bancos de dados como SwissProt, Pfam, Rfam, KEGG e Gene Ontology. O sequenciamento gerou 93.356.722 leituras pareadas, resultando em 124.580 contigs de transcritos. Dentre as sequências anotadas, foram identificados genes associados à biotransformação de xenobióticos, como enzimas da superfamília do Citocromo P450, Glutathione-S-Transferases e transportadores ABC; enzimas antioxidantes, como Glutathione Peroxidases e Superóxido Dismutases; receptores nucleares, como Receptor de Aroclor-hidrocarboneto e receptor de estrogênio; e genes associados ao sistema imunológico, como interleucinas e receptores toll-like. Esta pesquisa ressalta a importância da biotecnologia ambiental e da ecotoxigenômica no monitoramento dos impactos das atividades antropogênicas nos ecossistemas marinhos, destacando o potencial dos dados gerados para a descoberta de novos biomarcadores de contaminação aquática. Além disso, ela cria um recurso valioso para a pesquisa em diversas outras áreas, como evolução molecular, genética, e imunologia, ampliando o alcance e a relevância dos resultados encontrados, principalmente pelo fato da baleia jubarte se tratar de um animal não modelo.

Palavras-chave: biomarcadores, mamíferos marinhos, bioinformática

ABSTRACT

Persistent organic pollutants (POPs) are xenobiotics commonly produced by humans, which generally exhibit high toxicity when bioavailable. These contaminants primarily accumulate in aquatic environments, which act as sinks due to their use for food, raw material extraction, sewage disposal, and recreation. In this context, water quality monitoring becomes essential and can be carried out through a combination of techniques, such as the use of aquatic contamination biomarkers. Among the main biochemical biomarkers are biotransformation enzymes, which make contaminants more easily excreted by organisms, and antioxidant enzymes, which protect the organism against reactive oxygen species. Additionally, molecular biomarkers work synergistically with biochemical ones, enabling the detection of mRNA from genes involved in xenobiotic metabolism, even at very low levels. Some POPs can act as endocrine disruptors, mimicking sexual steroid hormones by binding to nuclear receptors, thus disturbing homeostasis. Moreover, the immune system serves as a critical barrier in defending against these contaminants, contributing to the immune response and protection of organisms. Ecotoxicogenomics, through the assembly of transcriptomes, enables the characterization of gene expression patterns, which can be used to identify aquatic contamination, serving as a valuable tool in environmental monitoring. In this context, the humpback whale (*Megaptera novaeangliae*) is an excellent candidate for such studies, as it is a long-lived, filter-feeding animal with a large amount of body fat, characteristics that favor toxin accumulation. Therefore, this study aimed to assemble and annotate the skin transcriptome of *Megaptera novaeangliae*, focusing on analyzing the transcription of genes involved in xenobiotic metabolism. For this purpose, skin biopsies of humpback whales were obtained through the Cetacean Monitoring Project (PMC). RNA from the samples was extracted, purified, and used for cDNA library construction, which was sequenced on the Illumina NextSeq2000 platform with paired-end reads. De novo assembly was performed using the Trinity software, and functional annotation was done with Trinotate, integrating analyses such as BLAST+, HMMER, and Infernal, along with databases like SwissProt, Pfam, Rfam, KEGG, and Gene Ontology. The sequencing generated 93,356,722 paired reads, resulting in 124,580 transcript contigs. Among the annotated sequences, genes associated with xenobiotic biotransformation, such as *Cytochrome P450* superfamily enzymes, *Glutathione-S-Transferases*, and *ABC transporters*; antioxidant enzymes, such as *Glutathione Peroxidases* and *Superoxide Dismutases*; nuclear receptors, such as *Aryl Hydrocarbon Receptor* and estrogen receptor; and immune system-related genes, such as *interleukins* and *toll-like receptors*, were identified. This research highlights the importance of environmental biotechnology and ecotoxicogenomics in monitoring the impacts of anthropogenic activities on marine ecosystems, emphasizing the potential of the generated data for discovering new biomarkers of aquatic contamination. Moreover, it creates a valuable resource for research in various other fields, such as molecular evolution, genetics, and immunology, expanding the scope and relevance of the findings, especially considering that the humpback whale is a non-model animal.

Keywords: biomarkers; marine mammals; bioinformatics.

LISTA DE FIGURAS

Figura 1: <i>Megaptera novaeangliae</i>	15
Figura 2: Rota migratória da baleia jubarte no Brasil	15
Figura 3: Coleta de biópsias com balestra em uma baleia jubarte durante o PMC-BS.....	18
Figura 4: Mapa da localização das baleias cujas amostras foram utilizadas para à extração de RNA.....	19
Figura 5: Representação esquemática da metodologia utilizada	22
Figura 6: Gel de agarose com formaldeído 1,2% com as bandas 18S e 28S do RNA ribossomal.....	23
Figura 7: Perfil eletroforético de RNA TapeStation system das amostras de tecido tegumentar de <i>M. novaeangliae</i> e seus respectivos valores de RIN.....	24
Figura 8: Local onde a amostra de baleia jubarte usada no sequenciamento foi coletada	25
Figura 9: Histograma da qualidade das sequências brutas (senso) avaliado pelo FastQC	25
Figura 10: Histograma da qualidade das sequências brutas (antissenso) avaliado pelo FastQC	26
Figura 11: Histograma da qualidade das sequências filtradas (senso) avaliado pelo FastQC ..	27
Figura 12: Histograma da qualidade das sequências filtradas (antissenso) avaliado pelo FastQC	27
Figura 13: Perfil da expressão Ex contra ExN50	28
Figura 14: Distribuição dos 30 termos de ontologia gênica mais expressos	36

LISTA DE TABELAS

Tabela 1: Dados sobre a coleta das amostras de baleia jubarte para a extração de RNA.....	18
Tabela 2: Análise da concentração e pureza do RNA em espectrofotômetro	23
Tabela 3: Métricas da montagem via Trinity.....	29
Tabela 4: Identificação das ORFs completas, parciais 3' e 5', e internas via Transdecoder ...	30
Tabela 5: métricas de montagem via BUSCO.....	31
Tabela 6: Anotação funcional do transcriptoma tegumentar da baleia jubarte via Trinotate utilizando as maiores ORFs identificadas	32
Tabela 7: Genes envolvidos na biotransformação de fase 1.....	33
Tabela 8: Genes envolvidos na biotransformação de fase 2.....	33
Tabela 9: Genes envolvidos na biotransformação de fase 3.....	33
Tabela 10: Genes de receptores nucleares	34
Tabela 11: Genes do sistema antioxidante	34
Tabela 12: Genes do sistema imunológico	35
Tabela 13: Sequências senso (F) e antissenso (R) dos iniciadores e tamanho do produto (<i>amplicon</i>) para cada gene.....	37

LISTA DE ABREVIATURAS E SIGLAS

®: Marca comercial registrada, do inglês "Registered Trademark"
18S: RNA ribossomal 18S
28S: RNA ribossomal 28S
ABC: *ATP binding cassette*
ABCC1: *Proteína associada a multirresistência 1*
AhR: *Receptor de aril hidrocarboneto*
BLAST: Basic Local Alignment Search Tool
BUSCO: *Benchmarking Universal Single-Copy Orthologs*
CAT: *Catalase*
Cq: Ciclo de Quantificação
CYP: *Citocromo P450*
CYP1B: *Citocromo P450 família 1B*
DTT: Ditioneitol
eggNOG: banco de dados de grupos ortólogos e anotações funcionais
EX90N50: Métrica N50 da montagem considerando 90% da expressão total
ExN50: Métrica N50 da montagem considerando "X" % da expressão total
FMO: *Flavina Monooxigenase*
ESR1: *Receptor de Estrogênio*
GO: Ontologia gênica, do inglês "Gene Ontology"
GPx: *Glutathione peroxidase*
GR: *Glutathione Redutase*
GST: *Glutathione S-transferase*
GSTA2: *Glutathione S-transferase A2*
HMMER: Modelo oculto de Markov
HPA: Hidrocarbonetos Policíclicos Aromáticos
IL10: *Interleucina 10*
IUCN: União Internacional para a Conservação da Natureza
Infernal: *INFERENCE of RNA ALIGNment*
KEGG: Enciclopédia de Genes e Genoma de Kioto
LABCAI: Laboratório de Biomarcadores de Contaminação Aquática e Imunoquímica
LaCTAD: Laboratório Central de Tecnologias de Alto Desempenho em Ciências da Vida
N50: Métrica estatística do comprimento médio dos transcritos em 50% da montagem
NCBI: Centro Nacional de Informação Biotecnológica
ORFs: Janelas Abertas de Leitura, do inglês "Open Reading Frames"
Pb: pares de bases
PE: Extremidades Pareadas, do inglês "Paired ended"
PMC: Projeto de Monitoramento de Cetáceos
POPs: Poluentes Orgânicos Persistentes
PCB: Bifenilas Policloradas
RIN: RNA Integrity Number
SF: *Sulfotransferases*
SRA: Banco de dados *Sequence read archive*
SOD: *Superóxido dismutases*
SULT: *Sulfotransferases*
TPM: Transcritos Por Milhão
UFSC: Universidade Federal de Santa Catarina
Unicamp: Universidade Estadual de Campinas

SUMÁRIO

1	INTRODUÇÃO	12
1.1	OS POLUENTES ORGÂNICOS PERSISTENTES	12
1.2	OS BIOMARCADORES DE CONTAMINAÇÃO AQUÁTICA	13
1.3	A ECOTOXICOGENÔMICA.....	14
1.4	OS MAMÍFEROS MARINHOS	14
2	OBJETIVOS	17
2.1	OBJETIVO GERAL	17
2.2	OBJETIVOS ESPECÍFICOS	17
3	METODOLOGIA.....	17
3.1	A OBTENÇÃO DE AMOSTRAS	17
3.2	EXTRAÇÃO DO RNA	19
3.3	SEQUENCIAMENTO DO TRANSCRIPTOMA.....	19
3.4	MONTAGEM E ANOTAÇÃO.....	20
4	RESULTADOS E DISCUSSÃO.....	22
4.1	ANÁLISE DA QUALIDADE DO RNA.....	22
4.2	ANÁLISE DOS <i>READS</i> BRUTOS	25
4.3	MONTAGEM DO TRANSCRIPTOMA	27
4.4	ANOTAÇÃO FUNCIONAL.....	31
4.5	DESENHO DOS INICIADORES	36
5	CONCLUSÃO.....	37
6	PERSPECTIVAS FUTURAS	38
	REFERÊNCIAS.....	39

1 INTRODUÇÃO

1.1 OS POLUENTES ORGÂNICOS PERSISTENTES

Poluentes Orgânicos Persistentes (POPs) são xenobióticos, isto é, compostos exógenos aos seres vivos, que são produzidos pelos seres humanos e possuem grande toxicidade. Esses compostos são tipicamente hidrofóbicos e lipofílicos, o que faz com que eles se acumulem ao longo da cadeia trófica, já que eles são dificilmente metabolizados (Rieger, 2002). Por serem, frequentemente, semivoláteis, estão sujeitos a um transporte atmosférico de longo alcance, amplificando as consequências da sua persistência no meio ambiente (Jones; Voogt, 1999), podendo ser encontrados em diversos compartimentos ambientais, como rios, lagos, oceanos, aquíferos, solo e sedimentos (Sharma *et al.*, 2014). São exemplos de POPs, os pesticidas, como o Dicloro-Difenil-Tricloroetano (DDT); os diversos compostos químicos industriais, como as bifenilas policloradas (PCBs) e os Hidrocarbonetos Policíclicos Aromáticos (HPAs), lançados no ambiente através de incêndios florestais, queima incompleta de matéria orgânica ou de derivados do petróleo (Boreivik *et al.*, 2004).

A produção e o uso de diversos POPs foram restritos ou banidos na Convenção de Estocolmo (UNEP, 2001). Porém, eles ainda perduram no meio ambiente. Além do seu efeito tóxico imediato, esses componentes podem causar mal funcionamento dos órgãos, reduzir a eficiência do sistema imune e causar alterações nos níveis hormonais (Helle *et al.*, 1976; Bergman e Olsson, 1985; Jepson *et al.*, 2005; Kannan *et al.*, 2000).

A alteração dos níveis hormonais, como a dos hormônios esteroides, pode indicar uma diminuição do sucesso reprodutivo, devido a uma desregulação endócrina, possivelmente causada por xenobióticos, tendo em vista que as isoenzimas do citocromo P450 (CYP) são envolvidas no metabolismo desses hormônios. Animais de alto nível trófico comumente apresentam taxas reprodutivas limitadas e, geralmente, são os mais vulneráveis neste quesito (Van Der Oost; Beyer; Vermeulen, 2003). Ainda, alguns poluentes orgânicos persistentes podem ser desreguladores endócrinos, agindo de forma similar a hormônios esteroides sexuais, se ligando a receptores nucleares, como os receptores nucleares de estrogênio (ESRs), perturbando a homeostase (Fossi *et al.*, 2003).

Devido aos impactos nocivos causados pelos poluentes orgânicos persistentes nos ecossistemas e na saúde dos organismos, torna-se fundamental monitorar e avaliar as respostas biológicas que possam indicar a exposição a esses contaminantes. Para isso, o uso de biomarcadores é uma ferramenta crucial, pois permite a identificação de efeitos prejudiciais e alterações moleculares em estágios iniciais.

1.2 OS BIOMARCADORES DE CONTAMINAÇÃO AQUÁTICA

Os ambientes aquáticos constituem o sumidouro dos contaminantes e sofrem uma pressão das atividades humanas, como a sua utilização para a obtenção de alimento e de matéria-prima, o despejo de esgoto e a recreação, que acarretam perda da biodiversidade (Borgwardt, 2019). Sendo assim, torna-se necessário o monitoramento da qualidade da água através de uma combinação de técnicas, que incluem os biomarcadores de contaminação aquática, visto que estes podem elucidar os efeitos dos contaminantes nos organismos vivos, ao passo que apenas análises físico-químicas não o fazem. Os biomarcadores são quantificações em fluidos corporais, células ou tecidos que indicam modificações bioquímicas ou celulares frente à presença de contaminantes (Van Der Oost; Beyer; Vermeulen, 2003).

Dentre os principais biomarcadores bioquímicos, pode-se citar as enzimas de biotransformação, que tornam os xenobióticos mais hidrofílicos, e, portanto, mais facilmente excretáveis pelos organismos. A biotransformação é dividida em três fases: Fase I, em que ocorrem reações de oxidação, redução e hidrólise, como é o caso das reações envolvendo enzimas da superfamília do Citocromo P450 (CYPs); Fase II, em que ocorrem reações de conjugação, como é o caso das reações envolvendo Glutathione-S-Transferases (GSTs) e Sulfotransferases (SFs) e Fase III, em que ocorrem reações de transporte ativo de xenobióticos transmembrana, como é o caso das reações envolvendo os transportadores ABC (Omiecinski *et al.*, 2010). Em soma às enzimas de biotransformação, existem as enzimas antioxidantes, as quais defendem o organismo contra espécies reativas de oxigênio e são importantes na detoxificação de radicais a moléculas não reativas (Van Der Oost; Beyer; Vermeulen, 2003). Dentre elas, se podem citar as Superóxido Dismutases (SODs), Catalases (CATs) e Glutathione Redutase (GR).

Os biomarcadores moleculares podem indicar variações ecotoxicológicas ou estresse ambiental em estágios iniciais, agindo sinergicamente e aumentando a sensibilidade e especificidade dos biomarcadores convencionais de atividade enzimática. Para isso, técnicas como a reação em cadeia da polimerase quantitativa em tempo real (qPCR) tornam essa investigação possível, podendo detectar mRNA de genes transcritos em níveis muito baixos (Panti *et al.*, 2011).

Ademais, o sistema imunológico desempenha um papel crucial na resposta dos organismos aquáticos à contaminação, sendo uma barreira essencial contra agentes infecciosos e contaminantes presentes no ambiente. O sistema imunológico é frequentemente comprometido pela exposição a poluentes antropogênicos, como metais pesados, pesticidas e

microplásticos, levando a imunossupressão ou disfunções imunes. Essas alterações podem resultar em maior suscetibilidade a patógenos, comprometendo a saúde dos indivíduos e potencialmente reduzindo as populações de espécies aquáticas. Assim, biomarcadores imunológicos, como a atividade fagocítica, níveis de citocinas, como as interleucinas e alterações em linfócitos, são ferramentas valiosas para monitorar os impactos ecotoxicológicos, principalmente na pele, a primeira barreira de proteção do organismo (Kataoka e Kashiwada, 2021).

1.3 A ECOTOXICOGENÔMICA

O transcriptoma representa o perfil de expressão gênica, em uma célula, tecido ou organismo, em um determinado tempo. Ele consiste de todas as moléculas de RNA, incluindo as moléculas de RNA mensageiro, as quais constituem os componentes essenciais para traduzir o DNA em aminoácidos, para formar proteínas servidas de funções fisiológicas ou responsivas a estímulos externos, como os xenobióticos (Schirmer *et al.*, 2010). Esse método pode ser realizado através de três diferentes abordagens: montagem com base em um genoma de referência, montagem *de novo* e uma combinação das duas estratégias. A montagem *de novo* identifica sobreposições entre sequências de leituras curtas (*reads*), de tamanho definido em *k-mer*, resultando na construção dos transcritos originais. (Martin e Wang, 2011). A ecotoxicogenômica permite caracterizar padrões de expressão gênica que podem ser utilizados para identificar contaminação aquática afetando as espécies marinhas. Além disso, essa ciência contribui no desenvolvimento de novos biomarcadores moleculares e na possível determinação de novos genes alvo (Piña e Barata, 2011).

1.4 OS MAMÍFEROS MARINHOS

Os mamíferos marinhos são espécies sentinelas no contexto de monitoramento da contaminação aquática, uma vez que são animais longevos, de topo de cadeia e com quantidade significativa de gordura corporal, logo, acumulam grandes quantidades de toxinas (Borrell, 1993). A baleia jubarte, *Megaptera novaeangliae*, pertence a subordem dos mysticetos, cetáceos que possuem barbatanas, órgão filtrador que tem a função de reter os alimentos, enquanto a água é expelida pelas laterais da boca (**Figura1**).

Figura 1: *Megaptera novaeangliae*



Fonte: Martin Prochazkacz / Shutterstock.com

A maioria das populações de baleias jubarte são migratórias, reproduzindo-se em águas mais quentes, de baixas latitudes durante o inverno e alimentando-se em águas mais frias durante o verão. Nas áreas de criação, as baleias-jubarte frequentemente se concentram próximas à costa, ilhas e bancos de recifes. No Brasil, o principal local de reprodução da espécie se encontra no Arquipélago de abrolhos. A proximidade com a costa garante águas rasas e proteção contra predadores, oferecendo condições ideais para o nascimento e amamentação dos filhotes (Gonçalves *et al.*, 2018) (**Figura 2**):

Figura 2: Rota migratória da baleia jubarte no Brasil



Fonte: Portal Arquipélago Ilhabela

Dessa forma, a espécie se torna uma ótima candidata do monitoramento, já que seus representantes são predadores e filtradores, o que possibilita um maior contato com os contaminantes, não só através da cadeia trófica, mas também, da filtração da água (Jefferson *et al.*, 2011). Uma pesquisa bibliográfica realizada em 1997 revelou que 45% dos estudos de contaminação e de biomarcadores em animais marinhos foram conduzidos de forma destrutiva, utilizando cetáceos obtidos através da caça ou sacrificados para o propósito da pesquisa (Fossi e Marsili, 1997). Em contrapartida, a obtenção de amostras, como órgãos viscerais, de mysticetos que encalham na praia, geralmente é inviável, em função da raridade desses eventos e do estado avançado de decomposição em que a carcaça se encontra (Wiley *et al.*, 1995).

Posto isso, o monitoramento pode ser feito através do uso de biomarcadores não destrutivos associados a investigações avançadas, a nível molecular, visando a preservação desses animais, que, segundo a União Internacional para a Conservação da Natureza (IUCN), estão na lista de espécies ameaçadas, na categoria pouco preocupante (Cooke, 2018). Apesar de sua classificação atual ser menos alarmante, é imprescindível continuar os esforços para a conservação da espécie.

Nesse sentido, o uso de técnicas moleculares, como a análise da expressão gênica em biópsias de pele, tem se mostrado uma ferramenta valiosa. Diversos estudos já identificaram esses biomarcadores em cetáceos. Por exemplo, Righetti e colaboradores, em 2019, trabalharam com Golfinhos-nariz-de-garrafa, *Tursiops truncatus*; Fossi e colaboradores, em 2010 e 2014, trabalharam com a Baleia-fin, *Balaenoptera physalus* e com o Golfinho-riscado, *Stenella coeruleoalba*, respectivamente. Jia e colaboradores, em 2015, trabalharam com o Golfinho-corcunda-indopacífico, *Sousa chinensis*; Noel e colaboradores, em 2014, trabalharam com belugas, *Delphinapterus leucas*. Todos esses estudos demonstraram uma regulação positiva entre níveis de contaminantes e genes do metabolismo de xenobióticos. Isso se comprova uma alternativa eficaz e não destrutiva para o monitoramento da contaminação aquática. Portanto, o presente estudo busca realizar a montagem e anotação do transcriptoma tegumentar de *Megaptera novaeangliae*, visando identificar a transcrição de genes na pele da baleia jubarte, envolvidos no metabolismo de xenobióticos, por meio de um método de monitoramento menos invasivo. Ainda, essa pesquisa abre portas para áreas além da ecotoxigenômica, tais como genética, imunologia, evolução, entre outros campos de conhecimento, principalmente devido ao fato da baleia jubarte se tratar de um animal não modelo.

2 OBJETIVOS

2.1 OBJETIVO GERAL

Caracterizar o transcriptoma tegumentar da baleia jubarte, *Megaptera novaeangliae* e identificar níveis transcricionais nas vias de biotransformação, antioxidantes e relacionadas a desregulação endócrina.

2.2 OBJETIVOS ESPECÍFICOS

- 1- Realizar a montagem *de novo* do transcriptoma tegumentar de *Megaptera novaeangliae*, via Illumina NextSeq2000;
- 2- Avaliar a qualidade da montagem;
- 3- Realizar a anotação funcional do transcriptoma montado por homologia utilizando os bancos de dados públicos;
- 4- Selecionar transcritos de genes envolvidos em processos de biotransformação de xenobióticos e transcritos utilizados tradicionalmente como biomarcadores.

3 METODOLOGIA

3.1 A OBTENÇÃO DE AMOSTRAS

A obtenção de amostras é proveniente do Projeto de Monitoramento de Cetáceos na Bacia de Santos (PMC-BS), através de biópsias de pele retiradas usando dardos com ponteiros de 40 milímetros disparados com balestras de 150 Bar de pressão. Os dardos foram disparados a uma distância mínima de 4 metros do animal, e apontados para o meio do dorso, logo abaixo da nadadeira dorsal (**Figura 3**). O projeto foi concebido e é executado pela empresa de consultoria ambiental SOCIOAMBIENTAL para atender às demandas dos Processos de Licenciamento Ambiental da PETROBRAS, na Bacia de Santos. Os objetivos do projeto consistem na avaliação dos potenciais impactos da exploração, produção e escoamento de petróleo e gás natural sobre os cetáceos, na Bacia de Santos, a longo prazo. A área de abrangência se estende desde o município de Cabo Frio (RJ) a Florianópolis (SC) (Petrobras, 2020).

Figura 3: Coleta de biópsias com balestra em uma baleia jubarte durante o PMC-BS.



Fonte: Socioambiental Consultores Associados

Após a amostragem dos animais, as biópsias de pele são congeladas em nitrogênio líquido e armazenadas sob um código de biópsia. O material é encaminhado ao Laboratório de Biomarcadores de Contaminação Aquática e Imunoquímica (LABCAI), onde vai receber uma nova nomenclatura interna para as análises de biomarcadores bioquímicos e moleculares e vai ser armazenado em freezers a -80°C . Das amostras que chegaram ao laboratório, foram selecionadas seis para extração do RNA total, as quais obtiveram os códigos de biópsias 51, 52, 151, 157, 183 e 184. No LABCAI, foram atribuídos a elas os códigos MN04, MN03, MN11, MN20, MN21 e MN22, respectivamente (**Tabela 1**).

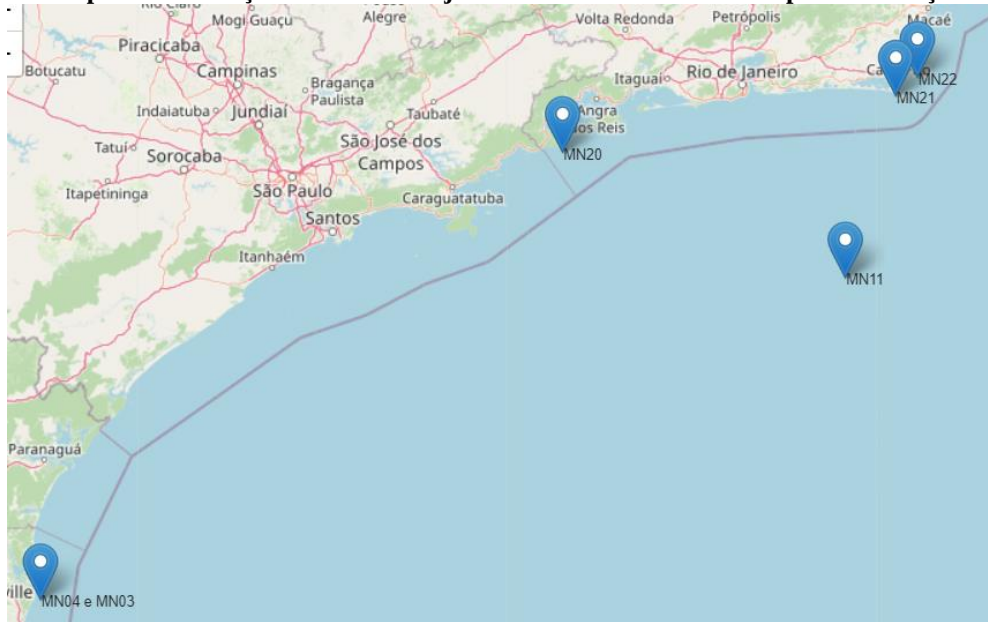
Tabela 1: Dados sobre a coleta das amostras de baleia jubarte para a extração de RNA

Biópsia	Código LABCAI	Data de Coleta	Data de recebimento	Sexo	Coordenadas geográficas	
					Latitude	Longitude
51	MN04	10/08/2016	05/12/2016	Macho	-26.48488	-48.55755
52	MN03	10/08/2016	05/12/2016	Macho	-26.48488	-48.55755
151	MN11	23/07/2017	10/10/2017	Macho	-24.26516	-42.4251
157	MN20	24/07/2017	10/10/2017	Macho	-23.39408	-44.5853
183	MN21	09/11/2017	02/03/2018	Fêmea	-22.99863	-42.04915
184	MN22	11/11/2017	02/03/2018	Macho	-22.86226	-41.88434

Fonte: acervo da autora

Essas amostras foram coletadas de baleias jubarte ao longo da costa brasileira conforme apresentado na **Figura 4**.

Figura 4: Mapa da localização das baleias cujas amostras foram utilizadas para à extração de RNA



Fonte: acervo da autora

3.2 EXTRAÇÃO DO RNA

O RNA tegumentar total foi extraído com o kit RNeasy Fibrous Tissue (Qiagen) seguindo as recomendações do fabricante. A concentração (260 nm) e pureza do RNA obtido foi avaliada por meio das razões 260/280 (qualidade aceitável $> 1,8$) e 260/230 (qualidade aceitável $> 1,7$) em espectrofotômetro NanoDrop® One (Thermo Fisher Scientific®) (Wieczorek; Delauriere; Schagat, 2012). Em seguida, a integridade foi avaliada através de eletroforese em gel de agarose (1,2%). As amostras com RNA de melhor qualidade foram enviadas para o Laboratório Central de Tecnologias de Alto Desempenho em Ciências da Vida (LaCTAD), da UNICAMP, para análise no instrumento TapeStation system, utilizando o RNA Screen Tape (Agilent Technologies). Essa tecnologia realiza uma eletroforese capilar e calcula o RIN (RNA Integrity Number), um valor entre 1 e 10, indicando a qualidade do RNA, em que 10 é o RNA de melhor qualidade (Schroeder, 2006).

3.3 SEQUENCIAMENTO DO TRANSCRIPTOMA

O LaCTAD realizou a construção da biblioteca de cDNA para o sequenciamento do transcriptoma. A biblioteca de cDNA foi preparada, segundo o protocolo do kit TruSeq

Stranded mRNA Library Prep (Illumina, Inc). A qualidade da biblioteca foi validada através do instrumento Bioanalyzer 2100 (Agilent Technologies).

A quantificação da biblioteca foi realizada através da técnica de qPCR, seguindo instruções do kit Sequencing Library qPCR Quantification (Illumina Inc.). Em seguida, foi diluída à concentração de 10 pMolar e então clusterizadas no cBot (Illumina Inc.), onde ocorre a amplificação dos fragmentos de cDNA, capturados por complementariedade aos oligonucleotídeos adaptadores ligados à superfície da flow cell. O sequenciamento do transcriptoma foi realizado no sequenciador de próxima geração Illumina NextSeq2000 (Illumina, Inc), com *reads* pareados PE 2 x 100 pb (100 milhões de reads ou 50 milhões de pares por amostra).

Todas as sequências (*reads*) obtidas foram depositadas para acesso público no serviço SRA (*Sequence Read Archive*) do NCBI (*National Center for Biotechnology Information*) sob o BioProject ID: PRJNA1182765. Em seguida, as leituras do sequenciamento foram avaliadas quanto à qualidade através do programa FASTQC (Versão 0.11.9) (Andrews, 2010). O programa Trimmomatic (Versão 0.39) (Bolger *et al.*, 2014) foi utilizado para eliminar leituras de sequenciamento de tamanho menores que 50 pb, *reads* de baixa qualidade ou cortes em posições nas leituras cuja qualidade foi menor que 30 para o valor de Phred, e ainda, a eliminação de possíveis adaptadores. Uma nova avaliação da qualidade foi realizada no FASTQC para confirmar a filtragem dos *reads*.

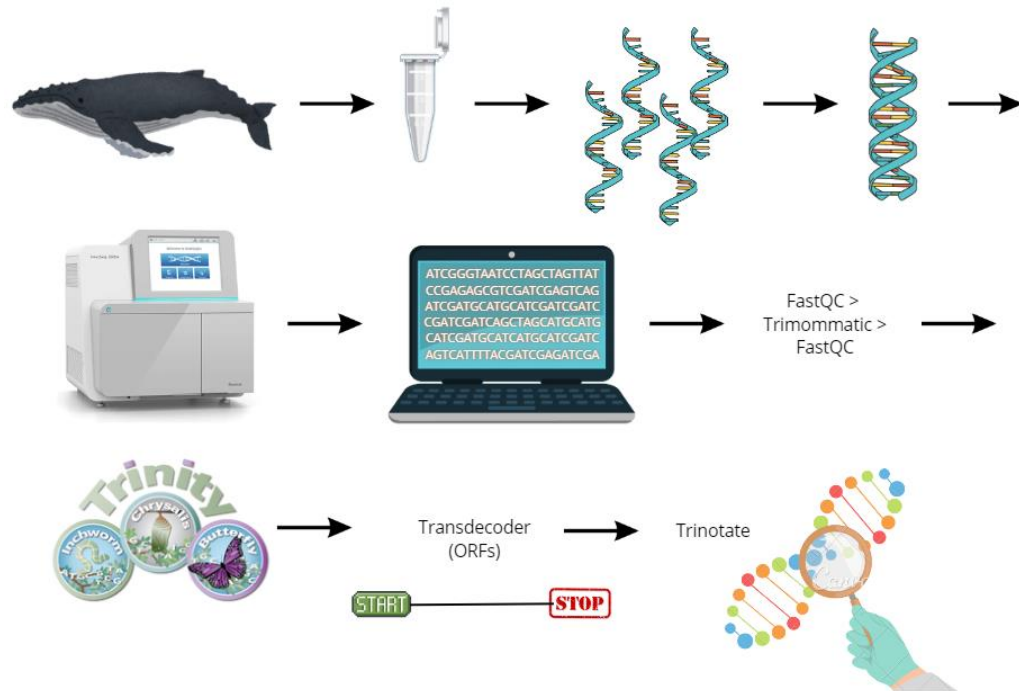
3.4 MONTAGEM E ANOTAÇÃO

A montagem *de novo* foi realizada utilizando o programa Trinity RNA-seq (Haas *et al.*, 2013) (Versão 2.15.1). Foram utilizadas apenas as leituras de alta qualidade resultantes das etapas anteriores para este procedimento. O tamanho de k-mer foi mantido padrão para a metodologia, em 25. As leituras foram normalizadas para posições com cobertura superior a 50, para melhoria de performance, sem prejuízo em relação aos resultados. A abundância dos transcritos foi estimada utilizando o método kallisto (Bray, 2016), disponível no pacote Trinity RNA-seq e estes foram normalizados em Transcritos Por Milhão (TPM). A qualidade da montagem foi verificada utilizando métricas como o número total de contigs ou transcritos, tamanho médio, valor de N50 e Ex90N50, executadas dentro do pacote Trinity. As janelas abertas de leitura (*open reading frames* ou ORFs) foram obtidas através do programa TransDecoder (Versão 5.0.1-3). As métricas das montagens foram acessadas por um *script* fornecido em conjunto com o pacote Trinity RNA-seq. Além disso, foi feita uma análise da

qualidade da montagem do transcriptoma através da ferramenta BUSCO (*Benchmarking Universal Single-Copy Orthologs*) (Waterhouse, 2018) (Versão 5.7.0), uma ferramenta de bioinformática amplamente utilizada para avaliar a completude de genomas ou transcriptomas. Ele utiliza conjuntos de genes conservados, conhecidos como *benchmarking sets* (conjuntos de referência), como o BUSCO ou OrthoDB, para comparar e determinar quais genes essenciais estão presentes no conjunto de dados analisado. Isso ajuda a avaliar a qualidade da montagem ou anotação, indicando quão bem um genoma ou transcriptoma captura genes conservados esperados para um grupo taxonômico específico. Foi feita uma busca por genes ortólogos tanto para Metazoa, quanto para Cetacea. No caso dos cetáceos, foi utilizado o conjunto de dados para a linhagem dos Cetartiodáctilos, disponibilizada pelo programa, clado que abrange os cetáceos e os artiodáctilos.

Os conjuntos de transcritos foram anotados funcionalmente utilizando o pacote Trinotate (Bryant *et al.*, 2017) (Versão 4.0.2), utilizando apenas as maiores ORFs encontradas. Nesta metodologia, os transcritos são anotados a partir de diversas bases de dados como SwissProt (Bairoch; Apweiler, 2000) e Pfam (Bateman *et al.*, 2004). Os resultados foram obtidos através de comparações entre os transcritos obtidos e os bancos de dados, no qual os programas que foram utilizados para tal foram o BLAST+ (Camacho *et al.*, 2009), com uma stringência de e-value de $1e-5$, além de HMMER (Finn *et al.*, 2011) e infernal (Nawrocki *et al.*, 2009). Todos os resultados foram armazenados em estrutura de banco de dados relacional formato SQL (*Structured Query Language*). Desta forma, através de referenciamento cruzado entre outros bancos de dados, como o *Gene Ontology* (Ashburner *et al.*, 2000), o KEGG (Kanehisa, 2000) e o EggNOG (Jensen *et al.*, 2008), o banco SQL foi complementado com informações funcionais sobre os transcritos. O *gene ontology* é uma ferramenta que padroniza a descrição das funções dos genes e seus produtos em diferentes organismos. Ele é dividido em três categorias principais: Função Molecular, que descreve as atividades específicas de um produto gênico; Processo Biológico, que abrange os processos nos quais o gene está envolvido; e Componente Celular, que indica onde na célula o gene exerce sua função. A **Figura 5** representa de forma esquemática a metodologia utilizada para realizar a montagem e anotação funcional do transcriptoma tegumentar da baleia jubarte.

Figura 5: Representação esquemática da metodologia utilizada



Fonte: acervo da autora

4 RESULTADOS E DISCUSSÃO

4.1 ANÁLISE DA QUALIDADE DO RNA

As amostras extraídas foram submetidas ao espectrofotômetro NanoDrop® One (Thermo Fisher Scientific®), para análise de qualidade. A razão 260/280 é utilizada para avaliar a pureza do RNA em relação à presença de proteínas. O RNA absorve luz a 260 nm, enquanto as proteínas absorvem a um comprimento de onda de 280 nm. Um valor ideal para a razão 260/280 está entre 1.8 e 2.0. Valores abaixo de 1.8 indicam que pode haver contaminação por proteínas ou outros compostos orgânicos, que podem prejudicar as análises moleculares. Em contrapartida, valores acima de 2.0 podem sugerir a presença de RNA degradado, que pode afetar negativamente os resultados experimentais. A razão 260/230 é utilizada para avaliar a presença de contaminantes como sais, fenol ou outros compostos que podem ser utilizados durante o processo de extração do RNA. Assim como na razão 260/280, o RNA também absorve luz a 260 nm, enquanto os contaminantes geralmente absorvem a 230 nm. Um valor adequado para a razão 260/230 varia entre 2.0 e 2.2. Valores abaixo de 2.0 podem indicar que há contaminantes presentes que podem interferir na eficácia das reações posteriores. Sendo, assim, foram selecionadas apenas as amostras que apresentaram valores de razões 260/280 e

260/230 adequados, ou seja, MN03, MN11 e MN21, indicando amostras sem contaminação e concentração suficiente para o sequenciamento do transcriptoma. (**Tabela 2**):

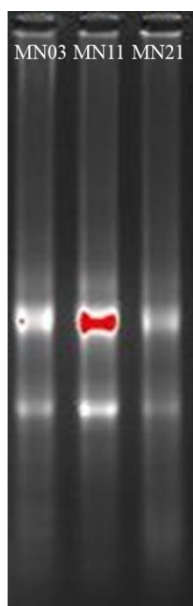
Tabela 2: Análise da concentração e pureza do RNA em espectrofotômetro

Biópsia	Código LABCAI	Concentração RNA ng/μL	A260/A280	A260/A230
51	MN04	18,09	1,98	1,30
52	MN03	979,82	2,09	2,17
151	MN11	1005,68	2,09	2,01
157	MN20	10,23	1,49	0,44
183	MN21	581,5	2,07	2,09
184	MN22	27,79	2,03	1,20

Fonte: acervo da autora

Em seguida, as amostras MN03, MN11 e MN21 foram submetidas à eletroforese em gel de agarose 1,2% com formaldeído (**Figura 6**). Todas as amostras se mostraram íntegras, com as duas bandas evidenciando as duas subunidades 28S e 18S do RNA ribossomal de eucariotos e foram enviadas para o LacTAD junto a Universidade Estadual de Campinas (Unicamp).

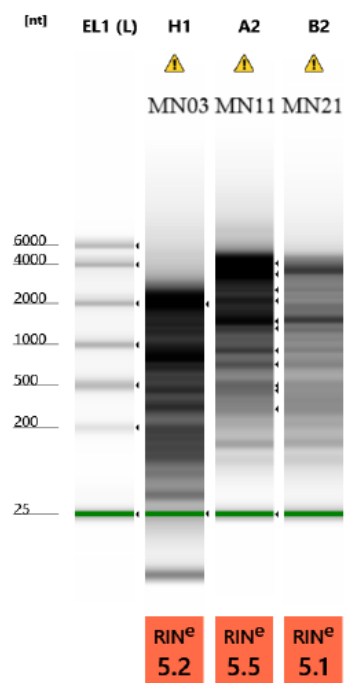
Figura 6: Gel de agarose com formaldeído 1,2% com as bandas 18S e 28S do RNA ribossomal



Fonte: acervo da autora

Por fim, as amostras foram analisadas através do instrumento TapeStation system, utilizando o RNA Screen Tape (Agilent Technologies). Números de Integridade de RNA (RINs) iguais ou superiores a oito geralmente são considerados bons parâmetros e podem ser usados para a subsequente construção de bibliotecas e análise de dados transcriptômicos. As amostras enviadas obtiveram RINs de 5,2; 5,5 e 5,1 e a amostra escolhida foi a com RIN de 5,1 (MN21), pois ela apresentou melhor perfil eletroforético em bandas mais definidas para os padrões de 28S e 18S (**Figura 7**).

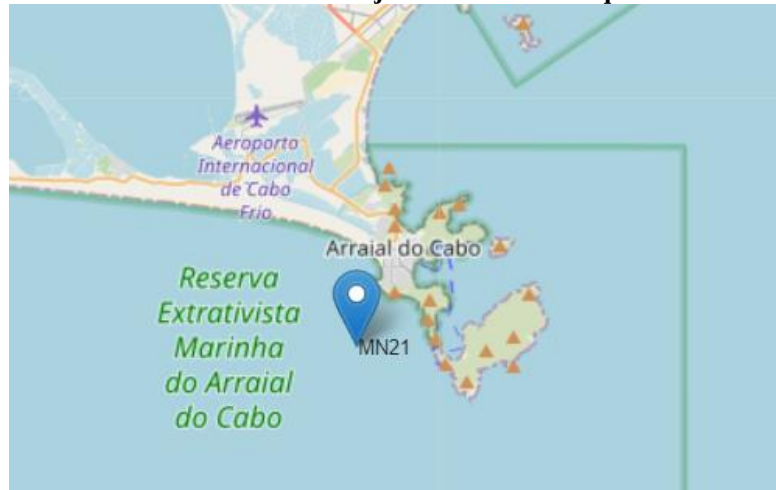
Figura 7: Perfil eletroforético de RNA TapeStation system das amostras de tecido tegumentar de *M. novaeangliae* e seus respectivos valores de RIN



Fonte: acervo da autora

Apesar desta não ser a melhor condição para o sequenciamento, optou-se por seguir com as análises tendo em vista a raridade das amostras. Além disso, estudos como o realizado por Lu, Zhou e Chen, em 2022, demonstraram que amostras de RNA degradadas, com RIN > 4,4, variaram pouco quanto à expressão de mRNA, quando comparadas a amostras de RNA intactos. Essa amostra representa uma fêmea, coletada em novembro de 2017, na Reserva Extrativista Marinha do Arraial do Cabo no estado do Rio de Janeiro (**Figura 8**).

Figura 8: Local onde a amostra de baleia jubarte usada no sequenciamento foi coletada



Fonte: acervo da autora

4.2 ANÁLISE DOS *READS* BRUTOS

A qualidade dos *reads* foi avaliada através do programa FastQC (Versão 0.11.9). Para as sequências senso, o total de *reads* foi de 93.356.722 e a porcentagem de guaninas e citosinas foi de 50%. A qualidade das sequências por base revelou excelentes notas de *phred*, em geral, acima de 30, exceto nas extremidades (**Figura 9**).

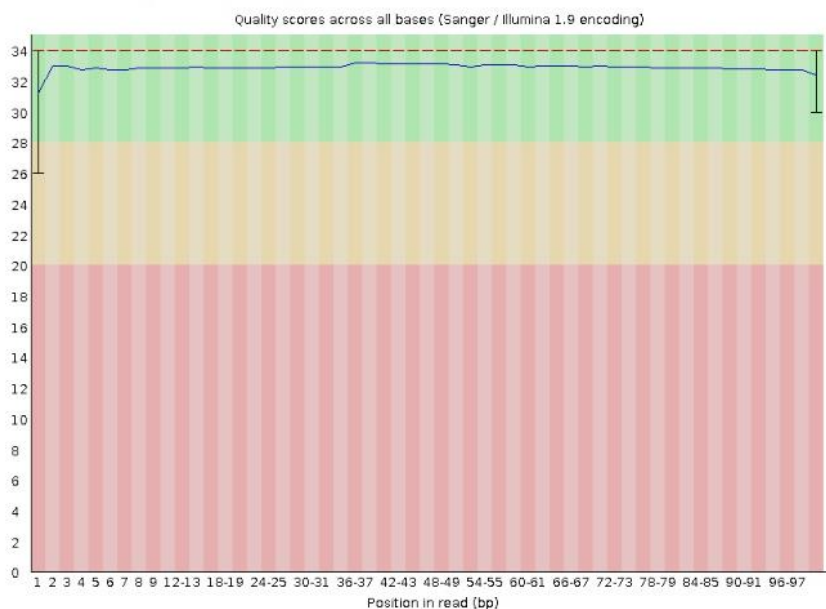
Figura 9: Histograma da qualidade das sequências brutas (senso) avaliado pelo FastQC



Fonte: acervo da autora

Já para as sequências antissenso, o total de *reads* foi o mesmo, 93.356.722, com nenhuma sequência de baixa qualidade e uma porcentagem de guaninas e citosinas de 51%. A qualidade das sequências por base nas sequências antissenso também obtiveram excelentes notas de *Phred*, em geral acima de 30, exceto nas extremidades (**Figura 10**):

Figura 10: Histograma da qualidade das sequências brutas (antissenso) avaliado pelo FastQC



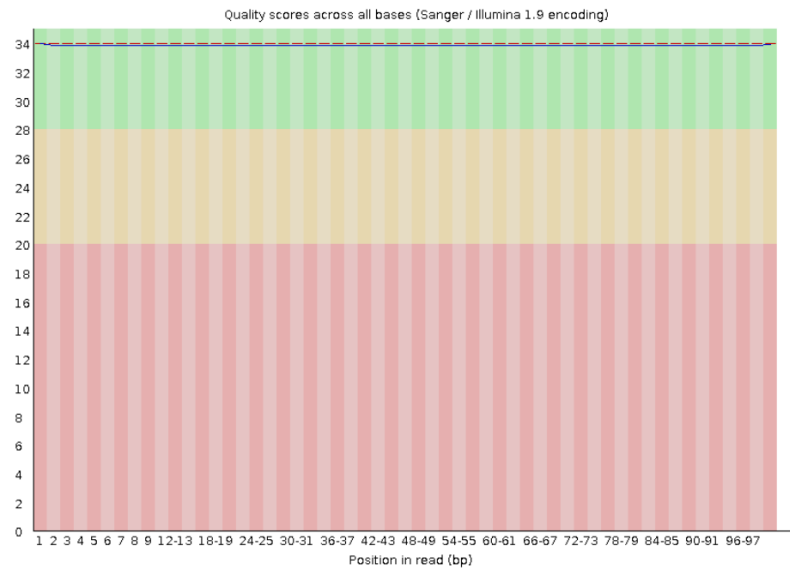
Fonte: acervo da autora

O número de *reads* nas sequências senso e antissenso é o mesmo porque, como se trata de um sequenciamento com leitura pareada, cada fragmento de cDNA é sequenciado nas duas extremidades (5' e 3'). Dessa forma, para cada molécula de RNA original, são geradas duas leituras complementares, uma no sentido da sequência (senso) e outra no sentido oposto (antissenso). Esse pareamento de *reads* garante que o número de leituras em ambas as direções seja equivalente, proporcionando uma cobertura balanceada da sequência, o que contribui para uma análise mais precisa e robusta. Ademais, as extremidades das sequências costumam apresentar notas de *Phred* mais baixas devido à degradação progressiva dos reagentes durante o sequenciamento. Essa degradação resulta em menor precisão na incorporação dos nucleotídeos nas bases finais. Adicionalmente, a presença de adaptadores nas extremidades pode causar erros durante a amplificação, contribuindo ainda mais para a diminuição da qualidade da leitura nestes locais especificamente.

Em seguida, foi feita uma filtragem, eliminando os *reads* com qualidade inferior a 30, através do programa Trimmomatic (Versão 0.39). Os *reads* filtrados foram submetidos ao programa FastQC novamente. Após o filtro, 42.752.635 *reads* foram considerados de alta

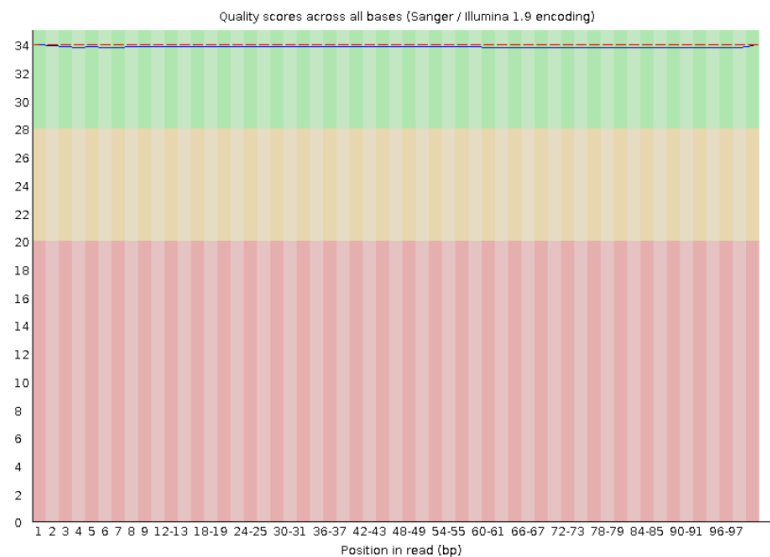
qualidade. Em geral, as sequências resultantes apresentaram excelentes notas de *phred*, por volta de 34 (**Figuras 11 e 12**):

Figura 11: Histograma da qualidade das sequências filtradas (senso) avaliado pelo FastQC



Fonte: Acervo da autora

Figura 12: Histograma da qualidade das sequências filtradas (antissenso) avaliado pelo FastQC



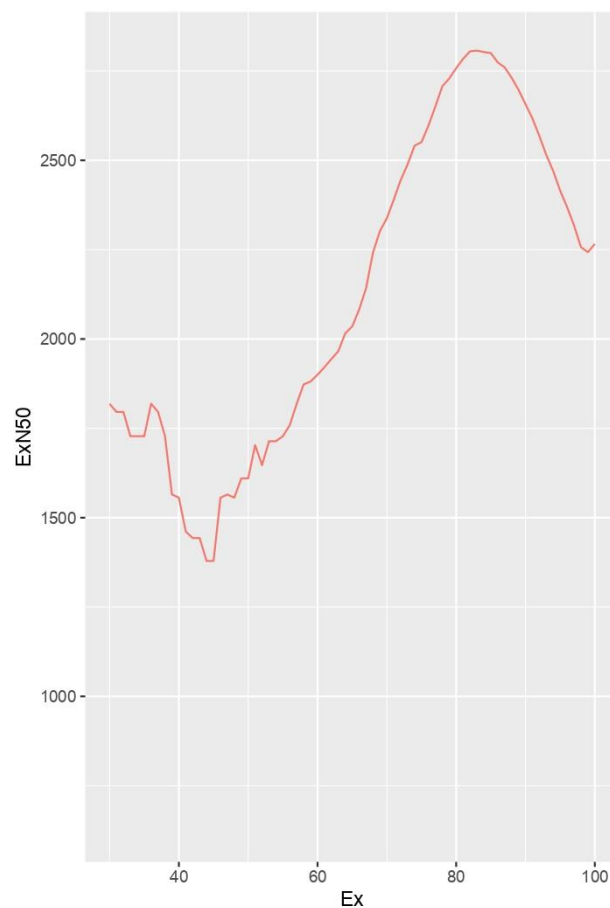
Fonte: acervo da autora

4.3 MONTAGEM DO TRANSCRIPTOMA

A montagem *de novo* do transcriptoma foi realizada através do programa Trinity RNA-seq (Versão 2.15.1), resultando em um total de 124.580 transcritos com uma média de 1.062,88 pb (**Tabela 3**). A métrica N50 indica que ao menos 50% dos nucleotídeos está em

contigs com pelo menos o comprimento do valor de N50, após organizar os contigs em ordem decrescente. Nessa montagem, o valor de N50 foi de 2.266, ou seja, ao menos metade dos nucleotídeos está em contigs com pelo menos 2.266 pb de comprimento. Uma forma mais apropriada para analisar os dados da montagem do transcriptoma é a estatística ExN50, uma vez que ela leva em consideração uma porcentagem específica do conjunto de sequências em vez de apenas metade do conjunto. Geralmente, afirmamos um valor de E90N50, pois ele abrange os transcritos altamente expressos que representam 90% dos transcritos, garantindo uma avaliação mais representativa da montagem (**Figura 13**):

Figura 13: Perfil da expressão Ex contra ExN50



Fonte: acervo da autora

A montagem revelou um valor de E90N50 de 2.657 pb, ou seja, entre os contigs que cobrem pelo menos 90% dos transcritos mais expressos, ao menos metade dos nucleotídeos nesse subconjunto está em contigs com ao menos 2.657 pb de comprimento. Isso indica uma montagem de boa qualidade, uma vez que o valor de Ex90N50 é substancialmente maior do que o tamanho médio dos transcritos, de 1.062,88 pb, e, portanto, sugere que a maioria dos

transcritos mais expressos está bem representada em contigs com comprimentos consideráveis, garantindo uma cobertura robusta e uma representação abrangente do transcriptoma. Para a anotação funcional, foi feita a identificação das ORFs pelo programa Transdecoder. Foram identificadas 134.241 ORFs longas das quais 25.525 são ORFs completas (**Tabela 3**). A discrepância entre o número de ORFs e o número de transcritos se deve à capacidade do TransDecoder de prever múltiplas ORFs a partir de transcritos parciais, isoformas alternativas ou regiões do RNA que ainda não são completamente expressas ou identificadas no conjunto de dados, além da possibilidade de identificar ORFs nos 6 quadros de leitura. Além disso, o TransDecoder pode prever ORFs em regiões não codificantes, artefatos do sequenciamento ou erros na montagem, contribuindo para o aumento do número total de ORFs.

Tabela 3: Métricas da montagem via Trinity

Métricas	Valores
Total de reads brutos	93.356.722
Total de reads filtrados	42.752.625
Conteúdo GC	52%
Total de sequências	124.580
Total de bases montadas	132.413.768
Tamanho médio dos transcritos	1062,88 pb
N50	2.266 pb
Ex90N50	2.657 pb
Total de trinity 'genes':	87.002
Total trinity transcritos:	124.580
Maiores ORFs identificadas	134.241
ORFs completas	25.525

Fonte: acervo da autora

As ORFs completas são sequências nucleotídicas que contêm tanto o códon de iniciação quanto o códon de parada, possibilitando a codificação de uma proteína completa. As ORFs parciais 5' e 3' referem-se a sequências que estão incompletas nas extremidades: a 5' parcial não possui o códon de iniciação, enquanto a 3' parcial é desprovida do códon de parada. Já as ORFs internas são aquelas localizadas dentro de uma sequência maior, sem apresentar os códons de iniciação e de parada (**Tabela 4**):

Tabela 4: Identificação das ORFs completas, parciais 3' e 5', e internas via Transdecoder

Tipo de ORF	Sequências
Completa	78.149
5' parcial	36.343
Interna	9.405
3' parcial	10.344
Total	134.241

Fonte: acervo da autora

A análise da ferramenta BUSCO demonstrou uma montagem de qualidade, tendo em vista que apenas 0,9% das sequências presentes em Metazoa estão faltando na montagem do Trinity, ou seja, a montagem realizada diz respeito a montagem de um transcriptoma de um animal, o que de fato é. Por outro lado, 39,7% das sequências presentes em Cetartiodáctilos estão faltando na montagem do Trinity, o que se mostra razoável, tendo em vista que a amostra estudada se trata apenas do tegumento da baleia e não do organismo inteiro (**Tabela 5**). Ainda, a montagem foi realizada utilizando apenas uma biblioteca, o que pode comprometer a cobertura e profundidade dos genes expressos, limitando a diversidade dos transcritos detectados e dificultando a reprodutibilidade do estudo. Essa abordagem única tende a introduzir viés na representação dos dados, já que a captura da variabilidade genética e expressão gênica pode ser limitada. Idealmente, o uso de múltiplas bibliotecas, aumentando o número de indivíduos na amostragem, permitiria abranger uma variedade genética, contribuindo para uma análise mais robusta e precisa do transcriptoma. Da mesma forma, o uso de múltiplas bibliotecas, a partir de diferentes tecidos de um mesmo animal, como fígado, rins e coração, incluiria não apenas uma maior abrangência de genes expressos, mas também a identificação de isoformas alternativas e genes com expressão diferencial, fornecendo uma visão mais detalhada da complexidade do transcriptoma. Entretanto, cumpre ressaltar a dificuldade de se coletar esses materiais, contrastando com a proposta do monitoramento não destrutivo.

Tabela 5: métricas de montagem via BUSCO

	Trinity	TransDecoder
BUSCO		
Metazoa	n= 954	n= 954
Cópias únicas completas	334 (35%)	474 (49,7%)
Cópias duplicadas	597 (62,6%)	424 (44,4%)
Fragmentadas	14 (1,5%)	28 (2,9%)
Faltando	9 (0,9%)	28 (3%)
BUSCO		
Cetartiodáctilos	n= 13335	n= 13335
Cópias únicas completas	3008 (22,6%)	3633 (27,2%)
Cópias duplicadas completas	4459 (33,4%)	3389 (25,4%)
Fragmentadas	577 (4,3%)	656 (4,9%)
Faltando	5291 (39,7%)	5657 (42,5%)

Fonte: acervo da autora

O elevado número de cópias duplicadas observadas tanto em Metazoa quanto em Cetartiodáctilos pode ser explicado pelo uso de ORFs longas, em vez de ORFs completas, no processo de montagem e anotação funcional. Essa abordagem contribui para a anotação de isoformas diferentes de um mesmo gene, levando a duplicações aparentes. Além disso, a metodologia empregada pelo BUSCO (Benchmarking Universal Single-Copy Orthologs) pode influenciar esses resultados, uma vez que a ferramenta busca ortólogos de cópia única em conjuntos de dados genômicos ou transcriptômicos, podendo identificar variantes isoformas ou duplicatas como cópias adicionais quando a qualidade ou a fragmentação dos dados impacta a análise.

4.4 ANOTAÇÃO FUNCIONAL

A anotação funcional foi realizada utilizando o programa Trinotate (Bryant *et al.*, 2017), com base nas ORFs longas obtidas. Das 124.580 sequências de transcritos montadas pelo Trinity, foram encontrados 52.225 hits significativos com o banco de dados de proteínas SwissProt através da ferramenta BLASTx. O BLASTx traduz as sequências de nucleotídeos em todas as seis possíveis molduras de leitura (três em cada direção) e, em seguida, compara cada tradução com o banco de dados de sequências de proteínas, permitindo a identificação de proteínas homólogas. A ferramenta BLASTp identificou 39.925 hits significativos contra o mesmo banco de dados SwissProt. O BLASTp compara diretamente sequências de proteínas com o banco de dados de outras proteínas, facilitando a anotação funcional baseada em

similaridades proteicas. No que diz respeito à identificação de domínios conservados, foram encontrados 31.780 hits significativos utilizando o software HMMER contra o banco de dados Pfam (*Protein Families Database*). O HMMER emprega modelos de perfil de ocultação (HMMs - *Hidden Markov Models*) para buscar correspondências em um banco de dados que contém múltiplos alinhamentos de sequências de proteínas. Além disso, foram encontrados 47.009 hits significativos contra o banco de dados KEGG (*Kyoto Encyclopedia of Genes and Genomes*). O KEGG é um recurso abrangente que auxilia na compreensão das vias metabólicas, funções e interações biológicas de células, organismos e ecossistemas, com base em informações genômicas e outros dados de alto nível. O banco de dados eggNOG também forneceu 39.003 hits significativos. O eggNOG fornece informações sobre grupos ortólogos e suas funções preditas, o que é útil para a anotação de genes e proteínas dentro de contextos evolutivos. Por fim, foram encontrados 1.284 hits significativos contra o banco de dados Rfam, através do Infernal (*INFERENCE of RNA ALIGNMENT*). O Infernal é um software dedicado à anotação de RNAs estruturados, como RNAs ribossômicos e RNAs de transferência. Utilizando modelos probabilísticos, especificamente Modelos de Covariância (CMs - *Covariance Models*), o Infernal identifica e alinha sequências de RNA com estruturas secundárias conservadas (**Tabela 6**):

Tabela 6: Anotação funcional do transcriptoma tegumentar da baleia jubarte via Trinotate utilizando as maiores ORFs identificadas

Métodos	Hits totais	Hits únicos
Blastx / SwissProt	52.225 (41,92%)	51.364 (41,23%)
Blastp / SwissProt	39.925 (31,05%)	35.602 (28,58%)
HMMER / Pfam	31.780 (25,51%)	27.964 (22,45%)
Infernal / Rfam	47.009 (37,74%)	16.750 (13,45%)
KEGG	1.284 (1,03%)	988 (0,79%)
eggNOG	39.003 (31,31%)	19.196 (15,41%)

Fonte: acervo da autora

Em busca por genes envolvidos no metabolismo de xenobióticos, foi feita uma busca manual por genes anotados funcionalmente via BLASTx. Foram encontradas 20 sequências de 8 isoformas de genes envolvidos na fase 1 de biotransformação, incluindo genes da superfamília do *Citocromo P450* e *Flavinas Monooxygenases* (**Tabela 7**). Foram encontradas 59 sequências de 16 isoformas de genes envolvidos na fase 2 de biotransformação, incluindo *glutathione S-transferases* e *sulfotransferases* (**Tabela 8**). Foram encontradas 91 sequências de 19 isoformas de genes envolvidos na fase 3 de biotransformação, como os genes pertencentes aos *transportadores do tipo ABC* (**Tabela 9**). Além disso, foram encontradas 42 sequências de 12

isoformas de receptores nucleares, incluindo o *receptor de hidrocarboneto arílico* e *receptor de estrogênio* (**Tabela 10**), 65 sequências de 17 isoformas de genes do sistema antioxidante, incluindo *catalase* e *glutathione redutase* (**Tabela 11**) e 131 sequências de 36 isoformas de genes do sistema imune, como *interleucinas* e *receptores toll-like* (**Tabela 12**).

Tabela 7: Genes envolvidos na biotransformação de fase 1

Função Biológica	Gene
Biotransformação Fase 1	<i>Cytochrome P450 1B1 (CYP1B1)</i>
	<i>Cytochrome P450 2C15 (CYP2C15)</i>
	<i>Cytochrome P450 2J2 (CYP2J2)</i>
	<i>Cytochrome P450 2S1 (CYP2S1)</i>
	<i>Cytochrome P450 2U1 (CYP2U1)</i>
	<i>Cytochrome P450 3A12 (CYP3A12)</i>
	<i>Cytochrome P450 3A5 (CYP3A5)</i>
	<i>Dimethylaniline monooxygenase (FMO4)</i>

Fonte: acervo da autora

Tabela 8: Genes envolvidos na biotransformação de fase 2

Função Biológica	Gene
Biotransformação Fase 2	<i>Glutathione S-transferase 3 (GST3)</i>
	<i>Glutathione S-transferase A2 (GSTA2)</i>
	<i>Glutathione S-transferase A4 (GSTA4)</i>
	<i>Glutathione S-transferase kappa (GSTK1)</i>
	<i>Glutathione S-transferase Mu 1 (GSTM1)</i>
	<i>Glutathione S-transferase Mu 3 (GSTM3)</i>
	<i>Glutathione S-transferase Mu 4 (GSTM4)</i>
	<i>Glutathione S-transferase omega-1 (GSTO1)</i>
	<i>Glutathione S-transferase omega-2 (GSTO2)</i>
	<i>Glutathione S-transferase P (GSTP1)</i>
	<i>Glutathione S-transferase theta-1 (GSTT1)</i>
	<i>Glutathione S-transferase theta-3 (GSTT3)</i>
	<i>Glutathione S-transferase theta-4 (GSTT4)</i>
	<i>Sulfotransferase 1A1 (SULT1A1)</i>
	<i>Sulfotransferase 2A1 (SULT2A1)</i>
	<i>Sulfotransferase 2B1 (SULT2B1)</i>

Fonte: acervo da autora

Tabela 9: Genes envolvidos na biotransformação de fase 3

Função Biológica	Gene
Transportadores ABC	<i>ABC-type oligopeptide transporter ABCB9 (ABCB9)</i>
	<i>ATP-binding cassette sub-family A member 2 (ABCA2)</i>
	<i>ATP-binding cassette sub-family A member 6 (ABCA6)</i>
	<i>ATP-binding cassette sub-family B member 1 (ABCB1)</i>
	<i>ATP-binding cassette sub-family B member 10 (ABCB10)</i>
	<i>ATP-binding cassette sub-family B member 6 (ABCB6)</i>
	<i>ATP-binding cassette sub-family C member 10 (ABCC10)</i>
	<i>ATP-binding cassette sub-family C member 3 (ABCC3)</i>

ATP-binding cassette sub-family C member 4 (ABCC4)
ATP-binding cassette sub-family C member 5 (ABCC5)
ATP-binding cassette sub-family C member 6 (ABCC6)
ATP-binding cassette sub-family D member 1 (ABCD1)
ATP-binding cassette sub-family D member 3 (ABCD3)
ATP-binding cassette sub-family E member 1 (ABCE1)
ATP-binding cassette sub-family F member 1 (ABCF1)
ATP-binding cassette sub-family F member 2 (ABCF2)
ATP-binding cassette sub-family F member 3 (ABCF3)
ATP-binding cassette sub-family G member 1 (ABCG1)
Multidrug resistance-associated protein 1 (ABCC1)

Fonte: acervo da autora

Tabela 10: Genes de receptores nucleares

Função Biológica	Gene
Receptores nucleares	<i>Aryl hydrocarbon receptor (AhR)</i>
	<i>Aryl hydrocarbon receptor repressor (AhRR)</i>
	<i>Steroid hormone receptor (ERR1)</i>
	<i>Estrogen receptor (ESR1)</i>
	<i>Peroxisome proliferator-activated receptor alpha (PPARA)</i>
	<i>Peroxisome proliferator-activated receptor delta (PPARD)</i>
	<i>Peroxisome proliferator-activated receptor gamma (PPARG)</i>
	<i>Retinoic acid receptor RXR-alpha (RXRA)</i>
	<i>Retinoic acid receptor RXR-beta (RXRB)</i>
	<i>Oxysterols receptor LXR-alpha (NR1H3)</i>
	<i>Oxysterols receptor LXR-beta (NR1H2)</i>
	<i>Growth hormone receptor (GHR)</i>
	Fonte: acervo da autora

Tabela 11: Genes do sistema antioxidante

Função Biológica	Gene
Sistema antioxidante	<i>Catalase (CAT)</i>
	<i>Glutathione reductase (GSR)</i>
	<i>Glutathione peroxidase 1 (GPX1)</i>
	<i>Glutathione peroxidase 3 (GPX3)</i>
	<i>Phospholipid hydroperoxide glutathione peroxidase (GPX4)</i>
	<i>Glutathione peroxidase 7 (GPX 7)</i>
	<i>Superoxide dismutase [Cu-Zn] (SOD1)</i>
	<i>Superoxide dismutase [Mn] (SOD2)</i>
	<i>Extracellular superoxide dismutase [Cu-Zn] (SOD 3)</i>
	<i>Peroxiredoxin-1 (PRDX1)</i>
	<i>Peroxiredoxin-2 (PRDX2)</i>
	<i>Peroxiredoxin-4 (PRDX4)</i>
	<i>Peroxiredoxin-5 (PRDX5)</i>
	<i>Peroxiredoxin-6 (PRDX6)</i>
	<i>Thioredoxin reductase 1 (TXNRD1)</i>
	<i>Thioredoxin reductase 2 (TXNRD2)</i>
	<i>Thioredoxin reductase 3 (TXNRD3)</i>
	Fonte: acervo da autora

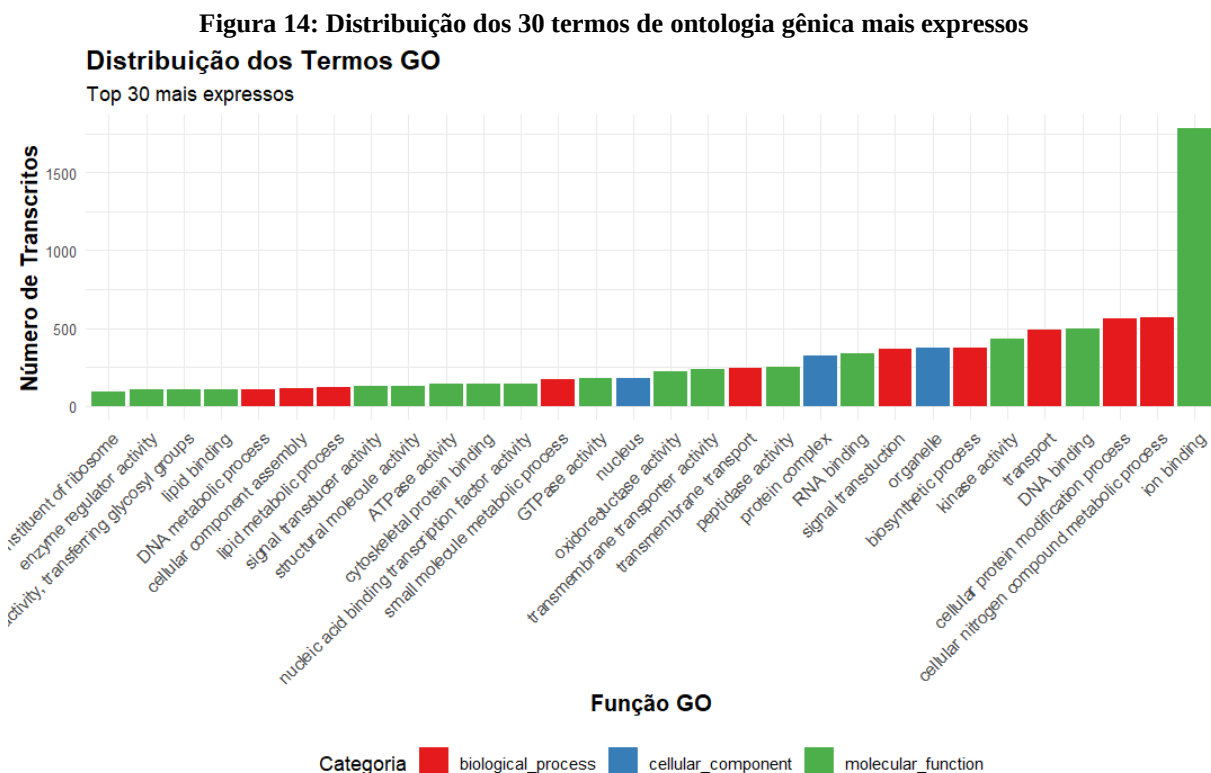
Tabela 12: Genes do sistema imunológico

Função Biológica	Gene
Sistema imune	<i>Complement C2 (C2)</i>
	<i>Complement C3 (C3)</i>
	<i>Complement C4-A (C4A)</i>
	<i>Interleukin-1 receptor type 1 (IL1R1)</i>
	<i>Interleukin-1 receptor type 1 (IL1R2)</i>
	<i>Interleukin-1 receptor-associated kinase 1 (IRAK1)</i>
	<i>Interleukin-1 receptor-associated kinase (IRAK2)</i>
	<i>Interleukin-1 receptor-associated kinase 3 (IRAK3)</i>
	<i>Interleukin-1 receptor-associated kinase (IRAK4)</i>
	<i>Interleukin-4 receptor subunit alpha (IL4R)</i>
	<i>Interleukin-6 receptor subunit alpha (IL6R)</i>
	<i>Interleukin-10 receptor subunit beta (IL10RB)</i>
	<i>Interleukin-11 receptor subunit alpha (IL11RA)</i>
	<i>Interleukin-13 receptor subunit alpha-1 (IL13RA1)</i>
	<i>Interleukin-1 receptor-associated kinase 1-binding protein (IRAK1BP1)</i>
	<i>Interleukin-10 receptor subunit beta (IL10RB)</i>
	<i>Interleukin-17 receptor A (IL17RA)</i>
	<i>Interleukin-17 receptor C (IL17RC)</i>
	<i>Interleukin-17 receptor D (IL17RD)</i>
	<i>Interleukin-17 receptor E (IL17RE)</i>
	<i>Interleukin-17B (IL17B)</i>
	<i>Interleukin-17C (IL17C)</i>
	<i>Interleukin-17D (IL17D)</i>
	<i>Interleukin-18-binding protein (IL28BP)</i>
	<i>Interleukin-20 receptor subunit alpha (IL20RA)</i>
	<i>Interleukin-20 receptor subunit beta (IL20RB)</i>
	<i>Interleukin-22 receptor subunit alpha-1 (IL22RA1)</i>
	<i>Interleukin-27 receptor subunit alpha (IL27RA)</i>
	<i>Interleukin-27 subunit beta (EBI3)</i>
	<i>Interleukin-34 (IL34)</i>
	<i>Interleukin-36 beta (IL36B)</i>
	<i>Toll-like receptor 1 (TLR1)</i>
	<i>Toll-like receptor 2 (TLR2)</i>
	<i>Toll-like receptor 3 (TLR3)</i>
	<i>Toll-like receptor 6 (TLR6)</i>
	<i>Toll-like receptor 9 (TLR9)</i>

Fonte: acervo da autora

Além disso, foi feita uma análise de ontologia gênica a partir dos dados gerados pelo Trinotate, classificando os transcritos como pertencentes a funções biológicas, processos moleculares ou componentes celulares. Das 30 funções mais expressas, foi observado que, aproximadamente, à 6.000 transcritos foram atribuídos funções moleculares, aproximadamente, à 3.500, processos biológicos e, aproximadamente, à 2.000, componentes celulares (**Figura 14**).

Ou seja, a maioria dos transcritos constituem funções moleculares, o que já era esperado tendo em vista que o objeto de estudo é um transcriptoma.



4.5 DESENHO DOS INICIADORES

Como um dos objetivos desse estudo foi a identificação de genes envolvidos no metabolismo de xenobióticos e genes do sistema imunológico, foram desenhados iniciadores para genes utilizados classicamente em trabalhos de monitoramento ambiental, no contexto de desregulação endócrina, da biotransformação de xenobióticos, do estresse oxidativo e de respostas inflamatórias (Sabat, 2010; Costa, 2012 e Serafini, 2024).

Foram escolhidos dois receptores nucleares (*AhR* e *ESR1*), um gene envolvido na fase um de biotransformação (*CYP1B*), um gene envolvido na fase dois de biotransformação (*GSTA2*), um gene envolvido na fase três de biotransformação (*ABCC*), um gene do sistema antioxidante (*CAT*) e um gene pertencente ao sistema imunológico (*IL10*). Os iniciadores foram desenhados com as sequências das janelas abertas de leitura, com o auxílio de ferramentas como *PrimerQuest* e *OligoAnalyzer* (Integrated DNA Technology) (**Tabela 13**):

Tabela 13: Sequências senso (F) e antissenso (R) dos iniciadores e tamanho do produto (*amplicon*) para cada gene

Gene	Sequência (5' – 3')	Amplicon (pb)
<i>AhR</i>	F - TTACGCACAGAACTTTATCC R - TAGAAGAAGGTGTAGGGTATG	119
<i>ESR1</i>	F - GTTTGCTCCTAACTTGCTCCTG R - ATACGGAACCGAGACGATGTAG	102
<i>CYP1B</i>	F - CTGTCTTGGGCTACCACATC R - TGGATCAAAGTCTTCTGGGTTAG	104
<i>GSTA2</i>	F - CAAGATGACCCAGATCAAAG R - CATAGTAGAGAAGTTCAACCAG	143
<i>ABCC1</i>	F - GCTTCAAGATCACCATCATC R - AGACCTCTTCATCCGAATAC	99
<i>CAT</i>	F - GTACTGAATGAGGAGCAGAG R - GACATCGCTGAAGTTCTTAAC	108
<i>IL10</i>	F - TCTTGACGGAATGTGATTTC R - GGACAGAAGGTGATGTTTATC	112

Fonte: acervo da autora

5 CONCLUSÃO

No presente estudo, foi realizada a caracterização do transcriptoma tegumentar da baleia jubarte, *Megaptera novaeangliae* e a anotação funcional. Além disso foi dada ênfase na identificação de transcritos envolvidos no metabolismo de xenobióticos, especialmente nas vias de biotransformação de fase 1 e 2 e receptores nucleares, alvos pré-estabelecidos para o monitoramento ambiental desta espécie no contexto do Projeto de Monitoramento de Cetáceos (PMC). Foram identificados genes pertencentes a todas as fases de biotransformação, dentre eles, genes da superfamília *CYP450*, uma *Flavina Monoxigenase*, *Glutathione-S-transferases*, *Sulfotransferases* e *Transportadores ABC*. Além disso, foram identificados receptores nucleares, como o *Receptor Aril hidrocarboneto* e o *Receptor de Estrogênio*, além de genes do sistema imunológico, como *Sistema complemento*, *Interleucinas* e *Receptores Toll-Like*. Ainda, foram encontrados genes de enzimas antioxidantes como a *Catalase*, *Glutathione Peroxidases*, *Superóxido Dismutases* e *Tiorredoxina Redutases*. Ademais, esta pesquisa será aplicada no monitoramento ambiental, uma vez que o transcriptoma serve como base para o desenho de primers específicos para genes de interesse comumente usados no processo de monitoramento ambiental, com amostras provenientes do Projeto de Monitoramento de Cetáceos (PMC). Esse monitoramento é uma etapa fundamental para o licenciamento ambiental da Petrobras,

permitindo a exploração de petróleo e gás natural de maneira regulada e com o compromisso de minimizar impactos ambientais, conforme as exigências de licenciamento ambiental. Neste estudo, o foco se constituiu na busca por genes relacionados ao metabolismo de xenobióticos e genes do sistema imunológico. No entanto, os dados obtidos e as análises realizadas apresentam um potencial de aplicação que vai além dessa temática específica, podendo ser úteis para pesquisas em outras áreas, como evolução molecular, genética, e imunologia, ampliando o alcance e a relevância dos resultados encontrados, principalmente pelo fato da baleia jubarte se tratar de um animal não modelo.

6 PERSPECTIVAS FUTURAS

Para as próximas etapas, é essencial dar continuidade ao trabalho com a realização da extração de RNA das amostras já disponíveis no laboratório, garantindo a qualidade e a integridade do material biológico. Em seguida, será feita a síntese de cDNA para as análises de qPCR (Reação em Cadeia da Polimerase em Tempo Real). Isso possibilitará uma avaliação quantitativa precisa da expressão gênica. Com os resultados das qPCR, será possível identificar a presença de genes envolvidos no metabolismo de xenobióticos, fornecendo dados relevantes para o monitoramento ambiental e contribuindo para uma maior compreensão dos impactos de contaminantes no ambiente marinho. Além disso, esses dados serão utilizados para a emissão de laudos com informações de biomarcadores moleculares no contexto do Projeto de Monitoramento de Cetáceos.

REFERÊNCIAS

- ANDREWS, S. 2010. FastQC: **a quality control tool for high throughput sequence data**. Disponível online em: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
- ASHBURNER, M. *et al.* Gene ontology: tool for the unification of biology. **Nature genetics**, v. 25, p. 25-29, 2000.
- BAIROCH, A.; APWEILER, R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. **Nucleic Acids Research**, v. 28, p. 45-48, 2000.
- BATEMAN, A. *et al.* The Pfam protein families database. **Nucleic Acids Research**, v. 32, p. D138-D141, 2004.
- BERGMAN, A.; OLSSON, M. Pathology of Baltic grey seal and ringed seal females with special reference to adrenocortical hyperplasia: Is environmental pollution the cause of a widely distributed disease syndrome. **Finnish Game Res**, v. 44, p. 47-62, 1985.
- BOLGER, A.M.; LOHSE, M.; USADEL, B. Trimmomatic: a flexible trimmer for Illumina sequence data. **Bioinformatics**, v. 30, p. 2114-2120, 2014.
- BORGWARDT, F. *et al.* Exploring variability in environmental impact risk from human activities across aquatic ecosystems. **The Science of the total environment**, v. 652, p. 1396–1408, 2019.
- BORRELL, A. PCB and DDTs in blubber of cetaceans from the Northeastern North-Atlantic. **Marine Pollution Bulletin**, v. 26, n. 3, p. 146–151, 1993.
- BREIVIK, K. *et al.* Primary sources of selected POPs: regional and global scale emission inventories. **Environmental pollution (Barking, Essex: 1987)**, v. 128, n. 1–2, p. 3–16, 2004.
- BRYANT, D.M. *et al.* A tissue-mapped axolotl de novo transcriptome enables identification of limb regeneration factors. **Cell Reports**, v. 18, p. 762-776, 2017.
- CAMACHO, C. *et al.* BLAST+: architecture and applications. **BMC bioinformatics**, v. 10, n. 421, 2009.
- COOKE, J. IUCN Red List of Threatened Species: Megaptera novaeangliae. **IUCN Red List of Threatened Species**, 2018. Disponível em: <https://www.iucnredlist.org/species/13006/50362794>. Acesso em: 14 nov. 2024.
- COSTA, J. *et al.* Gene expression analysis of ABC efflux transporters, CYP1A and GST α in Nile tilapia after exposure to benzo(a)pyrene. **Comparative biochemistry and physiology. Toxicology & pharmacology: CBP**, v. 155, n. 3, p. 469–482, 2012.
- FINN, R.D.; CLEMENTS, J.; EDDY, S.R. HMMER web server: interactive sequence similarity searching. **Nucleic Acids Research**, v. 39, p. W29-W37, 2011.

FOSSI, M. C.; MARSILI, L. The use of non-destructive biomarkers in the study of marine mammals. **Biomarkers: biochemical indicators of exposure, response, and susceptibility to chemicals**, v. 2, n. 4, p. 205–216, 1997.

FOSSI, M. C. *et al.* The use of a non-lethal tool for evaluating toxicological hazard of organochlorine contaminants in Mediterranean cetaceans: new data 10 years after the first paper published in MPB. **Marine pollution bulletin**, v. 46, n. 8, p. 972–982, 2003.

FOSSI, M. C. *et al.* A multi-trial diagnostic tool in fin whale (*Balaenoptera physalus*) skin biopsies of the Pelagos Sanctuary (Mediterranean Sea) and the Gulf of California (Mexico). **Marine environmental research**, v. 69 Suppl, p. S17-20, 2010.

FOSSI, M. C. *et al.* An “*ex vivo*” model to evaluate toxicological responses to mixtures of contaminants in cetaceans: Integumentum biopsy slices: “Ex Vivo” Models for Monitoring Toxicological Responses in Cetaceans. **Environmental toxicology**, v. 29, n. 10, p. 1107–1121, 2014.

GONÇALVES, M. I. C. *et al.* Movement patterns of humpback whales (*Megaptera novaeangliae*) reoccupying a Brazilian breeding ground. **Biota neotropica**, v. 18, n. 4, 2018.

HAAS, B.J *et al.*, *De novo* transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. **Nature Protocols**, v. 8, p. 1494-1512, 2013.

HELLE, E.; OLSSON, M.; JENSEN, S. PCB levels correlated with pathological changes in seal uteri. **Ambio**, p. 261-262, 1976.

JEFFERSON, T. A. *et al.* **Marine Mammals of the World: A Comprehensive Guide to Their Identification**. 1. ed. California, USA: Academic Press, 2011.

JENSEN, L. J. *et al.* eggNOG: automated construction and annotation of orthologous groups of genes. *Nucleic acids research*, v. 36, n. Database issue, p. D250-4, 2008.

JEPSON, P. D. *et al.* Relationships between polychlorinated biphenyls and health status in Harbour porpoises (*Phocoena phocoena*) stranded in the United Kingdom. **Environ. Toxicol. Chem**, v. 24, p. 238–248, 2005.

JIA, K. *et al.* In vitro assessment of environmental stress of persistent organic pollutants on the Indo-Pacific humpback dolphin. **Toxicology in vitro: an international journal published in association with BIBRA**, v. 30, n. 1, p. 529–535, 2015.

JONES, K. C.; DE VOOGT, P. Persistent organic pollutants (POPs): state of the science. **Environmental pollution (Barking, Essex: 1987)**, v. 100, n. 1–3, p. 209–221, 1999.

KANEHISA, M.; GOTO, S. KEGG: kyoto encyclopedia of genes and genomes. **Nucleic acids research**, v. 28, n. 1, p. 27–30, 2000.

KANNAN, K. *et al.* Toxicity reference values for the toxic effects of polychlorinated biphenyls to aquatic mammals. **Human and Ecological Risk Assessment**, v. 6, n. 1, p. 181-201, 2000.

- KATAOKA, C.; KASHIWADA, S. Ecological risks due to immunotoxicological effects on aquatic organisms. **International journal of molecular sciences**, v. 22, n. 15, p. 8305, 2021.
- LU, W.; ZHOU, Q.; CHEN, Y. Impact of RNA degradation on next-generation sequencing transcriptome data. *Genomics*, v. 114, n. 4, p. 110429, 2022.
- MARTIN, J. A.; WANG, Z. Next-generation transcriptome assembly. **Nature reviews. Genetics**, v. 12, n. 10, p. 671–682, 2011.
- NAWROCKI, E. P.; KOLBE, D. L.; EDDY, S. R. Infernal 1.0: inference of RNA alignments. **Bioinformatics** (Oxford, England), v. 25, n. 10, p. 1335–1337, 2009.
- NICOLAS, L.; BRAY, H. Páll Melsted and Lior Pachter, Near-optimal probabilistic RNA-seq quantification. **Nature Biotechnology**, v. 34, p. 525–527, 2016.
- NOEL, M. *et al.* PCBs are associated with altered gene transcript profiles in arctic beluga whales (*Delphinapterus leucas*). **Environmental science & technology**, v. 48, n. 5, p. 2942–2951, 2014.
- OMIECINSKI, C. J. *et al.* Xenobiotic metabolism, disposition, and regulation by receptors: from biochemical phenomenon to predictors of major toxicities. **Toxicological sciences: an official journal of the Society of Toxicology**, v. 120 Suppl 1, n. Supplement 1, p. S49-75, 2011.
- PANTI, C. *et al.* Ecotoxicological diagnosis of striped dolphin (*Stenella coeruleoalba*) from the Mediterranean basin by skin biopsy and gene expression approach. **Ecotoxicology (London, England)**, v. 20, n. 8, p. 1791–1800, 2011.
- PETROBRAS. **Plano de Meio Ambiente para a Bacia de Santos**. Revisão 01. 2020. Disponível em: https://comunicabaciadesantos.petrobras.com.br/documents/d/comunica-bacia-de-santos/pe_pmc-bs_rev01_medio_prazo_29-12-2020. Acesso em: 28/10/2024
- PIÑA, B.; BARATA, C. A genomic and ecotoxicological perspective of DNA array studies in aquatic environmental risk assessment. **Aquatic toxicology (Amsterdam, Netherlands)**, v. 105, n. 3–4, p. 40–49, 2011.
- RIEGER, P. G. *et al.* Xenobiotics in the environment: present and future strategies to obviate the problem of biological persistence. **Journal of biotechnology**, v. 94, n. 1, p. 101–123, 2002.
- RIGHETTI, B. P. H. *et al.* Biochemical and molecular biomarkers in integument biopsies of free-ranging coastal bottlenose dolphins from southern Brazil. **Chemosphere**, v. 225, p. 139–149, 2019.
- SABAT, R. *et al.* Biology of interleukin-10. **Cytokine & growth factor reviews**, v. 21, n. 5, p. 331–344, 2010.
- SCHIRMER, K. *et al.* Transcriptomics in ecotoxicology. **Analytical and bioanalytical chemistry**, v. 397, n. 3, p. 917–923, 2010.

SCHROEDER, A. *et al.* The RIN: an RNA integrity number for assigning integrity values to RNA measurements. **BMC molecular biology**, v. 7, n. 1, 2006.

SERAFINI, P. P. *et al.* Biochemical and molecular biomarkers and their association with anthropogenic chemicals in wintering Manx shearwaters (*Puffinus puffinus*). **Marine pollution bulletin**, v. 203, n. 116398, p. 116398, 2024.

SHARMA, B. M. *et al.* Environment and human exposure to persistent organic pollutants (POPs) in India: a systematic review of recent and historical data. **Environment international**, v. 66, p. 48–64, 2014.

UNEP, 2001. Final Act of the Plenipotentiaries on the Stockholm Convention on Persistent Organic Pollutants. **United Nations Environment Program Chemicals**, Geneva, Switzerland, p. 445.

VAN DER OOST, R.; BEYER, J.; VERMEULEN, N. P. E. Fish bioaccumulation and biomarkers in environmental risk assessment: a review. **Environmental toxicology and pharmacology**, v. 13, n. 2, p. 57–149, 2003.

WATERHOUSE, R. M. *et al.* BUSCO applications from quality assessments to gene prediction and phylogenomics. **Molecular biology and evolution**, v. 35, n. 3, p. 543–548, 2018.

WIECZOREK, D.; DELAURIERE, L.; SCHAGAT, T. **Methods of RNA Quality Assessment**. 2012. Disponível em: <https://www.promega.com.br/resources/pubhub/methods-of-rna-quality-assessment/>. Acesso em: 10 out. 2024.

WILEY, D. N.; ASMUTIS, R. A.; PITCHFORD, T. D. Stranding and mortality of humpback whales, *Megaptera novaeangliae*, in the mid-Atlantic and southeast United States, 1985–1992. **Fishery Bulletin**, v. 93, p. 196–205, 1995.