UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”
Instituto de Biociências
Câmpus do Litoral Paulista

Transcriptomic and proteopeptidomic approach for the identification of proteins and peptides with pharmacological potential in the marine gastropod *Olivancillaria urceus*

GABRIEL MARQUES DE BARROS

SÃO VICENTE - SP
2024

UNIVERSIDADE ESTADUAL PAULISTA
“Júlio de Mesquita Filho”

INSTITUTO DE BIOCIÊNCIAS
CÂMPUS DO LITORAL PAULISTA

Transcriptomic and proteopeptidomic approach for the identification of proteins and peptides with pharmacological potential in the marine gastropod *Olivancillaria urceus*

GABRIEL MARQUES DE BARROS
DR. LEANDRO MANTOVANI DE CASTRO

Dissertação apresentada ao Instituto de Biociências, Câmpus do Litoral Paulista, UNESP, para obtenção do título de Mestre no Programa de Pós-Graduação em Biodiversidade de Ambientes Costeiros.

SÃO VICENTE - SP
2024

D286t

DE BARROS, GABRIEL MARQUES
Transcriptomic and proteopeptidomic approach for the identification of proteins and peptides with pharmacological potential in the marine gastropod Olivancillaria urceus / GABRIEL MARQUES DE BARROS. -- São Vicente, 2024
44 p.

Dissertação (mestrado) - Universidade Estadual Paulista (UNESP), Instituto de Biociências, São Vicente
Orientador: LEANDRO MANTOVANI DE CASTRO

1. Biodiversidade. 2. Biologia Molecular. 3. Proteômica. 4. Transcriptoma. 5. Gastropode. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Dados fornecidos pelo autor(a).

RESUMO

Os ecossistemas aquáticos contêm uma grande variedade de espécies, desde microorganismos até grandes mamíferos, e são importantes para a manutenção da biodiversidade, além de benefícios para a espécie humana. No entanto, nosso conhecimento sobre as espécies marinhas é ainda limitado, cerca de 10% do total estimado. A bioprospecção de moléculas com potencial terapêutico em organismos marinhos tem sido um campo de pesquisa crescente nas últimas décadas e a produção de moléculas nestes organismos marinhos está relacionada à sua interação com o ambiente, incluindo interações simbióticas com outros organismos. Estudos demonstraram a correlação entre as características inatas de organismos marinhos, interações e produção de uma variedade de moléculas aprimoradas ao longo da evolução, sendo uma das principais moléculas, os peptídeos, que são considerados moléculas multifuncionais, dado as características de poderem agir sistemicamente como moléculas sinalizadoras, sobre rotas metabólicas e somarem diferentes mecanismos de ação. A atividade de pesca, especialmente a pesca de arrasto de fundo, foi desenvolvida para atender à crescente demanda da população mundial por produtos marinhos. No entanto, esse método é de baixa seletividade e resulta em uma captura significativa de fauna acompanhante, incluindo uma ampla variedade de organismos que são descartados como resíduos devido à falta de valor econômico. Pesquisadores identificaram uma vasta biodiversidade de espécies de fauna acompanhante, incluindo o gastrópode Olivancillaria urceus, que é um dos principais componentes da pesca de arrasto de fundo amplamente utilizada na pesca de camarão na costa sudeste brasileira. Esta espécie é endêmica do Atlântico Sudeste, encontrada em águas rasas da zona nerítica a profundidades de 0 a 53 metros. O potencial de biomoléculas dentro deste gênero de gastrópode descartados foi demonstrado em estudos anteriores utilizando extratos da espécie congênere, a Olivancillaria hiatula. No entanto, atualmente não há pesquisas sobre o potencial de moléculas bioativas em O. Urceus. Neste estudo, uma abordagem ômica, incluindo transcriptômica e proteopeptidômica, foi aplicada para explorar O. urceus em nível molecular. O transcriptoma do pé/manto muscular resultou na anotação de 19.097 genes pela ontologia genética, com a identificação de 20 transcritos semelhantes a toxinas, considerando a classe dos gastrópodes. A fração do proteoma confirmou 2.179 transcritos, incluindo sequências com atividade tóxica, como precursor de conotoxina, Conodipina-P3 e proteína contendo o domínio BPTI/Kunitz. Adicionalmente, 9.663 peptídeos de 1.484 proteínas precursoras foram detectados na fração peptídica, incluindo duas sequências semelhantes a neurotoxinas. A identificação dessas sequências pode levar à descoberta de novas moléculas com potencial terapêutico.

Palavras-chaves: bioprospecção, multiômica, biodiversidade, peptídeos

ABSTRACT

Aquatic ecosystems contain a wide variety of species, from microorganisms to large mammals, and are important for maintaining biodiversity as well as offering benefits to humankind. However, our knowledge of marine species is still limited, covering about 10% of the estimated total. The bioprospecting of molecules with therapeutic potential in marine organisms has been a growing field of research in recent decades, and the production of molecules in these marine organisms is related to their interaction with the environment, including symbiotic interactions with other organisms. Studies have shown a correlation between the innate characteristics of marine organisms, interactions, and the production of a variety of molecules refined through evolution, with peptides being one of the main types. Peptides are considered multifunctional molecules, as they can act systemically as signaling molecules, affect metabolic pathways, and exhibit various mechanisms of action. Fishing activity, especially bottom trawling, was developed to meet the growing global demand for marine products. However, this method is low in selectivity and results in a significant bycatch, including a wide range of organisms that are discarded as waste due to their lack of economic value. Researchers have identified a vast biodiversity of bycatch species, including the gastropod *Olivancillaria urceus*, which is one of the main components of bottom trawling widely used in shrimp fishing along the southeastern Brazilian coast. This species is endemic to the South Atlantic, found in shallow waters of the neritic zone at depths of 0 to 53 meters. The potential of biomolecules within this discarded gastropod genus has been demonstrated in previous studies using extracts from a congener species, *Olivancillaria hiatula*. However, there is currently no research on the potential of bioactive molecules in O. urceus. In this study, an omics approach, including transcriptomic and proteopeptidomic, was applied to explore O. urceus at the molecular level. The transcriptome of the muscular foot/mantle resulted in the annotation of 19,097 genes through gene ontology, with the identification of 20 toxin-like transcripts, considering the gastropod class. The proteome fraction confirmed 2,179 transcripts, including sequences with toxic activity, such as conotoxin precursor, Conodipine-P3, and proteins containing the BPTI/Kunitz domain. Additionally, 9,663 peptides from 1,484 precursor proteins were detected in the peptide fraction, including two neurotoxin-like sequences. The identification of these sequences could lead to the discovery of new molecules with therapeutic potential.

Keywords: bioprospecting, multi-omics, biodiversity, peptides

SUMÁRIO

1. INTRODUÇÃO

1.1 A BIODIVERSIDADE MARINHA E BIOPROSPECÇÃO DE PRODUTOS NATURAIS MARINHOS

Os ecossistemas aquáticos abrigam uma vasta biodiversidade de espécies que incluem desde a vida microbiana até grandes mamíferos aquáticos, sendo, portanto, uma fonte de benesses para a manutenção de toda a biosfera, inclusive para espécie humana [1-2]. Contudo, o entendimento que possuímos em relação aos espécimes presentes nos mares e oceanos é extremamente limitado. Dos trinta e cinco filos de animais estabelecidos, apenas um não possui representantes no ecossistema marinho e, quatorze são encontrados exclusivamente nos oceanos. Além disso, o número de espécies marinhas conhecidas é ainda relativamente baixo, cerca de 200 mil, o que representa 10% do total estimado, indicando uma necessidade de ampliação deste conhecimento [3-5].

Paralelamente a catalogação das espécies marinhas, houve nas últimas décadas um crescimento nas pesquisas de caráter exploratório, em especial no que se refere à bioprospecção de moléculas com potencial terapêutico em organismos que compõem esta vasta biodiversidade. Tanto espécies bem caracterizadas como aquelas passíveis de descrição tornaram-se fontes interessantes para a identificação dos chamados produtos marinhos naturais (PMNs) [6]. A produção destas moléculas, muitas das quais com características estruturais únicas, têm correlação com o papel desempenhado por um determinado organismo em seu habitat e sua interação com o mesmo, pois são necessários mecanismos adaptativos para a garantia de sua sobrevivência em um ambiente que pode ser considerado muitas vezes inóspito e de condições extremas, como variações na temperatura, luminosidade, salinidade, etc. Sendo assim a resposta de um organismo a estes estímulos externos, influenciam diretamente na estrutura química de metabólitos primários e secundários [7-8].

Por exemplo, a interação simbiótica entre microrganismos e macrorganismos marinhos são importantes para a manutenção de certas espécies marinhas, em grande parte, porque são através destas trocas que há metabolização de nutrientes necessários para o desenvolvimento, proteção química e mecânica contra patógenos, e, em contrapartida, são essenciais no descobrimento de novos PNMs [9]. Um estudo conduzido por Bhushan e colaboradores (2017), demonstra esta correlação existente entre as características inatas das esponjas-do-mar, como a ausência de mobilidade e proteção mecânica, somada a função de reciclagem do carbono e nitrogênio marinho, e a interação com outros organismos (macro e micro), que levam a produção de uma vasta variedade de moléculas aprimoradas ao longo dos milhões de anos de evolução.

Altmann (2017), aponta um crescimento exponencial de caracterização de PMNs, em uma taxa de 1000 novos compostos por ano. Esta taxa está atrelada aos seguintes pontos: (1) descoberta de novas espécies; (2) novas tecnologias para obtenção de amostras; (3) ganho de eficiência nos métodos de sequenciamento genético e avanços nas metodologias de espectrometria.

Dentre as classes conhecidas de PMNs, Dahiya e colaboradores (2021) e Nweze e colaboradores (2021) comentam sobre os aspectos biológicos e moleculares de peptídeos de origem marinha. Estes peptídeos têm sido utilizados como fármacos adjuvantes em terapias contra neoplasias, antimicrobianos de largo espectro, antivirais, antiparasitários e inibidores da geração de leucotrienos e prostaglandinas. Além disso, a utilização não se limita somente ao campo biomédico, podendo ser utilizado em outras áreas, como por exemplo na agricultura, auxiliando no controle de pragas das lavouras [14].

Haque e colaboradores (2022), listam uma diversidade de peptídeos que estão disponíveis no mercado e em utilização clínica, segregando-os em relação ao organismo fonte da molécula, mecanismo de ação e tamanho da molécula. Há também uma variabilidade importante, no que tange o mecanismo de ação e organismo de origem, pois, há inibidores de síntese de DNA, alquilante de DNA, inibidores de microtúbulos e, até mesmo como moléculas adjuvantes na produção de outras drogas, estes peptídeos provém de organismos que variam entre espécies de peixes, moluscos, esponjas-do-mar, entre outros [16-17].

1.2 PEPTÍDEOS BIOLOGICAMENTE ATIVOS ISOLADOS DE ORGANISMOS MARINHOS

Peptídeos são considerados moléculas multifuncionais, dado as características de poderem agir sistemicamente como moléculas sinalizadoras, sobre rotas metabólicas e somarem diferentes mecanismos de ação [18-19]. Zhang e colaboradores (2021), demonstram a presença de dois peptídeos com potencial antineoplásico, o SCH-P9 (Leu-Pro-Gly-Pro) e SCH-P10 (Asp-Tyr-Val-Pro), em extratos de *Sinonovacula constricta*, espécie de molusco marinho, que exibiram efeito in vitro contra linhagens DU-145 e PC-3, induzindo apoptose e inibição de ciclo celular, ainda no mesmo estudo fora encontrado efeito similar para o composto denominado YALPAH, extraído de *Setipinna taty* (Anchova escamosa), que exibiu citotoxicidade contra a linhagem PC-3, linhagem originária de células neoplásicas prostáticas [21].

Outro grupo de peptídeos citotóxicos, que foram isolados do gastrópode marinho *Dolabella Auricularia*, denominados de Dolastatinas, em especial a Dolastatina 10, são caracterizados por serem peptídeos lineares com aminoácidos distintos, tendo como mecanismo principal descrito a inibição da polimerização dos microtúbulos [22]. Geralmente esses peptídeos com potencial neoplásico tem como características tamanho entre 2-20 aminoácidos com ação correlata a sequência e composição destes, exibindo ações citotóxicas, antioxidantes, antiproliferativas e antitubulina [23].

Há na prática clínica peptídeos oriundo de espécies marinhas, onde Choudhary e colaboradores (2017) relata sobre a existência de 7 PMNs que foram aprovados pela agência americana FDA (*Food and Drug Administration*). Sendo que o primeiro peptídeo marinho aprovado foi um antineoplásico oriundo de uma esponja-do-mar caribenha (*Cryptotheca crypta*), que está disponível desde o ano de 1969. Somado a esta lista temos outros macrolídeos antineoplásicos provenientes de esponjas-do-mar e analgésicos derivados de peptídeos extraído de gastrópodes, que, de acordo com Duggan e colaboradores (2015), provém da conotoxina, que são peptídeos neurotóxicos, variando de 8 a 30 aminoácidos, possuindo em sua estrutura uma ou mais pontes dissulfídicas, isolados do gênero *Conus*, possuem alta eficiência na analgesia, sendo que naturalmente é utilizado como ferramenta predatória pelo caracol, porém, atualmente, é considerado, devido ao caráter específico de bloqueador de canais iônicos, uma molécula importante em estudos de neurofisiologia e neurofarmacologia.

No campo microbiológico, há diversos peptídeos que são cotados como boas prospecções no manejo de bactérias gram-positivas e negativas com baixa suscetibilidade a antibióticos, podendo ser citados as moléculas: Rodriguesines A e B, isolados de *Didemnum sp.*; Cyclo-(S-Pro-R-Leu), isolados de *Haliclona oculata*; Desmethylisaridin G, Desmethylisaridin C1, Isaridin e Isaridin G isolados de *B. felina* EN-135; Desotamide e Desotamide B isolados de *Streptomyces scopuliridis* SCSIO ZJ46 e Cyclo-(L-valyl-D-proline) e Cyclo-(L-phenylalanyl-D-proline) isolados de *Rheinheimera japonica*, obtendo bons resultados, inclusive contra *E. coli*, *S. aureus* e *S. pneumoniae* [24]. A demonstração do potencial de prospecção de peptídeos com atividade biológica importante é notória sendo, portanto, o ambiente marinho uma fonte inestimável destas moléculas [26].

1.3 ABORDAGENS MULTI-ÔMICAS PARA CONHECIMENTO DA BIODIVERSIDADE E DESCOBERTA DE MOLÉCULAS BIOATIVAS

Quando nos referimos a abordagem ômicas estamos fazendo alusão a aquisição de um amplo conhecimento nos diversos níveis dentro de um sistema biológico, entre eles os dados da sequência do DNA (genômica) [27], níveis de expressão de RNA (transcriptômica) [28], a abundância de proteínas e peptídeos (proteômica e peptidômica) [29-30], e também os níveis de micromoléculas envolvidas no metabolismo (metabolômica) [31].

Decorrente do desenvolvimento de abordagens multiômicas tem-se quebrado os paradigmas adotados para a análise de problemas-chave no campo biológico. Com a integração da genômica, transcriptômica, proteômica e metabolômica pode-se ampliar a capacidade de categorizar as interações moleculares que, utilizando-se de metodologias analíticas, especialmente a espectrometria de massas e o sequenciamento de nova geração, que permitem a quantificação e caracterização de novas moléculas e consequentemente novas espécies, preditas através de análises bioinformáticas com os extensos bancos de dados gerados durante estes processos. Assim, esse tipo de abordagem favorece o descobrimento de novas moléculas bioativas. A bioprospecção e caracterização de novas moléculas, incluindo peptídeos, são, portanto, uma das maneiras de agregar conhecimento e valor à diversidade marinha, através dessa abordagem torna-se possível a utilização de espécies consideradas subprodutos de práticas extrativistas [30,32-34]

O sequenciamento de RNA (RNA-Seq) é uma das tecnologias mais utilizadas do sequenciamento de nova geração [35]. Esta técnica resulta em uma amostragem de todo o transcriptoma analisado em grande profundidade sob uma determinada condição experimental particular, incluindo todas as sequências expressas, o que é uma representação reduzida do genoma na sua parte codificante de proteínas (~1%). O RNA-Seq tem muitas outras implicações interessantes, incluindo para estudos de filogenia, descoberta de polimorfismos de nucleotídeo único (SNP), perfil de micro RNA, novo transcrito ou descoberta de variante de splice. Permite ainda a comparação de perfis de mRNAs em condições ambientais distintas, por exemplo, condições benignas versus ecologicamente estressantes [35-36]. Assim, O RNA-Seq é uma ferramenta poderosa em pelo menos três situações: 1) é capaz de capturar a expressão de praticamente todos os genes sob as condições experimentais específicas; 2) não requer qualquer informação genética prévia, o que é bem adequado para organismos não- modelo; 3) é relativamente barato e acessível para a maioria dos laboratórios quando levado em consideração a quantidade de dados gerados [37].

Somado ao potencial do RNA-Seq, a abordagem multiômica também contempla a utilização da proteômica [38]. Para organismos não-modelos é utilizada o método bottom-up que consiste na digestão completa das proteínas em peptídeos para que sejam analisadas por espectrometria de massas, com isso é possível a caracterização de grande parte do proteoma, sendo eficaz em estudos de organismos modelos [39-40], porém, existe a problemática, quando se aplica a organismos não-modelos, pelo fato de não haver dados suficientes de genoma e transcriptoma anotados em bases dados, para contornar isso, dado ao dogma central da biologia molecular, podemos utilizar os dados adquiridos no transcriptoma, através do RNA-seq, para gerar sequências de proteínas previstas para que possamos confirmar pela proteômica posteriormente, existindo limitações, sendo a principal a qualidade dos dados gerados por RNA-seq [41-42]. Outra abordagem é a proteômica *De Novo*, abordagem que se concentra na identificação de peptídeos sem uso de informações prévias, ou seja, somente através da análise dos espectros formados na espectrometria de massas frente a espectros conhecidos dos aminoácidos, para deduzir suas sequencias [43-44]. Ambas as abordagens podem ser utilizadas para organismos não-modelos, inclusive para detecção de modificações pós-traducionais [45].

1.4 A FAUNA ACOMPANHANTE DA PESCA DE ARRASTO DE FUNDO COMO RECURSO PARA EXPLORAÇÃO

A utilização do ecossistema marinho como fonte de recursos remonta a origem da sociedade humana, sendo a pesca uma das principais formas de extração. Este tipo de atividade constitui uma parcela significativa dos produtos gerados para uma população em constante crescimento, com uma produção estimada no ano de 2016 de 110,2 milhões de toneladas de produtos aquícolas [46]. Devido à necessidade de maior volume extrativista e a superação de entraves tecnológicos, diversas formas de pesca foram desenvolvidas, sendo uma das principais, a pesca de arrasto de fundo, que é aplicada em larga escala para a comercialização do camarão-sete-barbas (*Xiphopenaeus kroyeri*) e camarão-rosa (*Farfantepenaeus brasiliensis* e *Farfantepenaeus paulensis*). Porém, trata-se de um método de baixíssima seletividade, levando a um grande número de outras espécies que são capturadas não intencionalmente e conjuntamente ao camarão, que é chamada de fauna acompanhante (bycatch) [47-49].

A fauna acompanhante é constituída de organismos dos mais diversos táxons, incluindo, crustáceos, esponjas, peixes, corais e, principalmente, moluscos, como não sendo alvo da pesca, são considerados subprodutos, sendo rejeitados, em sua grande maioria já sem vida, devido à ausência de valor técnico-econômico agregado, responsável pelo descarte de aproximadamente 1/3 dos 27 milhões de toneladas de rejeitos da pesca anualmente, acarretando em desequilíbrio ecológico nas áreas do extrativismo, tanto artesanal quanto industrial [48-52].

Alguns grupos de pesquisa têm se dedicado à identificação de espécies que compõem a fauna acompanhante. Em um estudo de amostragem da pesca do camarão sete barbas, em uma região que compreende seis municípios de Santa Catarina, 216 espécies foram identificadas, representando 105 famílias, distribuídas entre: cnidários, moluscos, crustáceos, equinodermos e peixes [53]. Portanto, esses dados evidenciam uma vasta biodiversidade a ser explorada para isolamento de novos compostos bioativos, incluindo peptídeos, que podem levar ao desenvolvimento de novas drogas e agregar valor para este recurso descartado.

Para fomentar a utilização destes subprodutos da pesca de arrasto de fundo de camarão, se faz necessário a catalogação do potencial técnico-econômico destas espécies, em especial neste estudo o gastrópode *Olivancillaria urceus*, um dos principais constituintes da pesca de arrasto de fundo no litoral do sudeste brasileiro, molusco pertencente à classe *Gastropoda*, descrito em 1798 por Röding, P. F. [54-55].

Como demonstrativo do potencial de prospecção de biomoléculas destes subprodutos provenientes da pesca do fundo de arrasto, o gastrópode *Olivancillaria hiatula*, congênere à espécie alvo deste estudo, que demonstrado Gasu e colaboradores (2018), obteve, utilizando o peptídeos originados de um extrato desta espécie, uma concentração mínima inibitória (MIC) destes peptídeos, frente a bactéria gram-positivas e negativas na casa de 0,039 e 2,5 mg/mL, sendo corroborado pelo estudo subsequente, Gasu e colaboradores (2019) observando o potencial antimicrobiano do extrato peptídico desta espécie, sendo necessário 39 µg/mL para inibir em 50% a formação do biofilme da bactéria *Pseudomonas aeruginosa*. Em contrapartida, não há estudos, até o momento, de bioprospecção utilizando como objeto de estudo O. urceus.

1.4.1 *OLIVANCILLARIA* URCEUS (RODING, 1798)

A *O. urceus*, também conhecida como *O. brasiliensis*, é um molusco pertencente à família *Olividae* e é endêmico do Atlântico Sudeste, estendendo-se sua ocorrência desde o estado da Bahia, no Brasil, até o Golfo San Matias, na Argentina, habitando fundos de areia rasa em águas da zona nerítica, podendo ser encontrados em profundidades que variam de 0 a 53 metros. Em relação ao seu descritivo morfológico, apresentam pé e sifão de coloração rosa escuro, sua concha, possui formato subquadrangular, de coloração rosada suja com linhas axiais, espirais e uma faixa fasciolar marrom-ocre de padrão irregular podendo atingir até 63,5 mm de comprimento, sendo grossas, lisas e polidas, com espessura máxima de 1 mm [58-59].

Seu hábito alimentar é unicamente predatório, sendo majoritariamente o consumo de bivalves, e possui inter-relação simbiótica, devido ao habitat ao qual se encontram, com microrganismos do ambiente marinho, incluindo bactérias Gram-positivas (Firmicutes e Actinobacteria) e bactérias Gram-negativas (Proteobactérias) [54-55,59].

2. JUSTIFICATIVA

Dada a problemática exposta com relação a fauna acompanhante e a possibilidade de bioprospecção de moléculas no ambiente marinho, este projeto almeja agregar valor comercial de um organismo considerado descarte, bem como ampliar o conhecimento científico acerca deste gastrópode ainda pouco estudado. Neste contexto, o projeto se enquadra, segundo a classificação das Áreas de Tecnologias Prioritárias definidas pelo Ministério da Ciência, Tecnologia, Inovações e Comunicações (MCTIC), como um estudo na Área de Tecnologias Habilitadoras, pois tem como objetivo destacar novas perspectivas de caráter econômico e científico. Além disso, o desenvolvimento das metodologias e técnicas aplicadas neste estudo, como a espectrometria de massas e sequenciamento de nova geração (NGS) e a obtenção destes dados moleculares em larga escala e com alto grau de confiança, contribuirão tanto para a formação de recursos humanos especializados como para o enriquecimento de bases de dados que poderão futuramente serem utilizadas em novas pesquisas sobre características funcionais destas moléculas, além de promover uma abordagem não convencional ao estudo de novas espécies endêmicas dentro da biodiversidade marinha brasileira.

3. ARTIGO

Type of the Paper (Article)

Transcriptomic and proteopeptidomic approach for the identification of proteins and peptides with pharmacological potential in the marine gastropod *Olivancillaria urceus*

Gabriel de Barros[1,2], Letícia Gama[1,2], Felipe de Mello[1], Claudia Correa[1], Horácio Montenegro[3], Leandro de Castro[1,2,*]

[1] Department of Biological and Environmental Sciences, Bioscience Institute, Sao Paulo State University (UNESP), Sao Vicente 11330-900, SP, Brazil
[2] Biodiversity of Coastal Environments Postgraduate Program, Department of Biological and Environmental Sciences, Bioscience Institute, Sao Paulo State University (UNESP), Sao Vicente 11330-900, SP, Brazil
[3] NGS Soluções Genômicas, Piracicaba 13416-030, SP, Brazil
* Correspondence: leandro.mantovani@unesp.br

Abstract: The marine environment is a rich source of new biotechnologies and products. Bottom trawl fishing for shrimp species like *Xiphopenaeus kroyeri* and *Farfantepenaeus brasiliensis* leads to the unintentional capture of non-target species, known as bycatch, which includes a variety of marine life often discarded without economic value. An abundant bycatch species along Brazil's southeast coast is *Olivancillaria urceus*, a carnivorous gastropod that primarily feeds on bivalves. Despite its abundance, this species remains underexplored, especially for biotechnological applications. Gastropods such as Conus are known for their diverse and potent toxins, with vast potential for pharmacological discoveries. In this study, an omic approach, including transcriptomics and proteopeptidomic, were applied to explore *O. urceus* at the molecular level. The transcriptome from the muscular foot/mantle resulted in the annotation of 19,097 genes by gene ontology, with 20 toxin-like transcripts identified considering Gastropod class. The proteome fraction confirmed 2,179 transcripts, including sequences with toxin activity, like conotoxin precursor, Conodipine-P3 and BPTI/Kunitz domain-containing protein. Additionally, 9663 peptides from 1484 precursor proteins were detected in the peptide fraction, including two sequences like neurotoxins. The identification of these sequences may lead to the discovery of new molecules with therapeutic potential.

Keywords: Omics; marine gastropod, bioprospection

1. Introduction

The marine environment constitutes a vast reservoir of organisms and therefore an invaluable source for the development of new biotechnologies and products. The bottom trawl fishing of *Xiphopenaeus kroyeri*, *Farfantepenaeus brasiliensis*, and *Farfantepenaeus paulensis* shrimp species is an extractivism method with low selectivity that leads to the unintentional capture of non-target species, known as bycatch, including fish, turtles, crustaceans, mollusks, sponges, corals, among others [1,2]. These incidental catches are considered fishing byproducts, lacking economic value, and are often discarded lifeless [3–6].

Among the species commonly captured as bycatch in bottom trawl fishing along Brazil's southeast coast is *Olivancillaria urceus* (*O. urceus*), a mollusk from the family Olividae and class Gastropoda, first described by Röding in 1798. Endemic to the Southeast Atlantic, this species has been documented as an abundant bycatch [1], range extends from the Bahia state in Brazil to the San Matias Gulf in Argentina, inhabiting shallow sandy bottoms in neritic zone waters, found at depths ranging from 0 to 53 meters [7]. These organisms are carnivorous and primarily feed on bivalves [8], living in symbiosis with other microorganisms, such as Gram-positive bacteria, Firmicutes and Actinobacteria, and Gram-negative bacteria, Proteobacteria [9]. Regarding their morphology, they have a dark pink foot and siphon, and their shell has a subquadrangular shape, dirty pink coloration with axial and spiral lines, and an irregular brown, ochre fasciolar band, which can reach

Citation: To be added by editorial staff during production.

Academic Editor: Firstname Lastname

Received: date
Revised: date
Accepted: date
Published: date

up to 63.5 mm in length, being thick, smooth, and polished with a maximum thickness of 1 mm (Figure 1A and B; [7]. Currently, there are not many studies on the prospecting of molecules from this neglected resource and no deep molecular characterization has been conducted.

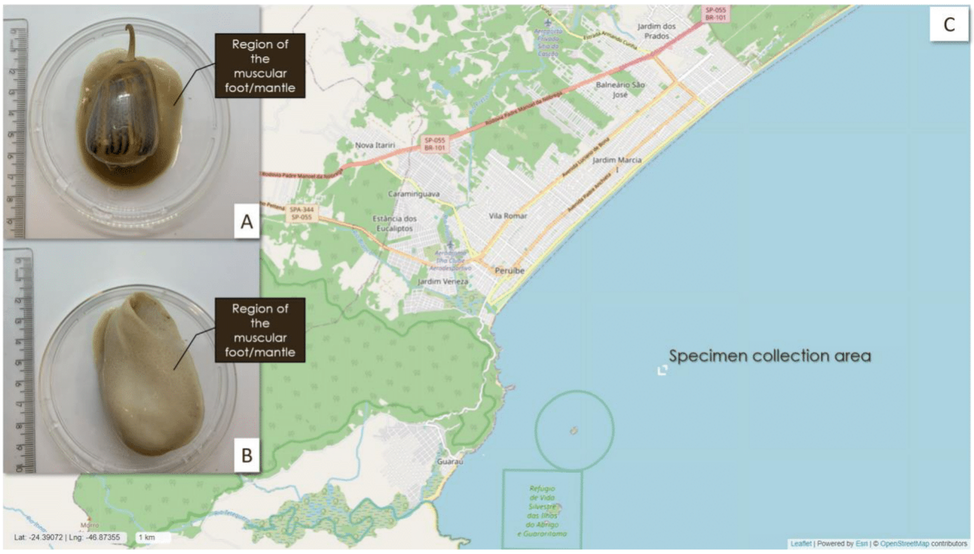


Figure 1. The images illustrate (A) the shell of the species and (B) the underside of the animal, highlighting the anatomical area from which the fragments were collected for transcriptomic and proteomic analyses. Map (C) shows the geographic region where the *O. urceus* samples were collected.

Recently, peptide extracts from whole body of congeneric species *O. hiatula* showed an antimicrobial activity against both gram-positive and gram-negative bacteria, with a minimum inhibitory concentration (MIC) ranged from 0.039 to 2.5 mg/mL [10]. Furthermore, another study at the same research group observed an antibiofilm activity with 50% inhibition against Pseudomonas aeruginosa, starting 39 µg/mL of peptide extract [11].

In the past twenty years, advancements in high-throughput sequencing techniques of nucleic acids (DNA/RNA) and proteins/peptides have allowed for a progress of knowledge in non-model species [12–15]. Most of these new studies have employed an integrated approach involving transcriptomics, utilizing RNA-seq to construct databases representing coding regions of the genome [16–18], along with proteomics and peptidomics data obtained through high-performance liquid chromatography coupled with mass spectrometry [19].

Using these molecular tools, natural marine products (NMPs) have been characterized as bioactive molecules, particularly for therapeutic use. Among the classes of NMPs are peptides, which exhibit a wide range of biological activities, including cytotoxic antineoplastic [20], antimicrobial [21], analgesic [22], antiviral [23], among others.

The study of NMPs in mollusks is exemplified by conotoxins, neurotoxic peptides isolated from marine gastropods of the Conus genus. Conotoxins are divided into several subfamilies based on their target, such as voltage-gated sodium channels (NaV), calcium channels (CaV), and nicotinic acetylcholine receptors (nAChRs) [24,25]. By selectively inhibiting these channels, conotoxins disrupt neurotransmission, leading to paralysis in prey, which is the primary function for Conus species in their natural predatory behavior. These peptides typically consist of 8 to 30 amino acids, and their structure is stabilized by multiple disulfide bridges, which impart a high degree of structural rigidity. This structural stability is key to their potent and precise interactions with ion channels, receptors, and transporters in the nervous system [26].

Beyond their utility in research, conotoxins hold immense potential for therapeutic applications. One of the most well-known conotoxins, Ziconotide (Prialt®), derived from Conus magus, has been approved for clinical use as a powerful non-opioid analgesic. Ziconotide works by blocking N-type voltage-gated calcium channels, effectively reducing chronic pain by inhibiting neurotransmitter release in pain pathway [27–29].

Considering the issue raised regarding bycatch fauna and the potential for molecule bioprospecting in the marine environment, this study aimed to add information to a discarded marine gastropod, through a set of integrated omics techniques.

2. Results

2.1 Transcriptome

Sequencing data were deposited at the National Center for Biotechnology Information (Bioproject: PRJNA1180544; Biosample: SAMN44522766). The Illumina sequencing generated a total of 239,009,224 paired reads (2x 119,504,612) of 100 bases each. Transcriptome assembly by Trinity resulted in 319,067 transcripts, and protein prediction by TransDecoder yielded 238,999 predicted peptides (108,636 complete proteins, 75,458 5' partial peptides, 28,826 3' partial peptides, and 26,079 internal peptides). We analyzed the BUSCO results based on the integrity of our transcriptome, comparing it with universally conserved orthologs. The analysis was carried out in two stages: one using Trinity to assemble the transcriptome and the other using TransDecoder to predict the proteins. We observed that, in the case of complete orthologs in a single copy, Trinity identified 524 (54.9%) of the 954 BUSCO orthologs, while TransDecoder found 540 (56.6%), which indicated good integrity of the assembly and of the predicted proteins. Next, we checked for duplicate complete orthologs: Trinity detected 363 (38.1%) and TransDecoder 347 (36.4%). We believe that this duplication could be related to gene duplication events or possible assembly artifacts. We went on to analyze the fragmented genes, which were relatively few: 43 (4.5%) in Trinity and 47 (4.9%) in TransDecoder. Finally, we noticed that only 24 (2.5%) of the genes were not found in Trinity, while 20 (2.1%) did not appear in TransDecoder. We therefore concluded that our transcriptome was well assembled and translated, capturing most of the universal orthologs, which allowed us to be confident in the high quality of the assembly and protein prediction.

Due to protein prediction by Trinotate, the total number of transcripts with any prediction exceeds the number of transcripts identified by Trinity, because a single transcript sequence can generate more than one possible protein isoform. Additionally, a high number of transcripts were observed to have no annotation. Out of the 378,795 identified transcripts, 301,180 did not match any sequences in the databases included in this analysis (approximately 75.51% of the total transcriptome; (see **Supplementary Table 1**). The NR_Diamondx_BLASTX database had the highest number of hits, covering just over 18% of the transcriptome **(Figure 2**).

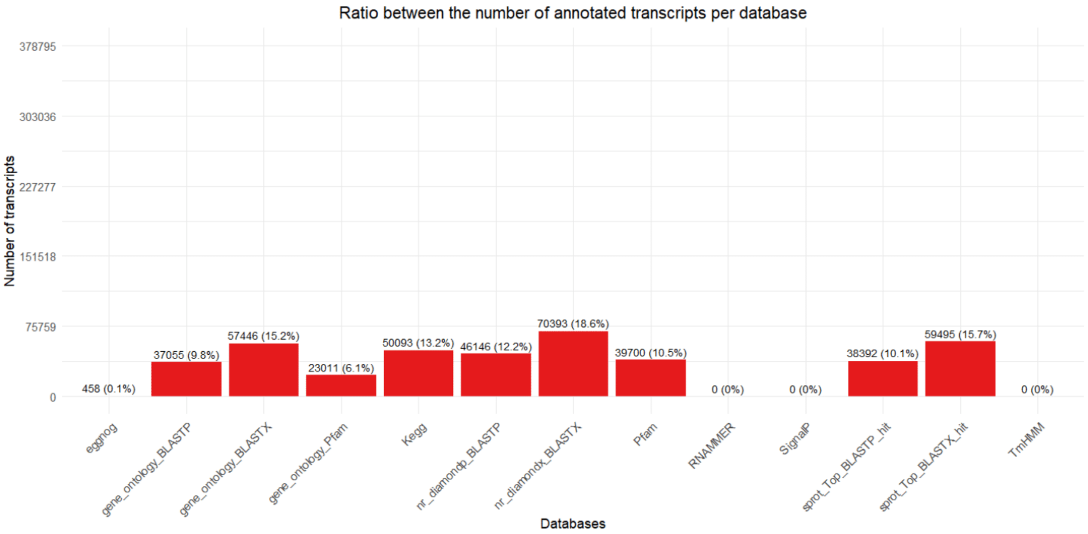


Figure 2. Percentage of transcripts from *O. urceus* identified and annotated per database. The databases included in the study were: eggNOG, gene_ontology_BLASTP, gene_ontology_BLASTX, Gene_Ontology_Pfam, KEGG, nr_diamond_BLASTP, nr_diamond_BLASTX, PFAM, RNAMMER, SignalP, sprot_Top_BLASTP_hit, sprot_Top_BLASTX_hit, and TrmHMM.centage of transcripts from O. urceus identified and annotated per database. The databases included in the study were eggNOG, gene_ontology_BLASTP, gene_ontology_BLASTX, Gene_Ontology_Pfam, KEGG, nr_diamond_BLASTP, nr_diamond_BLASTX, PFAM, RNAMMER, SignalP, sprot_Top_BLASTP_hit, sprot_Top_BLASTX_hit, and TrmHMM.

From these data, a filtering process was applied to obtain unique genes related to Gene Ontology (GO), resulting in 19,097 genes characterized by GO summarized by the WEGO. Among these, 16,357 were associated with cellular components, 16,182 with biological processes, and 16,869 with molecular functions, totaling 49,408 GO terms. Subcategorization in biological processes showed that most sequences are involved in cellular processes with 14532, metabolic process with 11061 and biological regulation with 9465 **(Figure 3A)**. Subcategorization of the annotation based on GO terms provided by WEGO demonstrated consistent distribution with respect to cellular components, with the majority originating from the terms "Cell" with 15112 genes and "Organelle" with 12405 genes **(Figure 3B)**. Finally, regarding molecular function, the highest concentration of sequences obtained the highest percentages associated with "binding" with 13717 genes and "catalytic activity" with 7908. Additionally, sequences associated with "toxin activity" were observed, guiding information of interest for bioprospecting with a total of 46 genes (**Figure 3C**). To assess the expression of these toxins across all transcripts, we conducted an analysis using transcripts per million (TPM). This allowed us to evaluate the presence of toxin-related transcripts within the Gastropoda class. The distribution of these transcripts is shown in **Table 1**, where the highest TPM values were observed for transcripts DN27199 c2 g1 i2 (371,809), DN16518 c0 g1 i1 (36,110), and DN3301 c0 g1 i1 i2 (10,069).

In addition, an analysis was conducted to assess the phylogenetic proximity of proteins predicted by TransDecoder, aiming to characterize genera closely related to the genus Olivancillaria. A primary filter was applied to the Gastropoda class, followed by a separation by genus **(Figure 4)**.

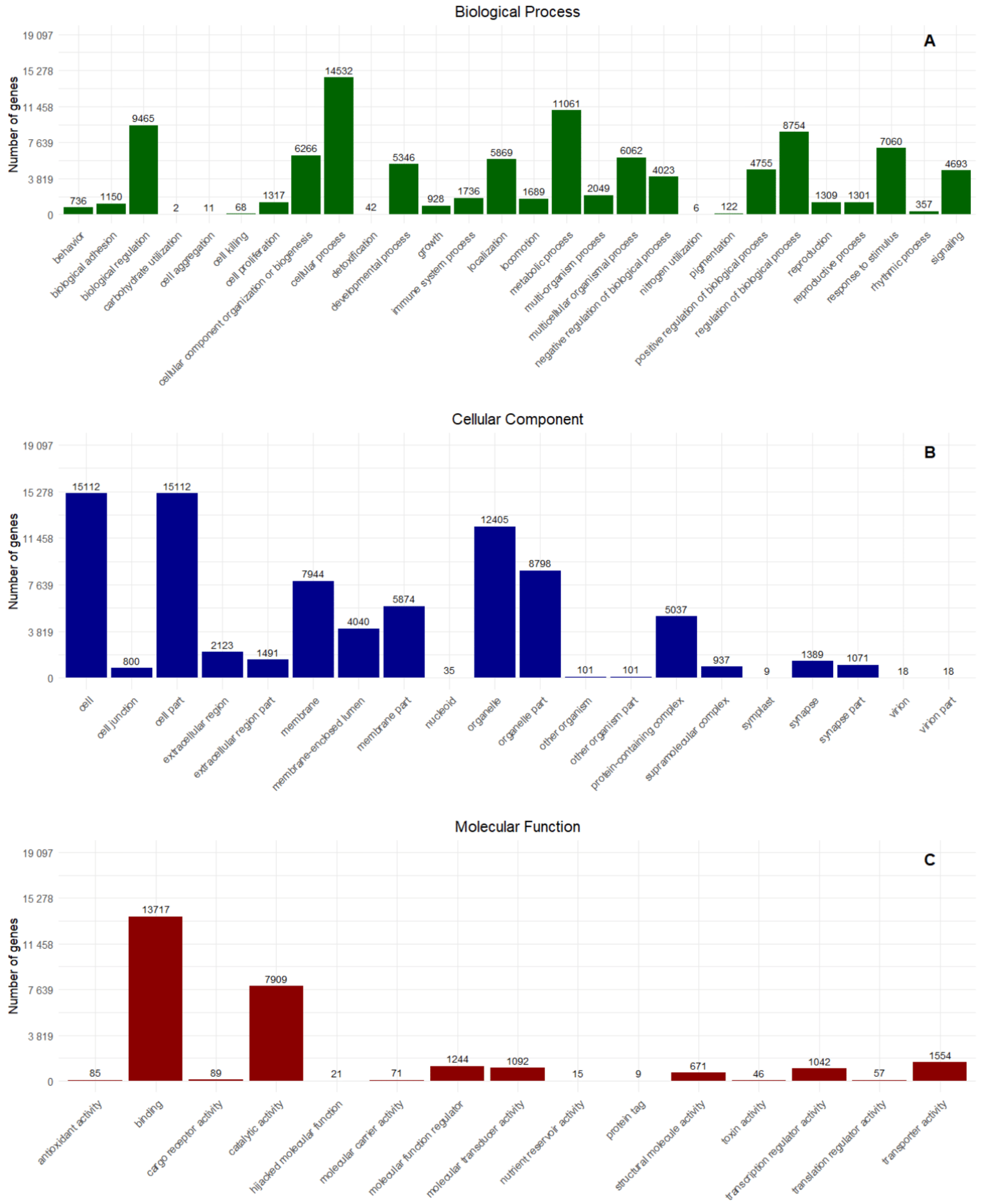


Figure 3. Analysis of the functional categories based on gene ontology of transcripts obtained from *O. urceus*. (A) Biological Processes, (B) Cellular Components and (C) Molecular Functions.

Table 1. Transcripts Identified in the *O. urceus* Transcriptome and Annotated Genes from UniProt/Swiss-Prot BLASTP Related to Toxin Activity by GO Term in Gastropoda Class

Transcript ID (Trinity ID)	**Uniprot/Swiss-Prot ID**	**Alignment Region in the Sequence**	**Similarity**	**P-Value**	**Description**	**TPM**
DN27199 c2 g1 i2	DAZ86947.1	Q:6-98,H:3-86	33.1%	1e-06	Conotoxin precursor Pmag02	371809
DN16518 c0 g1 i1	COP3_CONPU	Q:37-94,H:33-86	44.8%	1.3e-06	Conodipine-P3	36110
DN3301 c0 g1 i1	KCP_HALAI	Q:132-479,H:5-120	49.1%	5.53e-41	BPTI/Kunitz domain	10069
DN126505 c0 g1 i1	CUE3_CONIM	Q:9-44,H:38-73	47.2%	1.38e-06	Conotoxin Im14.3	29
DN13158 c3 g1 i1	COP1_CONPU	Q:27-93,H:28-90	44.8%	1.47e-12	Conodipine-P1	1679
DN132248 c0 g1 i1	CCAP_CONVL	Q:46-124,H:22-96	50.6%	1.15e-17	ConoCAP	28
DN133748 c0 g1 i1	CUE3_CONIM	Q:13-42,H:41-70	56.7%	5.69e-07	Conotoxin Im14.3	42
DN17279 c3 g1 i1	CUE3_CONIM	Q:21-50,H:44-73	56.7%	8.29e-07	Conotoxin Im14.3	27
DN109957 c0 g1 i1	CUE3_CONIM	Q:147-177,H:44-73	64.5%	6.39e-07	Conotoxin Im14.3	641
DN18169 c0 g1 i3	CELE_CONVC	Q:21-125,H:1-98	60%	1.33e-27	Elevenin-Vc1	100
DN18169 c0 g1 i2	CELE_CONVC	Q:21-125,H:1-98	60%	1.33e-27	Elevenin-Vc1	91
DN19184 c2 g1 i2	COP1_CONPU	Q:7-183,H:8-176	31%	7.61e-26	Conodipine-P1	68
DN19184 c2 g1 i1	COP1_CONPU	Q:7-183,H:8-176	31%	7.61e-26	Conodipine-P1	19
DN4345 c0 g2 i2	CCAP_CONVL	Q:14-48,H:155-189	71.4%	1.11e-12	ConoCAP	78
DN46921 c0 g1 i1	TU92_POLAB	Q:1-74,H:1-70	55.4%	2.78e-23	Turripeptide Pal9.2	25
DN55000 c0 g1 i1	ACTP1_TERAN	Q:43-230,H:1-190	70%	3.68e-98	Tereporin-Ca1	38
DN60625 c0 g1 i1	PV22_POMMA	Q:4-113,H:179-283	33.3%	6.28e-11	Perivitellin-2 31 kDa sub.	21
DN87107 c0 g1 i1	ACTP1_TERAN	Q:1-62,H:127-190	65.6%	248e-11	Tereporin-Ca1	34
DN94413 c0 g1 i1	CTHB5_CONVC	Q:55-149,H:1-96	80.2%	5.1e-51	Thyrostimulin beta-5 sub.	22

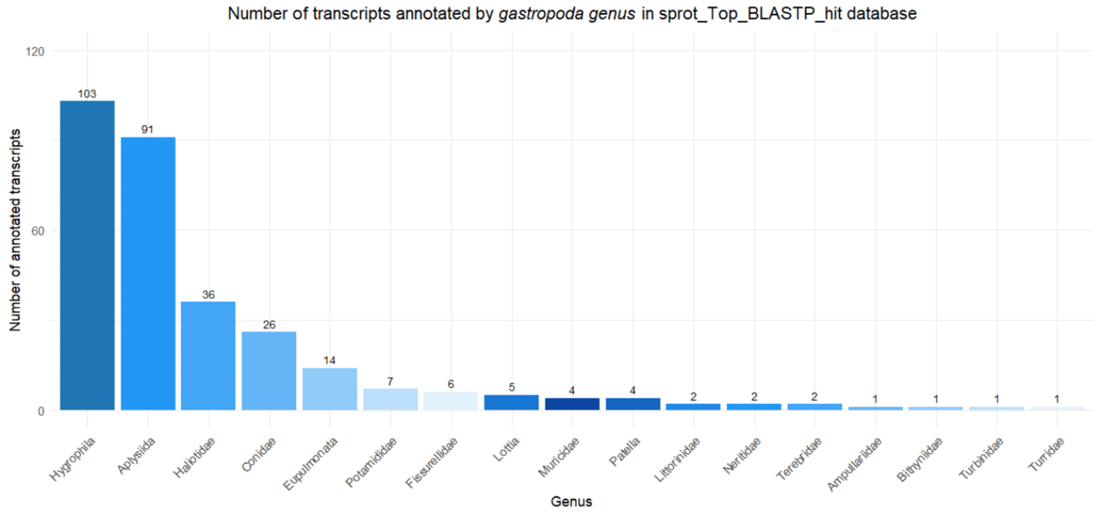


Figure 4. Number of annotated transcripts per Genus in the Sprot_BLASTP database.

2.2 Proteopeptidome

Proteomic analysis was performed on foot muscle/mantle protein extracts from *O. urceus*, with a single LC/MSMS run for each sample (n=4) of the fraction above 10 kDa. A total of 2179 proteins were identified, considering an FDR < 5%, using the database generated by Transdecoder based on protein prediction from the transcriptome **(Supplementary Table 2**). Out of the total proteins identified in the mass spectrometry, only 18.54% (404 proteins) did not have any known annotation and can be classified as unknown or uncharacterized. A total of 678 proteins detected by mass spectrometry were present in all groups. The demonstration of the overlapping relationships between different groups can be seen in **Figure 5**.

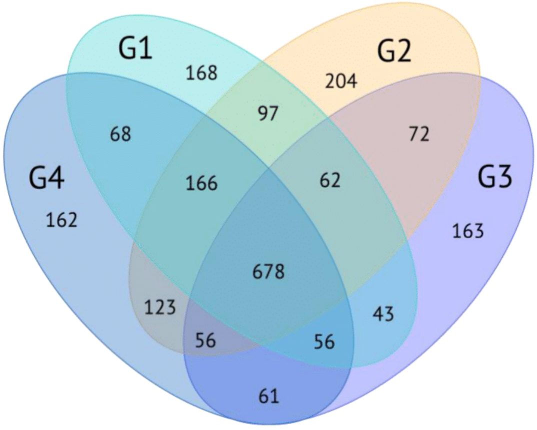


Figure 5. Number of identified proteins and overlap between sample groups in the proteomic analysis.

Regarding the databases, most proteins were identified through entries in the nr_diamond_BLASTX database and nr_diamond_BLASTP, accounting for approximately 89,1% of the total proteins (**Figure 6**).

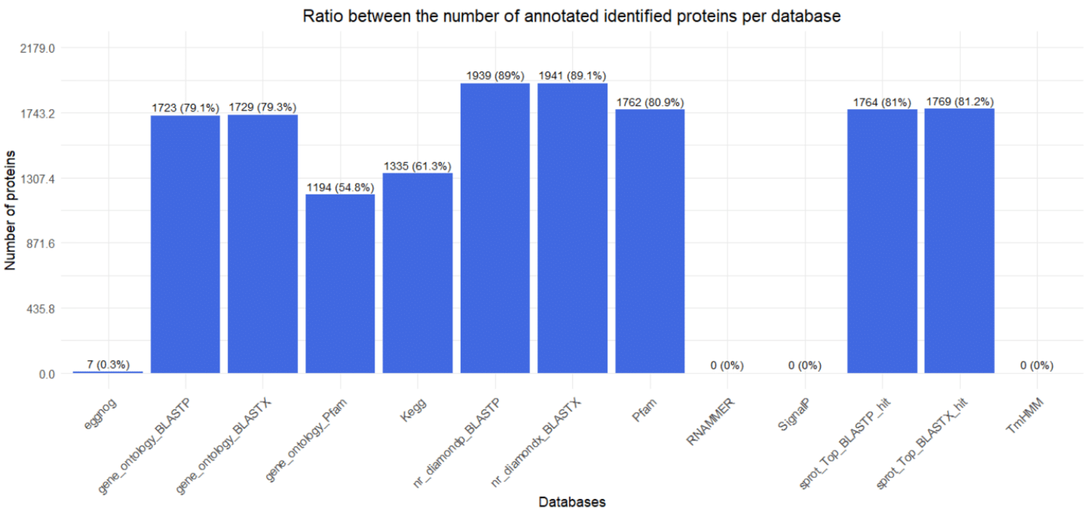


Figure 6. Percentage of proteins identified and annotated per database. The bar chart shows the number of proteins identified (Y-axis) in various databases (X-axis). The percentage within each bar represents the annotated proteins relative to the total number of proteins analyzed. The databases included in the study are eggNOG, gene_ontology_BLASTP, gene_ontology_BLASTX, Gene_Ontology_Pfam, KEGG, nr_diamond_BLASTP, nr_diamond_BLASTX, PFAM, RNAMMER, SignalP, sprot_Top_BLASTP_hit, sprot_Top_BLASTX_hit, and TrmHMM. This graph shows the coverage of each database in annotating proteins identified from the *O. urceus* transcriptome.

Using gene ontology (gene_ontology_BLASTP, gene_ontology_BLASTX, and gene_ontology_Pfam databases), it was able to classify proteins into three main categories: the cellular component, the biological process and the molecular function. For cellular components, were obtained 1601 gene ontology terms; for biological processes, were found 1497 terms; and for molecular functions, were identified 1659 terms. A total of 1067 proteins did not have any associated gene ontology terms (**Supplementary Figure 1**). Next, the five subcategories with the highest number of gene ontology occurrences for each previously established category are shown in **Figure 7**. Finally, proteomic analyses enabled the detection of three sequences related to toxin activity: DN16518 c0 g1 i1, DN27199 c2 g1 i2, and DN3301 c0 g1 i1 These toxins present in the proteome showed in the transcriptome a high expression in terms of transcripts per million compared to other toxins (**Table 1; Figure 8**).

To Peptidomic analysis, a LC-MS/MS run was conducted for each of the four samples of the fraction below 10 kDa, representing the naturally generated. A total of 9663 peptides from 1484 precursor proteins were found, considering an FDR < 5% and using the database generated by Transdecoder from the transcriptome. Of the total number of precursor proteins obtained by mass spectrometry, only 14.65% (314 precursor proteins) had no annotation and could be classified as unknown or uncharacterized. Regarding the databases, most of the precursor proteins were identified through entries in the nr_DIAMOND_BLASTX database, corresponding to a total of approximately 88% of the total proteins.

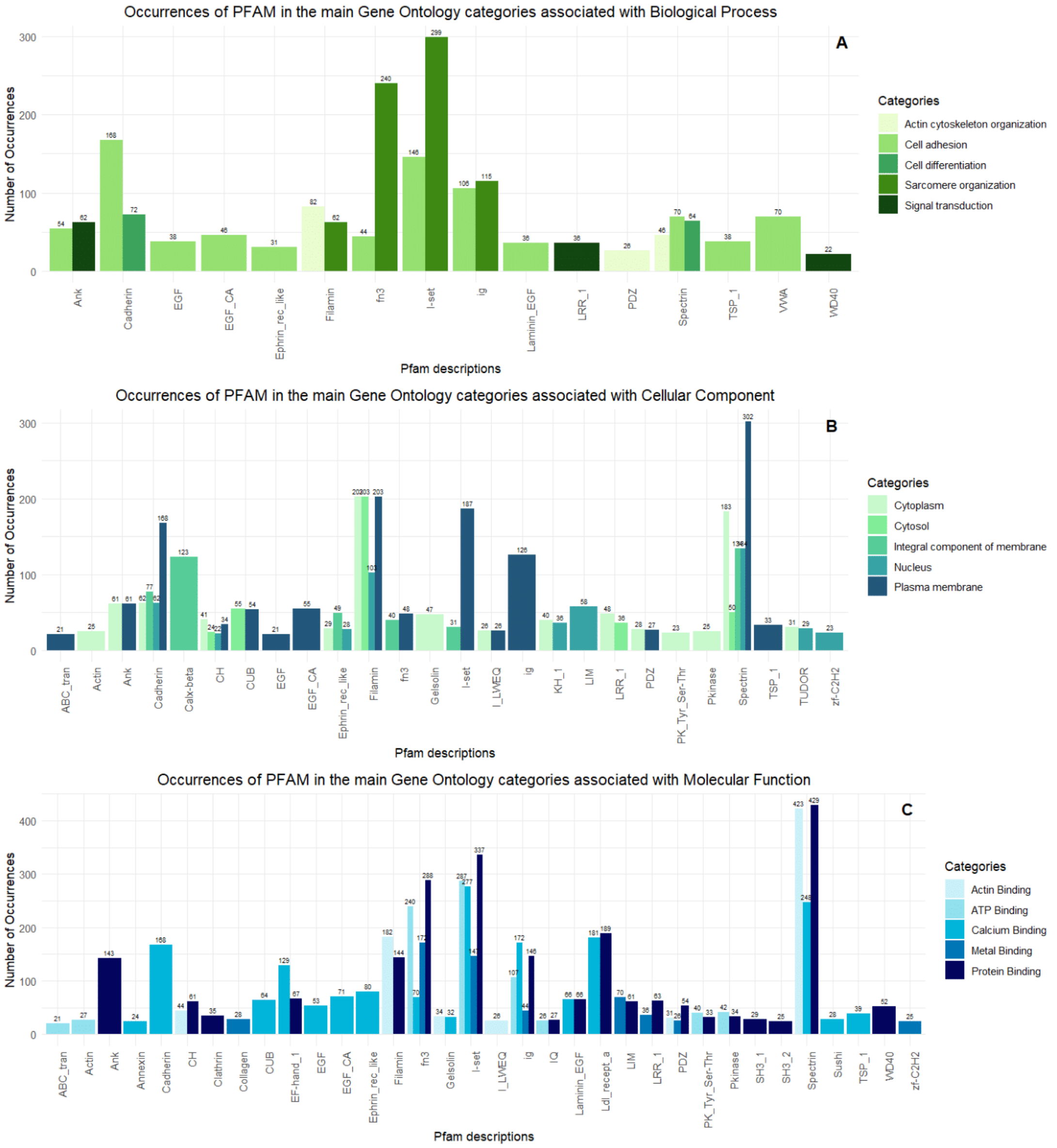


Figure 7. Occurrences of PFAM domains in the main Gene Ontology categories: (A) Biological Process, (B) Cellular Components, and (C) Molecular Functions observed in the proteome. In (A), the graph shows the number of occurrences in key biological processes, such as actin cytoskeleton organization, cell adhesion, cell differentiation, and signal transduction. In (B), the domains are presented in categories related to cellular components, including the cytoplasm, cytosol, integral membrane components, nucleus, and plasma membrane. Finally, (C) highlights domains involved in molecular functions, such as actin binding, ATP binding, calcium ion binding, metal ion binding, and protein binding.

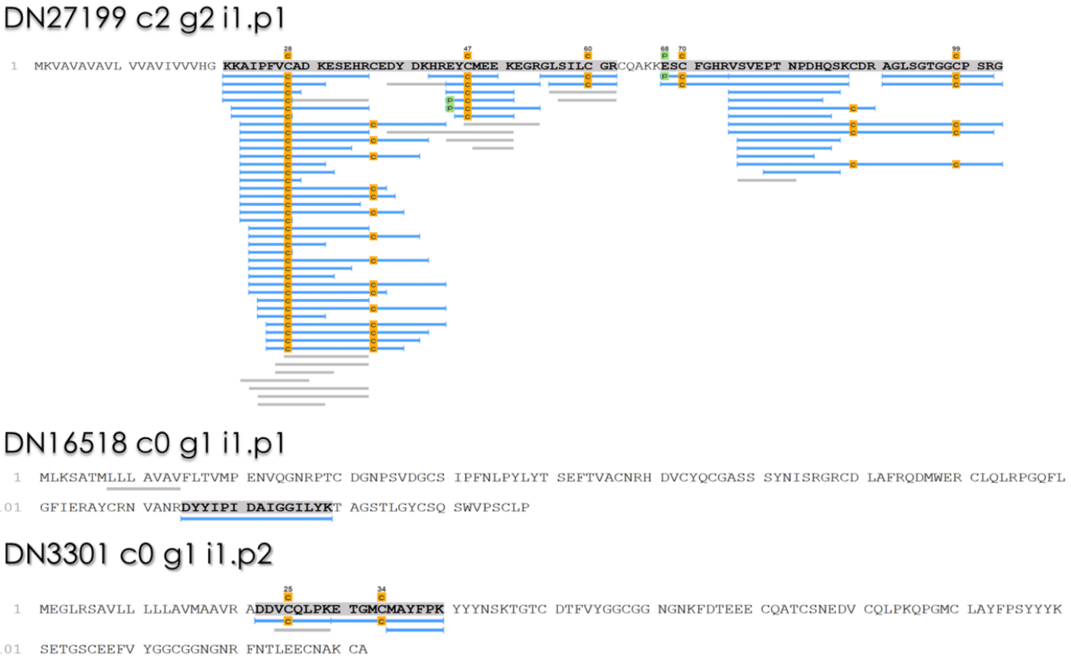


Figure 8. Peptide fragments identified as toxin-like proteins in the proteomic analysis of *O. urceus*. The top sequence, DN27199 c2 g2 i1.p1, shows similarity to DAZ86947.1, identified as a "Conotoxin precursor Pmag02." Highlighted peptide coverage includes *KATLF-PEAVD…AGLGTGYEC*, with post-translational modifications (carbamidomethylation at multiple cysteine residues). The middle sequence, DN16518 c0 g1 i1.p1, resembles COP3_CONPU, classified as "Conodipine-P3." Highlighted coverage includes the *DYIIPIDAIGGILYK* region. The bottom sequence, DN3301 c0 g1 i1.p2, corresponds to KCP_HALAI, from the "BPTI/Kunitz domain" family. Highlighted peptide coverage includes *ADDVCQLPKETGMCMAYFPK*, with post-translational modifications (carbamidomethylation at positions C25 and C34).

At similar way to the proteomic analysis, we used in the peptidome the gene ontology (gene_ontology_BLASTP, gene_ontology_BLASTX and gene_ontology_Pfam databases), separating the peptides into three main categories: cellular component, biological process and molecular function, followed by subcategories for each of these factors. For cellular components, we obtained 1061 gene ontology terms, biological process 1015 terms were found and for molecular function 1089 terms. A total of 754 peptides did not have any type of gene ontology associated with them. **Supplementary figure 2** shows the distribution of the subcategories that obtained the highest number of occurrences according to the main categories, considering, for better graphic visualization, categories with occurrences greater than 30.

Deepening the analysis of subcategories, we selected the five subcategories with the highest number of occurrences of gene ontology for each category previously established, and using the Pfam database, we observed which families and/or protein domains these subcategories are associated with. To better visualize the data, we only considered categories with occurrences greater than 20 for the cellular component and molecular function categories, while for biological process, it was necessary to adjust to occurrences greater than 4, since the data is relatively scarcer than the previous categories (**Figure 9**

).

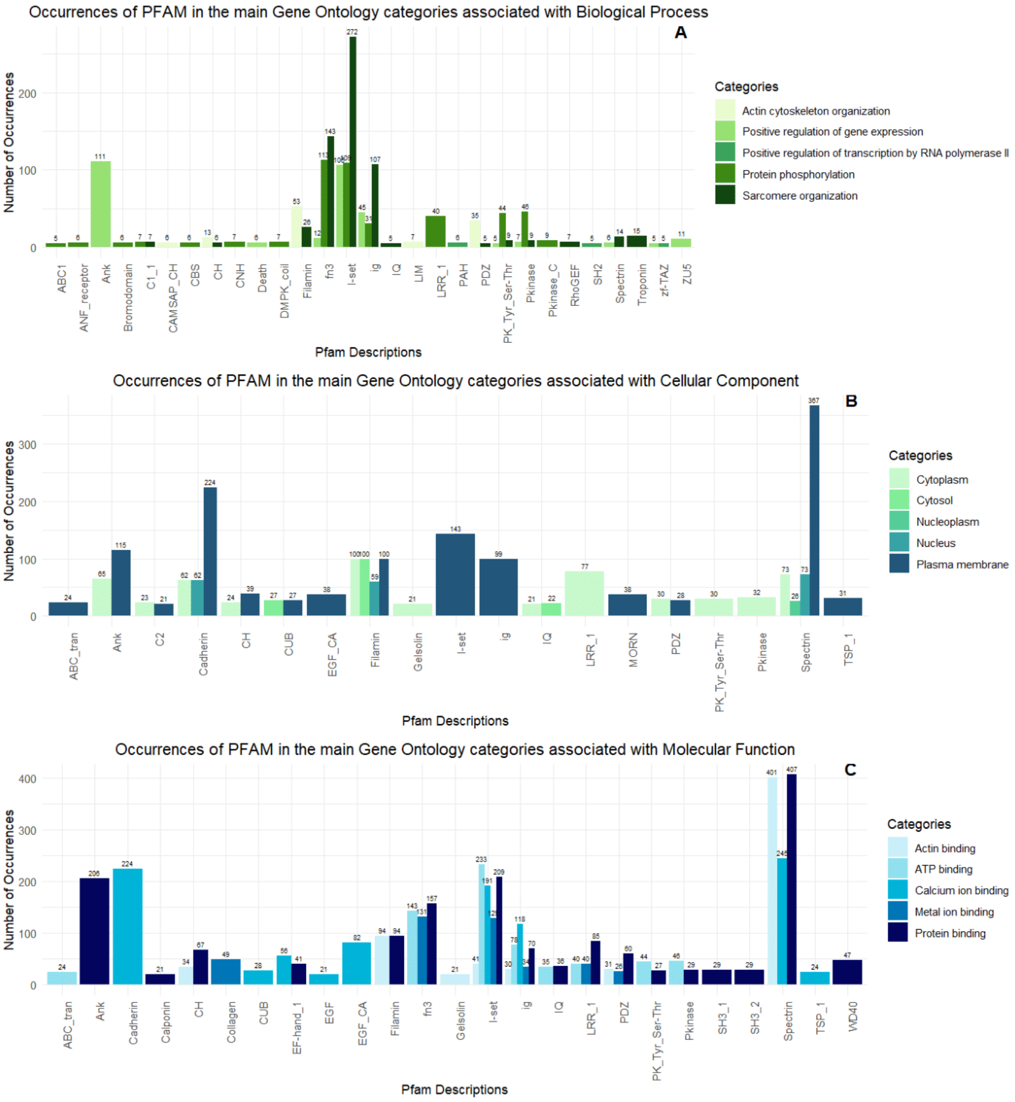
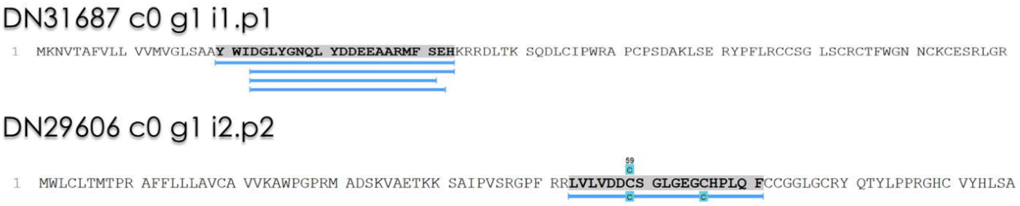


Figure 9. Occurrences of PFAM domains in the main Gene Ontology (GO) categories identified in the *O. urceus* peptidome. A. Biological Process, highlighting the predominance of proteins involved in actin cytoskeleton organization, regulation of transcription by RNA polymerase II, and protein phosphorylation. B. Cellular Component. C. Molecular Function, emphasizing proteins associated with actin, ATP, calcium ions, and metals.

In addition, we conducted a more comprehensive analysis, checking other databases in addition to SWISS-Prot/UniProt, as a means of checking for probable toxins present in the peptidomics. We found two transcripts with descriptions like toxins, DN31687 c0 g1 i1 and DN29606 c0 g1 i2 (**Figure 10**).



DN31687 c0 g1 i1.p1

1 MKNVTAFVLL VVMVGLSAA**Y WIDGLYGNQL YDDEEAARMF SEH**KRRDLTK SQDLCIPWRA PCPSDAKLSE RYPFLRCCSG LSCRCTFWGN NCKCESRLGR

DN29606 c0 g1 i2.p2

1 MWLCLTMTPR AFFLLLAVCA VVKAWPGPRM ADSKVAETKK SAIPVSRGPF RR**LVLVDDCS GLGEGCHPLQ F**CCGGLGCRY QTYLPPRGHC VYHLSA

Figure 10. Peptide sequences identified from the peptidomic analysis of *O. urceus*. The top sequence, DN31687 c0 g1 i1.p1, shares 60.4% identity with XP_025098946, an uncharacterized protein from *Pomacea canaliculata*. The highlighted region indicates the peptide sequence coverage: YWIDGLYGNQLYDDEEAARMFSEH. The bottom sequence, DN29606 c0 g1 i2.p2, corresponds to PF07740, part of the "Toxin_12" family. The peptide sequence coverage is highlighted as LVLVDDCSGLGGECCHPLQF, with detected post-translational modifications (carbamidomethylation at positions C57 and C66).

3. Discussion

This study is the first to describe a dataset of transcripts, proteins, and peptides for the marine gastropod species *O. urceus*, specifically focusing on the foot muscle/mantle region, using different omics. Regarding general transcriptome data, it was observed a high number of transcripts that did not have any type of annotation (approximately 75.51% of the total transcriptome), this characteristic of the data was expected, since this organism does not have any molecular database, as well as data of few nearby species and genus. This characteristic found in the analysis of the transcriptome reaffirms the absence of molecular data from non-model organisms, so there was a need to evaluate the data by phylogenetic proximity considering, mainly, the Gastropoda class.

The transcriptome, comprising 235,860 polypeptides predicted using the TransDecoder tool, provided a crucial dataset for analysis. The experimental validation of these predictions was supported by proteomic and peptidomic data obtained through mass spectrometry, which revealed 2,179 proteins —1,656 unique to the proteome and 961 unique to the peptidome. It is worth highlighting that 523 proteins were common across all three datasets (**Figure 11**), underscoring their robust expression and consistency across different analytical approaches.

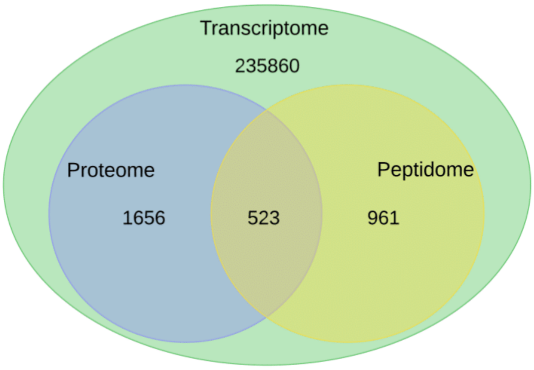


Figure 11. Integrated omics from foot muscle/mantle of *O. urceus*. Comparative analysis of transcripts confirmed by proteomic and peptidomic techniques.

Primarily we identified an abundance of proteins and enzymes mainly related to biological processes such as actin cytoskeleton organization, cell proliferation, cell adhesion, sarcomere components, and signal transduction. Among the classes of proteins found, we can mention: filamins, spectrins, ankyrins, cadherins, actins, myosins, tropomyosins and tubulins as well as domains related to calcium ion binding, such as calmodulin, troponin C, EF-Hand. In addition, the presence of a high amount of collagen was also detected. The high expression of these proteins is justified by the type of tissue analyzed, predominantly composed of the muscular foot and mantle regions, which consists mainly of muscle and connective tissue [30,31]. However, these tissues in gastropods are also covered by a layer of epidermis, with different types of epithelial cells, ranging from cuboidal to columnar shapes, which can be ciliated or mucus-secreting, and are related to functions such as locomotion, predation, and protection [32].

Additionally, our analyses also allowed the identification of proteins and peptides with pharmacological potential, such as domains related to toxin activity. The availability of molecular data on toxins from the genus Conus played an important role in the study of potential toxins present in the transcriptome of *O. urceus*. Of the transcripts related to toxin activity, 18 showed similarities with the genus Conus and are discussed below.

3.1 Transcripts and polipeptides related to toxin activity identified in O. urceus

Transcript DN16518 c0 g1 i1 shows high expression and sequence similarity to Conodipine-P1, while transcripts DN19184 c2 g1 i1 and DN19184 c2 g1 i2 exhibit similarity to Conodipine-P1. Additionally, transcript DN13158 c3 g1 i1 shows similarity to Conodipine-P3. These bioactive proteins, originally isolated from the venom of Conus purpurascens, are members of the phospholipase A2 family, specifically the Group IX subfamily. Structurally, both Conodipine-P1 and P3 are heterodimers, consisting of alpha and beta chains connected by disulfide bridges, which are crucial for maintaining protein stability and functionality. The disulfide bridges also confer resistance to proteolytic degradation, enhancing the functional integrity of the toxins in diverse environments [33].

The DN27199 c2 g2 i1 transcript correlates with a conotoxin precursor, sharing structural motifs typical of conotoxins, including a hydrophobic signal domain, a propeptide, and a cysteine-rich mature peptide. Post-translational modifications, such as cleavage into three domains and disulfide bridge formation, are critical for its maturation and biological activity [34,35].

The transcript DN3301 c0 g1 i1 exhibited a sequence like Kunitz-type domains found in various biological systems, such as animals, plants and fungi, primarily described as serine protease inhibitors and toxins in venomous animals [36–38]. Structurally, this type of domain consists of 50 to 70 amino acid residues, organized into two antiparallel β-sheets, stabilized by three disulfide bridges with a connection pattern of C1-C6, C2-C4, C3-C5 [39]. As observed in the amino acid sequence, the transcript detected in *O. urceus* contains two of these domains and the region F*Y*GC****N*F*****C, which is a signature sequence of the Kunitz family [40]. Additionally, a BLAST analysis showed 55.47%, 49.26%, and 47.79% identity with the TFPI-like multiple Kunitz type protease inhibitor sequence from the salivary gland of the non-conoidean neogastropod *Colubraria reticulata* (SPP68599.1; [41]) and with conkunitizins from *Conus magus* (DAC80551.1) and *Conus ermineus* (AXL95648.1), respectively [39,40].

The DN109957 c0 g1 i1, DN126505 c0 g1 i1, DN133748 c0 g1 i1, and DN17279 c3 g1 i1 transcripts are like Conotoxin Im14.3 from *Conus imperialis*, which is composed of 73 amino acids and stabilized by two disulfide bridges. The cysteine framework XIV (C-C-C-C) governs the arrangement of disulfide bonds, essential for maintaining the neurotoxin's three-dimensional conformation [42].

ConoCAP-related transcripts DN132248 c0 g1 i1 and DN4345 c0 g2 i2 encode a protein of 207 amino acids that is processed into three distinct chains: ConoCAP-a, ConoCAP-b, and ConoCAP-c and post-translational modifications, such as disulfide bridge formation and amidation, are important for the protein's functionality. ConoCAP-a and ConoCAP-b reduce heart rates in various organisms, they show differential potency and efficacy [43].

The DN18169 c0 g1 i3 and DN18169 c0 g1 i2 transcripts are associated with Elevenin-Vc1, a toxin consisting of 100 amino acids. Structurally, this toxin belongs to the Elevenin family, forming a monomer stabilized by disulfide bridges. Its C-C cysteine framework is essential for its neuropeptide-like function, affecting the nervous system of prey [44].

Tereporin-Ca1, identified from the DN55000 c0 g1 i1 and DN87107 c0 g1 i1 transcripts, is a 190-amino acid pore-forming protein. This toxin, classified within the actinoporin family, forms 1 nm-wide pores in lipid membranes, a structural feature associated with cytolytic and cardiac stimulatory functions [45].

Turripeptide Pal9.2, associated with the DN46921 c0 g1 i1 transcript, comprises 70 amino acids and contains three disulfide bridges. This neurotoxin targets ion channels and shares structural features with Kazal-type serine protease inhibitors, which confer additional protease resistance [46].

The DN60625 c0 g1 i1 transcript correlates with Perivitellin-2, a heterodimeric protein composed of a 31 kDa tachylectin subunit and a 67 kDa MACPF subunit. Structurally, the two subunits form a pore-forming complex with an internal diameter of approximately 5.6 nm, contributing to its cytotoxic effects in mammalian cells [47].

The DN31687 c0 g1 i1 transcript was identified in the peptidome fraction with a 60.4% identity similarity to the LOC112566798 (XP_025098946) protein from *Pomacea canaliculata* and obtained a TPM of 3706. This uncharacterized protein was predicted by automated computational analysis and is part of the genomic annotation of BioProject PRJNA472795 [48]. The 98 amino acid sequence shows a remarkable region between residues 73 and 92, with similarities to spider neurotoxins such as agatoxin, purotoxin and ctenitoxin, suggesting a possible neurotoxic function [48].

The DN29606 c0 g1 i2 transcript was identified from the Pfam database, associated with the PF07740 domain, with a TPM of 457, described as an ion channel inhibitory toxin. This family of potent toxins acts as ion channel inhibitors, blocking different types of ion channels [49,50].

Table 2. Summary of the putative toxins from *O. urceus* identified by omics, compared with toxins described in Conus and their respective actions.

Transcript ID – *O. urceus*	***Name of the toxin in Conus***	**Probable toxic activity**	**Author**
DN16518 c0 g1 i1	Conodipine-P3	Phospholipase A2	[33]
DN 13158 c3 g1 i1	Conodipine-P1		
DN27199 c2 g2 i1	Conotoxin precursor Pmag02	Conotoxin	[51,52]
DN3301 c0 g1 i1	Kunitz-type domain	Serine protease inhibitors/toxin	[39]
DN109957 c0 g1 i1	Conotoxin Im14.3	Likely as a neurotoxin	[25]
DN126505 c0 g1 i1			
DN133748 c0 g1 i1			
DN17279 c3 g1 i1			
DN132248 c0 g1 i1	ConoCAP	Decreases heart rate	[43]
DN4345 c0 g2 i2			
DN18169 c0 g1 i3	Elevenin-Vc1	Toxin induces hyperactivity	[44]

DN18169 c0 g1 i2			
DN12249 c1 g1 i1	Thyrostimulin alpha-2 subunit	Function of toxin by similarity	[53]
DN94413 c0 g1 i1	Thyrostimulin beta-5 subunit		
DN55000 c0 g1 i1	Tereporin-Ca1	Function of a pore-forming protein	[45]
DN87107 c0 g1 i1			
DN46921 c0 g1 i1	Turripeptide Pal9.2 toxin	Inhibiting ion channels by similarity with other similar toxins	[46]
DN60625 c0 g1 i1	Perivitellin-2 protein (31 kDa subunit)	Cytotoxicity	[47]

3.3 Limitations of omics analyses

Although we identified 20 toxin transcripts in the Gastropoda class from our transcriptome database, most of these transcripts represented the potential diversity of toxins expressed only at the mRNA level. When we correlated these transcripts with the data obtained by proteomic and peptidomic, we found some significant differences. Only the transcripts DN27199 c2 g2 i1.p1, DN16518 c0 g1 i1.p1, DN3301 c0 g1 i1.p2, DN31687 c0 g1 i1 and DN29606 were identified in the proteomics (**Figure 7**), which can be attributed to biological and technical factors.

Firstly, mRNA expression does not always translate directly into protein production. Post-transcriptional regulation, such as mRNA degradation, translation efficiency, and protein stability, play critical roles in determining the final proteins present in the cell. In addition, some proteins can be expressed at very low levels, making them difficult to detect [54,55].

In addition, the sensitivity and coverage of proteomics technologies can limit the detection of low-abundance or highly hydrophobic proteins, such as some toxins. Proteomic and peptidomic experiments confirmed the expression of many genes, particularly those with high transcript abundance per million. However, some less abundant transcripts were not detected and identified in these analyses, which can be in part explained by certain technical factors of the mass spectrometry method, such as the ionization capacity of specific peptide sequences, overlap of distinct peptides at the same elution time and priority for detecting ions in higher abundance [56–58]. Finally, the differential expression of toxins may be specific to certain physiological states, environmental conditions, stages of the life cycle, and sampling may not have captured these specific moments [59,60].

4. Materials and Methods

4.1 Collection and Maintenance of Specimens

The specimens of *O. urceus* were collected during the shrimp fishing carried out using the bottom trawling technique in Peruíbe County, on the south coast of São Paulo State (-24,346556, -46,956072; Figure 1C). The specimens were inserted on a recipient with seawater, sand, aerator and taken to the laboratory. The acclimatization period lasted 7 days with a photoperiod of 12 h, salinity of 33 ppm in artificial seawater, and 25 °C temperature. The specimens were fed two or three times a week with small fragments of squid.

4.2 RNA Extraction, mRNA Library Synthesis and Illumina Sequencing

For transcriptomic analysis, one specimen of *O. urceus* was placed on ice, and the muscular foot/mantle regions was removed using scissors. After the tissue collection, the sample was immediately stored in liquid nitrogen for shipping to NGS Soluções Genômicas, Piracicaba, Brazil. After adding Rnase-free Dnase I to eliminate genomic DNA, RNA extraction was performed following the protocol based on the Rneasy Lipid Tissue Kit. RNA integrity was evaluated with Bioanalyzer. The samples presenting an optic density ratio of 260/280 nm higher than 1,8 and RNA integrity higher than 7 were selected for the following experiments. The cDNA library was synthesized with ~1 ug of total RNA, following the recommendations of the RNA TrueSeq manufacturer. The resultant cDNA libraries were sequenced using an Illumina HiSeq 2500 platform. The Real-Time Analysis program (Illumina) performed the base calling of sequencing images, converting them into fastq sequences. Before assembly, raw reads were evaluated using FastQC [61].

4.3 Transcriptome Assembly, Annotation and Functional Enrichment

The resultant sequencing reads were submitted to de novo assembly using Trinity version 2.11.0, a pipeline for transcriptome assembly. Standard assembly parameters were used with a minimum length of 50 bp. Trinity filtered bases of low quality, removed Illumina adapters with Trimmomatic version 0.36 [62], counted k-mers present in sequencing reads with Jellyfish version 2.3.0 [63] and performed digital normalization to reduce data volume, without affecting the assembly's accuracy. Next, the modules Inchworm, Chrysalis, and Butterfly assemble the transcripts, reconstruct possible transcript isoforms, and group them into "genes". Trinity was executed with default options plus the argument "--trimmomatic --SS_lib_type RF". Annotation of the transcriptome was performed using the Trinotate pipeline version 3.2.2, which integrates various tools to predict proteins from transcriptome assembly to obtain functional annotation across multiple gene categories. Initially, TransDecoder version 5.5.0 [17] was used to extract the longest open reading frames (ORFs) from the transcripts and translate these ORFs in silico into their respective peptides. The peptides obtained in the previous step were used in similarity searches performed with DIAMOND version 2.0.11 [64] against the SwissProt/UniProt protein database [65], HMMER version 3.2.2 [66], and the Pfam database [67]. Peptides longer than 50 amino acids and peptides showing similarity to the databases were retained for further analysis. The retained transcripts were analyzed for the prediction of start and stop codons, and possible splicing sites, resulting in the final prediction of peptides. The transcripts assembled by the Trinity methodology and the peptides predicted by TransDecoder were compared in similarity searches performed with DIAMOND against the SwissProt and NCBI NR databases [68].

The assessment of the assembled transcriptome quality was obtained using the BUSCO program version 5.1.3 [69], which evaluates genome quality by the presence of an expected set of conserved single-copy genes in the assembly. Evaluation gene sets were derived from the OrthoDB database version 10.0 [70], representing a list of single-copy orthologous genes found in all representatives of certain taxonomic groups. The ortholog database used for analysis pertained to the Metazoa kingdom. The assembled transcriptome by Trinity (with the parameter "--mode trans") and the predicted proteome by TransDecoder (with the parameter "--mode prot") were also analyzed.

Functional annotations were performed by comparing the sequences with those in public databases. Gene Ontology (GO) terms were assigned to each unigene based on the GO terms annotated to their corresponding homologs. Unigenes were classified according to GO terms within molecular functions, cellular components, and biological process categories, and additionally plotted using the WEGO (Web Gene Ontology Annotation Plot) software with default parameters.

4.4 Proteomic and Peptidomic procedures

4.4.1 Protein and Peptide Fractions Preparation

For proteomics and peptidomics, we used a total of 12 specimens, divided into four groups based on size, with group 1 containing the largest specimens and group 4 the smallest.

Protein and peptide fractions of foot muscle/mantle from *O. urceus were* prepared as previously described with some modifications [13]. Tissues collected were resuspended in 10 mL of deionized water at 80 °C each, homogenized with a mechanical tissue homogeneizer (Thomas Scientific, Logan, NJ, USA) and kept at the same temperature for 20 min. After this incubation, the samples were cooled at 4 °C on ice for 40 min and then acidified with HCl to a final concentration of 10 mM. Cooled samples were sonicated 10 times with 20 pulses of 1s at 4 Hz and the homogenates centrifuged at 5000 × g for 40 min at 4 °C. After this point, supernatants were filtered through a Millipore membrane that allows the passage of molecules with a molecular weight of less than 10 Kda (Amicon Ultra, Millipore, Burlington, MA, USA). Eluates not retained by the membrane (peptide fraction) were loaded onto C18 pre-columns (OASIS—Waters, UK), washed with deionized water, and eluted within 100 % acetonitrile containing 0.15% trifluoroacetic acid (TFA). Sample volumes were concentrated to 5 μL in a vacuum centrifuge and stored at −80°C. Samples retained by the membrane (protein fraction) were processed for proteomic analysis. A solution of 8 M Urea was added to the incubation sample to a final concentration of 4 M, followed by the addition of dithiothreitol (DTT) to a final concentration of 5 mM. The mixture was incubated at 65 °C for 60 min. Iodoacetamide (IAA) was then added to a final concentration of 15 mM, and the samples were incubated for 60 min at room temperature in the dark. To quench the excess of IAA, DTT was added to a final concentration of 10 mM; proteins were digested with 1:50 trypsin:sample (Promega, Madison, WI, USA) overnight at 37 °C, acidified with formic acid, and desalted.

4.4.2 NanoLC and Mass Spectrometry

Peptide mixtures were suspended in 0.1% formic acid and analyzed as follows. An UltiMate 3000 Basic Automated System (Thermo Fisher®, Waltham, MA, USA) was set up and connected online with a Fusion Lumos Orbitrap mass spectrometer (Thermo Fisher®, Waltham, MA, USA) at the mass spectrometry facility RPT02H/Carlos Chagas Institute—Fiocruz, Paraná. Peptide mixtures was chromatographically separated on a column (15 cm in length with an internal diameter of 75 μm), packed in-house with ReproSil-Pur C18-AQ 3 μm resin (Dr. Maisch GmbH HPLC) with a flow rate of 250 nL/min of 5% to 38% ACN in 0.1% formic acid on a 120 min gradient. The Fusion Lumos Orbitrap was placed in data-dependent acquisition (DDA) mode to automatically turn between full-scan MS and MS/MS acquisition with 60 s dynamic exclusion. Survey scans (300–1500 m/z) were acquired in the Orbitrap system with a resolution of 120,000 at m/z 200. The most intense ions captured in a 2 s cycle time were chosen, excluding those which were unassigned or had a 1+ charge state. The selected ions were then isolated in sequence and fragmented using HCD (higher-energy collisional dissociation) with normalized collision energy of 30%. The fragment ions were analyzed with a resolution of 50,000 at 200 m/z. The general mass spectrometric conditions were as follows: 2.3 kV spray voltage, no sheath or auxiliary gas flow, heated capillary temperature of 175 °C, predictive automatic gain control (AGC) enabled, and an S-lens RF level of 30%. Mass spectrometer scan functions and nLC solvent gradients were regulated using the Xcalibur 4.1 data system (Thermo Fisher®, Waltham, MA, USA).

4.4.3 Protein and Peptide Identification

Raw data files (.raw) generated by the mass spectrometer were searched in the Transdecoder database built from the O. urceus, using the PEAKS Studio software (version 8.5; Bioinformatics Solution, Waterloo, ON, Canada; [71,72]. The research parameters used were as follows: no enzymatic specificity to peptidome fractions, trypsin to proteome fractions; precursor mass tolerance adjusted to ±15 ppm and fragmentation ion mass (tolerance ±0.5 Da); oxidized methionine (+15.99 Da) and acetylation (+42.01 Da) were defined as variable modifications. Carbamidomethylation (+57.02) was also added as a variable modification for alkylated and reduced samples. The identified peptides were then sorted by their mean local confidence to select the best spectra for annotation and filtered by FDR ≤5%.

4.4.4 Integrative data analyses of the transcriptome and proteome

To analyze and integrate the transcriptome and proteome data, the R programming language was used in the RStudio environment. In addition to the base package, the “ggplot2” version 3.5.1 [73], “dplyr” version 1.1.4 [74], https://github.com/tidyverse/dplyr, https://dplyr.tidyverse.org.), “tidyverse” version 2.0.0 [75], “readxl” version 1.4.3 [76] and “stringr” version 1.5.1 [77,78]. R package version 1.5.1 [79], https://github.com/tidyverse/stringr, https://stringr.tidyverse.org.) packages were used. Initially, the data was imported using the readxl package, which makes it easier to read Excel files, allowing large volumes of transcriptome and proteome data to be imported. Next, the dplyr package was used to efficiently manipulate this data, enabling operations such as data selection, filtering and aggregation. The stringr package was used to manipulate strings, which is essential for processing and standardizing transcript and protein identifiers. Data integration was carried out using tidyverse, a set of packages that offers tools for data science, allowing data to be organized and analyzed in a structured way. To visualize the results, we used the ggplot2 package, which allows graphs to be created and is used to visually represent the analyses, making it easier to interpret the data.

5. Conclusions

Supplementary Materials: The following supporting information can be downloaded at: /xxx/s1, Supplementary Table 1, Supplementary Table 2, Supplementary Figure 1: Supplementary Figure 2.

Author Contributions: Conceptualization, L.d.C. and G.d.B.; methodology, G.d. B.,L.G., F.d.M, C.C. and, L.d. C..; software, L.d.C., G.d.B and H.M.; validation, L.d.C., G.d.B and H.M; formal analysis, L.d.C., G.d.B and H.M.; investigation, L.d.C. and G.d.B; resources, L.d.C. and G.d.B; writing—original draft preparation, L.d.C. and G.d.B; writing—review and editing, G.d.B.,L.G., F.d.M, C.C., H.M. and, L.d. C.; visualization, L.d.C. and G.d.B; supervision, L.d.C.; project administration, L.d.C.; funding acquisition, L.d.C. All authors have read and agreed to the published version of the manuscript.”

Funding: This research was funded by the São Paulo Research Foundation (FAPESP), process number: 2019/16023-6 (Leandro Mantovani de Castro).

Data Availability Statement: The data are contained within the article, and mass spectrometry supporting information (raw files) can be downloaded at http://massive.ucsd.edu/ProteoSAFe/status.jsp?task=ad3339bfa37545b59c52484bd5517460

Acknowledgments: We acknowledge the São Paulo Research Foundation (FAPESP) for the funding and São Paulo State University, especially the LABPEPMAR group, and NGS Solutions for the support. The authors also thank FIOCRUZ for use of the Network of Technological Platforms.

Conflicts of Interest: The authors declare no conflicts of interest

References

1 Branco, J.O., Freitas Júnior, F. and Christoffersen, M.L. (2015) Bycatch Fauna of Seabob Shrimp Trawl Fisheries from Santa Catarina State, Southern Brazil. *Biota Neotropica*, **15**. https://doi.org/10.1590/1676-06032015014314.

2 De Castro Mendonça, L.M., Guimarães, C.R.P. and Lima, S.F.B. (2019) Mollusk Bycatch in Trawl Fisheries Targeting the Atlantic Seabob Shrimp Xiphopenaeus Kroyeri on the Coast of Sergipe, Northeastern Brazil. *Papeis Avulsos de Zoologia*, **59**. https://doi.org/10.11606/1807-0205/2019.59.33.

3 Herrera-Valdivia, E., López-Martínez, J., Vargasmachuca, S.C. and García-Juárez, A.R. (2016) Taxonomic and Functional Diversity of the Bycatch Fishes Community of Trawl Fishing from Northern Gulf of California, Mexico. *Revista de Biologia Tropical*, **64**. https://doi.org/10.15517/rbt.v63i3.15852.

4 Mendo, J., Mendo, T., Gil-Kodaka, P., Martina, J., Gómez, I., Delgado, R., Fernández, J., Travezaño, A., Arroyo, R., Loza, K. and James, M.A. (2022) Bycatch and Discards in the Artisanal Shrimp Trawl Fishery in Northern Peru. *PLoS ONE*, **17**. https://doi.org/10.1371/journal.pone.0268128.

5 Rodrigues-Filho, J.L., Couto, E. da C.G., Barbieri, E. and Branco, J.O. (2016) Seasonal Cycles of the Carcinofauna Caught as Bycatch in Sea-Bob Shrimp, Xiphopenaeus Kroyeri Fishery on Santa Catarina's Coast. *Boletim do Instituto de Pesca*, **42**. https://doi.org/10.20950/1678-2305.2016v42n3p648.

6 Barrilli, G.H.C., Filho, J.L.R., do Vale, J.G., Port, D., Verani, J.R. and Branco, J.O. (2021) Role of the Habitat Condition in Shaping of Epifaunal Macroinvertebrate Bycatch Associated with Small-Scale Shrimp Fisheries on the Southern Brazilian Coast. *Regional Studies in Marine Science*, **43**. https://doi.org/10.1016/j.rsma.2021.101695.

7 Teso, V. and Pastorino, G. (2011) A Revision of the Genus Olivancillaria (Mollusca: Olividae) from the Southwestern Atlantic. *Zootaxa*. https://doi.org/10.11646/zootaxa.2889.1.1.

8 Arrighetti, F., Teso, V., Brey, T. and Penchaszadeh, P.E. (2019) Gastropod Relevance in Predator-Prey Interactions on a Benthic Shallow Sandy Ecosystem at Mar Del Plata, Argentina (38°S). *Journal of the Marine Biological Association of the United Kingdom*, **99**. https://doi.org/10.1017/S0025315418000036.

9 Tangerina, M.M.P., Correa, H., Haltli, B., Vilegas, W. and Kerr, R.G. (2017) Bioprospecting from Cultivable Bacterial Communities of Marine Sediment and Invertebrates from the Underexplored Ubatuba Region of Brazil. *Archives of Microbiology*, **199**. https://doi.org/10.1007/s00203-016-1290-9.

10 Gasu, E.N., Ahor, H.S. and Borquaye, L.S. (2018) Peptide Extract from Olivancillaria Hiatula Exhibits Broad-Spectrum Antibacterial Activity. *BioMed Research International*, **2018**. https://doi.org/10.1155/2018/6010572.

11 Gasu, E.N., Ahor, H.S. and Borquaye, L.S. (2019) Peptide Mix from Olivancillaria Hiatula Interferes with Cell-to-Cell Communication in Pseudomonas Aeruginosa. *BioMed Research International*, **2019**. https://doi.org/10.1155/2019/5313918.

12 Beaulieu, L. (2019) Insights into the Regulation of Algal Proteins and Bioactive Peptides Using Proteomic and Transcriptomic Approaches. *Molecules*, **24**. https://doi.org/10.3390/molecules24091708.

13 Mazzi Esquinca, M.E., Correa, C.N., Marques de Barros, G., Montenegro, H. and Mantovani de Castro, L. (2023) Multiomic Approach for Bioprospection: Investigation of Toxins and Peptides of Brazilian Sea Anemone Bunodosoma Caissarum. *Marine Drugs*, **21**. https://doi.org/10.3390/md21030197.

14 Moutinho Cabral, I., Madeira, C., Grosso, A.R. and Costa, P.M. (2022) A Drug Discovery Approach Based on Comparative Transcriptomics between Two Toxin-Secreting Marine Annelids: Glycera Alba and Hediste Diversicolor. *Molecular Omics*, **18**. https://doi.org/10.1039/d2mo00138a.

15 Zheng, J.W., Lu, Y., Yang, Y.F., Huang, D., Li, D.W., Wang, X., Gao, Y., Yang, W.D., Guan, Y. and Li, H.Y. (2023) Systematic Dissection of Genomic Features Determining the Vast Diversity of Conotoxins. *BMC Genomics*, **24**. https://doi.org/10.1186/s12864-023-09689-4.

16 Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., Di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N. and Regev, A. (2011) Full-Length Transcriptome Assembly from RNA-Seq Data without a Reference Genome. *Nature Biotechnology*, **29**. https://doi.org/10.1038/nbt.1883.

17 Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., Macmanes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., Leduc, R.D., Friedman, N. and Regev, A. (2013) De Novo Transcript Sequence Reconstruction from RNA-Seq Using the Trinity Platform for Reference Generation and Analysis. *Nature Protocols*, **8**. https://doi.org/10.1038/nprot.2013.084.

18 Hrdlickova, R., Toloue, M. and Tian, B. (2017) RNA-Seq Methods for Transcriptome Analysis. Wiley Interdisciplinary Reviews: RNA. https://doi.org/10.1002/wrna.1364.

19 Slattery, M., Ankisetty, S., Corrales, J., Marsh-Hunkin, K.E., Gochfeld, D.J., Willett, K.L. and Rimoldi, J.M. (2012) Marine Proteomics: A Critical Assessment of an Emerging Technology. Journal of Natural Products. https://doi.org/10.1021/np300366a.

20 Ahmed, S., Khan, H., Fakhri, S., Aschner, M. and Cheang, W.S. (2022) Therapeutic Potential of Marine Peptides in Cervical and Ovarian Cancers. Molecular and Cellular Biochemistry. https://doi.org/10.1007/s11010-021-04306-y.

21 Ribeiro, R., Pinto, E., Fernandes, C. and Sousa, E. (2022) Marine Cyclic Peptides: Antimicrobial Activity and Synthetic Strategies. Marine Drugs. https://doi.org/10.3390/md20060397.

22 Lee, H.K., Zhang, L., Smith, M.D., Walewska, A., Vellore, N.A., Baron, R., McIntosh, J.M., White, H.S., Olivera, B.M. and Bulaj, G. (2015) A Marine Analgesic Peptide, Contulakin-G, and Neurotensin Are Distinct Agonists for Neurotensin Receptors: Uncovering Structural Determinants of Desensitization Properties. *Frontiers in Pharmacology*, **6**. https://doi.org/10.3389/fphar.2015.00011.

23 Sukmarini, L. (2022) Antiviral Peptides (AVPs) of Marine Origin as Propitious Therapeutic Drug Candidates for the Treatment of Human Viruses. Molecules. https://doi.org/10.3390/molecules27092619.

24 Jin, A.H., Vetter, I., Himaya, S.W.A., Alewood, P.F., Lewis, R.J. and Dutertre, S. (2015) Transcriptome and Proteome of Conus Planorbis Identify the Nicotinic Receptors as Primary Target for the Defensive Venom. *Proteomics*, **15**. https://doi.org/10.1002/pmic.201500220.

25 Jin, A.H., Muttenthaler, M., Dutertre, S., Himaya, S.W.A., Kaas, Q., Craik, D.J., Lewis, R.J. and Alewood, P.F. (2019) Conotoxins: Chemistry and Biology. Chemical Reviews. https://doi.org/10.1021/acs.chemrev.9b00207.

26 Duggan, P.J. and Tuck, K.L. (2015) Bioactive Mimetics of Conotoxins and Other Venom Peptides. Toxins. https://doi.org/10.3390/toxins7104175.

27 Deer, T.R., Pope, J.E., Hanes, M.C. and McDowell, G.C. (2019) Intrathecal Therapy for Chronic Pain: A Review of Morphine and Ziconotide as Firstline Options. Pain Medicine (United States). https://doi.org/10.1093/pm/pny132.

28 Matis, G., De Negri, P., Dupoiron, D., Likar, R., Zuidema, X. and Rasche, D. (2021) Intrathecal Pain Management with Ziconotide: Time for Consensus? Brain and Behavior. https://doi.org/10.1002/brb3.2055.

29 McGivern, J.G. (2007) Ziconotide: A Review of Its Pharmacology and Use in the Treatment of Pain. Neuropsychiatric Disease and Treatment. https://doi.org/10.2147/nedt.2007.3.1.69.

30 Bairati, A., Comazzi, M. and Gioria, M. (2001) An Ultrastructural Study of Connective Tissue in Mollusc Integument: II. Gastropoda. *Tissue and Cell*, **33**. https://doi.org/10.1054/tice.2001.0197.

31 Rigon, F., Mânica, G., Guma, F., Achaval, M. and Faccioni-Heuser, M.C. (2010) Ultrastructural Features of the Columellar Muscle and Contractile Protein Analyses in Different Muscle Groups of Megalobulimus Abbreviatus (Gastropoda, Pulmonata). *Tissue and Cell*, **42**. https://doi.org/10.1016/j.tice.2009.08.001.

32 Dennis, M.M., Molnár, K., Kriska, G. and Lőw, P. (2021) Mollusca. Invertebrate Histology, Wiley, 87–132. https://doi.org/10.1002/9781119507697.ch4.

33 Möller, C., Clay Davis, W., Clark, E., DeCaprio, A. and Marí, F. (2019) Conodipine-P1-3, the First Phospholipases A2 Characterized from Injected Cone Snail Venom. *Molecular and Cellular Proteomics*, **18**. https://doi.org/10.1074/mcp.RA118.000972.

34 Robinson, S.D. and Norton, R.S. (2014) Conotoxin Gene Superfamilies. Marine Drugs. https://doi.org/10.3390/md12126058.

35 Terlau, H. and Olivera, B.M. (2004) Conus Venoms: A Rich Source of Novel Ion Channel-Targeted Peptides. Physiological Reviews. https://doi.org/10.1152/physrev.00020.2003.

36 Bártová, V., Bárta, J. and Jarošová, M. (2019) Antifungal and Antimicrobial Proteins and Peptides of Potato (Solanum Tuberosum L.) Tubers and Their Applications. Applied Microbiology and Biotechnology. https://doi.org/10.1007/s00253-019-09887-9.

37 Bonturi, C.R., Teixeira, A.B.S., Rocha, V.M., Valente, P.F., Oliveira, J.R., Filho, C.M.B., Batista, I. de F.C. and Oliva, M.L.V. (2022) Plant Kunitz Inhibitors and Their Interaction with Proteases: Current and Potential Pharmacological Targets. International Journal of Molecular Sciences. https://doi.org/10.3390/ijms23094742.

38 Mourão, C.B.F. and Schwartz, E.F. (2013) Protease Inhibitors from Marine Venomous Animals and Their Counterparts in Terrestrial Venomous Animals. Marine Drugs. https://doi.org/10.3390/md11062069.

39 Mishra, M. (2020) Evolutionary Aspects of the Structural Convergence and Functional Diversification of Kunitz-Domain Inhibitors. Journal of Molecular Evolution. https://doi.org/10.1007/s00239-020-09959-9.

40 Ranasinghe, S. and McManus, D.P. (2013) Structure and Function of Invertebrate Kunitz Serine Protease Inhibitors. Developmental and Comparative Immunology. https://doi.org/10.1016/j.dci.2012.10.005.

41 Modica, M.V., Lombardo, F., Franchini, P. and Oliverio, M. (2015) The Venomous Cocktail of the Vampire Snail Colubraria Reticulata (Mollusca, Gastropoda). *BMC Genomics*, **16**. https://doi.org/10.1186/s12864-015-1648-4.

42 Jin, A.H., Dutertre, S., Dutt, M., Lavergne, V., Jones, A., Lewis, R.J. and Alewood, P.F. (2019) Transcriptomic-Proteomic Correlation in the Predation-Evoked Venom of the Cone Snail, Conus Imperialis. *Marine Drugs*, **17**. https://doi.org/10.3390/md17030177.

43 Möller, C., Melaun, C., Castillo, C., Díaz, M.E., Renzelman, C.M., Estrada, O., Kuch, U., Lokey, S. and Marí, F. (2010) Functional Hypervariability and Gene Diversity of Cardioactive Neuropeptides. *Journal of Biological Chemistry*, **285**. https://doi.org/10.1074/jbc.M110.171397.

44 Krishnarjuna, B., Sunanda, P., Seow, J., Tae, H.S., Robinson, S.D., Belgi, A., Robinson, A.J., Safavi-Hemami, H., Adams, D.J. and Norton, R.S. (2023) Characterisation of Elevenin-Vc1 from the Venom of Conus Victoriae: A Structural Analogue of α-Conotoxins. *Marine Drugs*, **21**. https://doi.org/10.3390/md21020081.

45 Gorson, J., Ramrattan, G., Verdes, A., Wright, E.M., Kantor, Y., Srinivasan, R.R., Musunuri, R., Packer, D., Albano, G., Qiu, W.G. and Holford, M. (2015) Molecular Diversity and Gene Evolution of the Venom Arsenal of Terebridae Predatory Marine Snails. *Genome Biology and Evolution*, **7**. https://doi.org/10.1093/gbe/evv104.

46 Olivera, B.M., Watkins, M., Bandyopadhyay, P., Imperial, J.S., de la Cotera, E.P.H., Aguilar, M.B., Vera, E.L., Concepcion, G.P. and Lluisma, A. (2012) Adaptive Radiation of Venomous Marine Snail Lineages and the Accelerated Evolution of Venom Peptide Genes. *Annals of the New York Academy of Sciences*, **1267**. https://doi.org/10.1111/j.1749-6632.2012.06603.x.

47 Giglio, M.L., Ituarte, S., Ibañez, A.E., Dreon, M.S., Prieto, E., Fernández, P.E. and Heras, H. (2020) Novel Role for Animal Innate Immune Molecules: Enterotoxic Activity of a Snail Egg MACPF-Toxin. *Frontiers in Immunology*, **11**. https://doi.org/10.3389/fimmu.2020.00428.

48 Zhou, X., Chen, Y., Zhu, S., Xu, H., Liu, Y. and Chen, L. (2016) The Complete Mitochondrial Genome of Pomacea Canaliculata (Gastropoda: Ampullariidae). *Mitochondrial DNA*, **27**. https://doi.org/10.3109/19401736.2014.919488.

49 Dutertre, S. and Lewis, R.J. (2010) Use of Venom Peptides to Probe Ion Channel Structure and Function. Journal of Biological Chemistry. https://doi.org/10.1074/jbc.R109.076596.

50 Liu, Z., Dai, J., Chen, Z., Hu, W., Xiao, Y. and Liang, S. (2003) Isolation and Characterization of Hainantoxin-IV, a Novel Antagonist of Tetrodotoxin-Sensitive Sodium Channels from the Chinese Bird Spider Selenocosmia Hainana. *Cellular and Molecular Life Sciences*, **60**. https://doi.org/10.1007/s00018-003-2354-x.

51 Robinson, S.D., Safavi-Hemami, H., McIntosh, L.D., Purcell, A.W., Norton, R.S. and Papenfuss, A.T. (2014) Diversity of Conotoxin Gene Superfamilies in the Venomous Snail, Conus Victoriae. *PLoS ONE*, **9**. https://doi.org/10.1371/journal.pone.0087648.

52 Pardos-Blas, J.R., Irisarri, I., Abalde, S., Tenorio, M.J. and Zardoya, R. (2019) Conotoxin Diversity in the Venom Gland Transcriptome of the Magician's Cone, Pionoconus Magus. *Marine Drugs*, **17**. https://doi.org/10.3390/md17100553.

53 Robinson, S.D., Li, Q., Bandyopadhyay, P.K., Gajewiak, J., Yandell, M., Papenfuss, A.T., Purcell, A.W., Norton, R.S. and Safavi-Hemami, H. (2017) Hormone-like Peptides in the Venoms of Marine Cone Snails. *General and Comparative Endocrinology*, **244**. https://doi.org/10.1016/j.ygcen.2015.07.012.

54 Tian, Y., Yu, A.M., Yin, C. and Qian, A. (2022) Editorial: Post-Transcriptional Regulation and Its Misregulation: From Molecular Basis to Translational Medicine. Frontiers in Cell and Developmental Biology. https://doi.org/10.3389/fcell.2022.1101576.

55 Wu, Q., Medina, S.G., Kushawah, G., Devore, M.L., Castellano, L.A., Hand, J.M., Wright, M. and Bazzini, A.A. (2019) Translation Affects MRNA Stability in a Codon-Dependent Manner in Human Cells. *eLife*, **8**. https://doi.org/10.7554/eLife.45396.

56 Hawkridge, A.M. (2014) Practical Considerations and Current Limitations in Quantitative Mass Spectrometry-Based Proteomics. Quantitative Proteomics. https://doi.org/10.1039/9781782626985-00001.

57 Li, C., Chu, S., Tan, S., Yin, X., Jiang, Y., Dai, X., Gong, X., Fang, X. and Tian, D. (2021) Towards Higher Sensitivity of Mass Spectrometry: A Perspective From the Mass Analyzers. Frontiers in Chemistry. https://doi.org/10.3389/fchem.2021.813359.

58 Venkataramanan, K.P., Min, L., Hou, S., Jones, S.W., Ralston, M.T., Lee, K.H. and Papoutsakis, E.T. (2015) Complex and Extensive Post-Transcriptional Regulation Revealed by Integrative Proteomic and Transcriptomic Analysis of Metabolite Stress Response in Clostridium Acetobutylicum. *Biotechnology for Biofuels*, **8**. https://doi.org/10.1186/s13068-015-0260-9.

59 Caburatan, L. and Park, J. (2021) Differential Expression, Tissue-Specific Distribution, and Posttranslational Controls of Phosphoenolpyruvate Carboxylase. Plants. https://doi.org/10.3390/plants10091887.

60 Campos, B., Fletcher, D., Piña, B., Tauler, R. and Barata, C. (2018) Differential Gene Transcription across the Life Cycle in Daphnia Magna Using a New All Genome Custom-Made Microarray. *BMC Genomics*, **19**. https://doi.org/10.1186/s12864-018-4725-7.

61 Andrews, S. (2010) FastQC. *Babraham Bioinformatics*.

62 Bolger, A.M., Lohse, M. and Usadel, B. (2014) Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics*, **30**. https://doi.org/10.1093/bioinformatics/btu170.

63 Marçais, G. and Kingsford, C. (2011) A Fast, Lock-Free Approach for Efficient Parallel Counting of Occurrences of k-Mers. *Bioinformatics*, **27**. https://doi.org/10.1093/bioinformatics/btr011.

64 Buchfink, B., Xie, C. and Huson, D.H. (2014) Fast and Sensitive Protein Alignment Using DIAMOND. Nature Methods. https://doi.org/10.1038/nmeth.3176.

65 The UniProt Consortium. (2019) UniProt: A Worldwide Hub of Protein Knowledge | Nucleic Acids Research | Oxford Academic. *Nucleic Acids Research*, **47**.

66 Wheeler, T.J. and Eddy, S.R. (2013) Nhmmer: DNA Homology Search with Profile HMMs. *Bioinformatics*, **29**. https://doi.org/10.1093/bioinformatics/btt403.

67 El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L.J., Gustavo A Salazar, Alfredo Smart, Erik L L Sonnhammer, Layla Hirsh, Lisanna Paladin, Damiano Piovesan, Silvio C E Tosatto and Robert D Finn. (2019) Pfam Protein Families Database in 2019 | Nucleic Acids Research | Oxford Academic. *Nucleic Acids Research*, **47**.

68 Sayers, E.W., Bolton, E.E., Brister, J.R., Canese, K., Chan, J., Comeau, D.C., Farrell, C.M., Feldgarden, M., Fine, A.M., Funk, K., Hatcher, E., Kannan, S., Kelly, C., Kim, S., Klimke, W., Landrum, M.J., Lathrop, S., Lu, Z., Madden, T.L., Malheiro, A., Marchler-Bauer, A., Murphy, T.D., Phan, L., Pujar, S., Rangwala, S.H., Schneider, V.A., Tse, T., Wang, J., Ye, J., Trawick, B.W., Pruitt, K.D. and Sherry, S.T. (2023) Database Resources of the National Center for Biotechnology Information in 2023. *Nucleic Acids Research*, **51**. https://doi.org/10.1093/nar/gkac1032.

69 Waterhouse, R.M., Seppey, M., Simao, F.A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E. V. and Zdobnov, E.M. (2018) BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. *Molecular Biology and Evolution*, **35**. https://doi.org/10.1093/molbev/msx319.

70 Kriventseva, E. V., Kuznetsov, D., Tegenfeldt, F., Manni, M., Dias, R., Simão, F.A. and Zdobnov, E.M. (2019) OrthoDB V10: Sampling the Diversity of Animal, Plant, Fungal, Protist, Bacterial and Viral Genomes for Evolutionary and Functional Annotations of Orthologs. *Nucleic Acids Research*, **47**. https://doi.org/10.1093/nar/gky1053.

71 Ma, B., Zhang, K., Hendrie, C., Liang, C., Li, M., Doherty-Kirby, A. and Lajoie, G. (2003) PEAKS: Powerful Software for Peptide de Novo Sequencing by Tandem Mass Spectrometry. *Rapid communications in mass spectrometry : RCM*, **17**, 2337–2342. https://doi.org/10.1002/rcm.1196.

72 Zhang, J., Xin, L., Shan, B., Chen, W., Xie, M., Yuen, D., Zhang, W., Zhang, Z., Lajoie, G.A. and Ma, B. (2012) PEAKS DB: De Novo Sequencing Assisted Database Search for Sensitive and Accurate Peptide Identification. *Molecular & Cellular Proteomics*, **11**, M111.010587-M111.010587. https://doi.org/10.1074/mcp.M111.010587.

73 Wickham, H. (2009) Ggplot2 by Hadley Wickham. Media.

74 Wickham, H., François, R., Henry, L. and Müller, K. (2021) A Grammar of Data Manipulation [R Package Dplyr Version 1.0.7]. *Media*.

75 Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K. and Yutani, H. (2019) Welcome to the Tidyverse. *Journal of Open Source Software*, **4**. https://doi.org/10.21105/joss.01686.

76 Wickham, H., Bryan, J., Kalicinski, M., Valery, K., Leitienne, C., Colbert, B., Hoerl, D. and Miller, E. (2019) Read Excel Files Package ‘Readxl.’ *Package ‘ readxl ’ Version 1.3.1*.

77 Wickham, H. (2019) Package Stringr: Simple, Consistent Wrappers for Common String Operations. Cran.

78 Wickham, H. (2022) CRAN - Package Stringr. 2019.

79 Wickham, H. and Grolemund, G. (2023) R for Data Science: Visualize, Model, Transform, Tidy, and Import Data. O’Reilly.

4. CONCLUSÃO

Em resumo, os resultados evidenciados através de uma abordagem *ômica* em *O. urceus* refletem um conjunto abrangente de adaptações moleculares que permitem sua sobrevivência em um ambiente marinho dinâmico e desafiador. Essas análises revelaram um perfil proteico dominado por várias famílias importantes essenciais aos processos celulares, como organização do citoesqueleto, sinalização celular e adesão, ressaltando os papéis críticos dessas funções na adaptação. Além disso, este estudo confirmou a detecção de transcritos e polipeptídeos relacionados à atividade da toxina, destacando a importância do uso de abordagens integrativas para estudar a expressão gênica em um organismo não modelo. Juntas, essas descobertas enfatizam não apenas as funções biológicas do *O. urceus*, mas também o potencial biotecnológico de suas proteínas para pesquisas e aplicações futuras.

REFERÊNCIAS

1. GLEICK Peter H, et al. Water in crisis: a guide to the world's fresh water resources. New York: Oxford University Press; 1993. World fresh water resources; p. 13-24.

2. Daniotti Sara, Re Ilaria, et al. Marine Biotechnology: Challenges and Development Market Trends for the Enhancement of Biotic Resources in Industrial Pharmaceutical and Food Applications. A Statistical Analysis of Scientific Literature and Business Models. Mar. Drugs. 2021;19(61). DOI: 10.3390/md19020061.

3. MORA Camilo, TITTENSOR Derek P., ADL Sina, et al. How Many Species Are There on Earth and in the Ocean?. Plos Biology. 2011;9(8):1-8. DOI: 10.1371/journal.pbio.1001127.

4. THOMPSON Cristiane C., KRUGER Ricardo H., THOMPSON, Fabiano L., et al. Unlocking Marine Biotechnology in the Developing World. Trends In Biotechnology. 2017;35(12):1119-1121. DOI: 10.1016/j.tibtech.2017.08.005.

5. LOTZE Heike K. Marine biodiversity conservation. Current Biology. 2021;31(19):1190-1195. DOI: 10.1016/j.cub.2021.06.084.

6. MOLINSKI Tadeusz F., et al. Drug development from marine natural products. Nature Reviews Drug Discovery. 2008;8(1):69-85. DOI: 10.1038/nrd2487.

7. BLUNT John W, et al. Marine natural products. Nat. Prod. Rep. 2013;30(2):237-323. DOI: 10.1039/c2np20112g.

8. KIURU Paula, et al. Exploring Marine Resources for Bioactive Compounds. Planta Medica. 2014;80(14):1234-1246. DOI: 10.1055/s-0034-1383001.

9. WAHL Martin, et al. The Second Skin: ecological role of epibiotic biofilms on marine organisms. Frontiers In Microbiology. 2012;3(1):1-21. DOI: 10.3389/fmicb.2012.00292.

10. BHUSHAN Agneya, et al. Entotheonella Bacteria as Source of Sponge-Derived Natural Products: opportunities for biotechnological production. Blue Biotechnology. 2017;55:291-314. DOI: 10.1007/978-3-319-51284-6_9.

11. ALTMANN Karl-Heinz, et al. Drugs from the Oceans: marine natural products as leads for drug discovery. Chimia. 2017;71(10):646-652. DOI: 10.2533/chimia.2017.646.

12. DAHIYA Rajiv, et al. Structural and biological aspects of natural bridged macrobicyclic peptides from marine resources. Archiv Der Pharmazie. 2021;1(1):1-16. DOI: 10.1002/ardp.202100034.

13. NWEZE Justus Amuche, et al. Potentials of marine natural products against malaria, leishmaniasis, and trypanosomiasis parasites: a review of recent articles. Infectious Diseases Of Poverty. 2021;10(1):1-19. DOI: 10.1186/s40249-021-00796-6.

14. SONG Chenggang, et al. Marine Natural Products: the important resource of biological insecticide. Chemistry & Biodiversity. 2021;18(5):1-18. DOI: 10.1002/cbdv.202001020.

15. HAQUE Neshatul, et al. Marine Natural Products in Clinical Use. Marine Drugs. 2022;20(8):1-40. DOI: 10.3390/md20080528.

16. ZHANG Bin, et al. Marine Sponge-Associated Fungi as Potential Novel Bioactive Natural Product Sources for Drug Discovery: a review. Mini-Reviews In Medicinal Chemistry. 2020;20(19):1966-2010. DOI: 10.2174/1389557520666200826123248.

17. LIU Miaomiao, et al. Potential of marine natural products against drug-resistant bacterial infections. The Lancet Infectious Diseases. 2019;19(7):237-245. DOI: 10.1016/S1473-3099(18)30711-4.

18. UDENIGWE Chibuike C, et al. Food Protein-Derived Bioactive Peptides: production, processing, and potential health benefits. Journal Of Food Science. 2011;77(1):11-24. DOI: 10.1111/j.1750-3841.2011.02455.x.

19. BALI Anjana, et al. Interplay between RAS and opioids: opening the pandora of complexities. Neuropeptides. 2014;48(4):249-256. DOI: 10.1016/j.npep.2014.05.002.

20. ZHANG Qi-Ting, et al. Recent Advances in Small Peptides of Marine Origin in Cancer Therapy. Marine Drugs. 2021;19(2):1-29. DOI: 10.3390/md19020115.

21. SUZUKI Eriko, et al. Direct cell–cell interaction regulates division of stem cells from PC-3 human prostate cancer cell line. Biochemical And Biophysical Research Communications. 2022;631:25-31. DOI: 10.1016/j.bbrc.2022.09.004.

22. Gao, Gang, et al. "Marine Antitumor Peptide Dolastatin 10: Biological Activity, Structural Modification and Synthetic Chemistry." Marine Drugs, vol. 19, no. 7, 2021, p. 363., DOI: 10.3390/md19070363.

23. SUAREZ-JIMENEZ Guadalupe-Miroslava, et al. Bioactive Peptides and Depsipeptides with Anticancer Potential: sources from marine animals. Marine Drugs. 2023 May 07;10(12):963-986. DOI: 10.3390/md10050963.

24. CHOUDHARY Alka, et al. Current Status and Future Prospects of Marine Natural Products (MNPs) as Antimicrobials. Marine Drugs. 2017;15(9):272-314. DOI: 10.3390/md15090272.

25. DUGGAN Peter, et al. Bioactive Mimetics of Conotoxins and other Venom Peptides. Toxins. 2015;7(10):4175-4198. DOI: 10.3390/toxins7104175.

26. GOGINENI Vedanjali, et al. Marine natural product peptides with therapeutic potential: chemistry, biosynthesis, and pharmacology. Biochimica Et Biophysica Acta (Bba) - General Subjects. 2018;1862(1):81-196. DOI: 10.1016/j.bbagen.2017.08.014.

27. GIACCO Luca del, et al. Introduction to Genomics. Methods In Molecular Biology. 2011;1(1):79-88. DOI: 10.1007/978-1-60327-216-2_6.

28. CHAMBERS Daniel C, et al. Transcriptomics and single-cell RNA-sequencing. Respirology. 2018;24(1):29-36. DOI: 10.1111/resp.13412.

29. BAGGERMAN Geert, et al. Peptidomics. J Chromatogr B Analyt Technol Biomed Life Sci. 2004;803(1):3-16. DOI: 10.1016/j.jchromb.2003.07.019.

30. ASLAM Bilal, et al. Proteomics: technologies and their applications. Journal Of Chromatographic Science. 2016;55(2):182-196. DOI: 10.1093/chromsci/bmw167.

31. RINSCHEN Markus M., et al. Identification of bioactive metabolites using activity metabolomics. Nature Reviews Molecular Cell Biology. 2019;20(6):353-367. DOI: 10.1038/s41580-019-0108-4.

32. KLOPFLEISCH Robert, et al. Anscriptome and Proteome Research in Veterinary Science: what is possible and what questions can be asked?. The Scientific World Journal. 2012;2012:1-14. DOI: 10.1100/2012/254962.

33. MANZONI Claudia, et al. Genome, transcriptome and proteom: the rise of omics data and their integration in biomedical sciences. Briefings In Bioinformatics. 2016;19(2):286-302. DOI: 10.1093/bib/bbw114.

34. KUMAR Dhirendra, et al. Integrating transcriptome and proteome profiling: strategies and applications. Proteomics. 2016;16(19):2533-2544. DOI: 10.1002/pmic.201600140.

35. OZSOLAK Fatih, et al. RNA sequencing: advances, challenges and opportunities. Nature Reviews Genetics. 2011;12(2):87-98. DOI: 10.1038/nrg2934.

36. WANG Zhong, et al. RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics. 2009;10(1):57-63. DOI: 10.1038/nrg2484.

37. MARTIN Jeffrey A, et al. Next-generation transcriptome assembly. Nature Reviews Genetics. 2011;12(10):671-682. DOI: 10.1038/nrg3068.

38. STEPHENSON Emily, et al. Ultiomics uncovers developing immunological lineages in human. European Journal Of Immunology. 2021;51(4):764-772. DOI: 10.1002/eji.202048769

39. MUNTEL Jan, et al. Surpassing 10 000 identified and quantified proteins in a single run by optimizing current LC-MS instrumentation and data analysis strategy. Molecular Omics. 2019;15(5):348-360. DOI : 10.1039/c9mo00082h

40. BEKKER-JENSEN Dorte B, et al. An Optimized Shotgun Strategy for the Rapid Generation of Comprehensive Human Proteomes. Cell Systems. 2017;4(6):587-599. DOI: 10.1016/j.cels.2017.05.009

41. GUO Qingxiang, et al. CCPRD: a novel analytical framework for the comprehensive proteomic reference database construction of nonmodel organisms. Acs Omega. 2020;5(25):15370-15384.DOI: 10.1021/acsomega.0c01278

42. AMARASINGHE Shanika L., et al. Opportunities and challenges in long-read sequencing data analysis. Genome Biology. 2020;21(1):1-16. DOI: 10.1186/s13059-020-1935-5

43. DEVABHAKTUNI Arun, et al. TagGraph reveals vast protein modification landscapes from large tandem mass spectrometry datasets. Nature Biotechnology. 2019;37(4):469-479. DOI: 10.1038/s41587-019-0067-5

44. MA Bin, et al. PEAKS: powerful software for peptidede novo sequencing by tandem mass spectrometry. Rapid Communications In Mass Spectrometry. 2003;17(20):2337-2342. DOI : 10.1002/rcm.1196

45. HECK Michelle, et al. Proteomics in Non-model Organisms: a new analytical frontier. Journal Of Proteome Research. 2020;19(9):3595-3606. DOI: 10.1021/acs.jproteome.0c00448.

46. FAO. FAO Yearbook: Fishery and Aquaculture Statistics 2018. 1st ed. Rome, Italy: FAO; 2020. 110 p. ISBN: 978-92-5-133371-6. DOI: 10.4060/cb1213t.

47. MENDO Jaime, et al. Bycatch and discards in the artisanal shrimp trawl fishery in Northern Peru. Plos One. 2022;17(6):1-13. DOI: 10.1371/journal.pone.0268128.

48. Alverson Dayton L., et al. A global assessment of fisheries bycatch and discards. Rome: FAO; 1994. ISBN: 92-5-103555-5.

49. GRAÇA LOPES Roberto, et al. FAUNA ACOMPANHANTE DA PESCA CAMAROEIRA NO LITORAL DO ESTADO DE SÃO PAULO, BRASIL. BVS-vet. 2002;28:173-188.

50. ROBERT Rafaela, et al. OS CAMARÕES DA FAUNA ACOMPANHANTE NA PESCA DIRIGIDA AO CAMARÃO-SETE-BARBAS (Xiphopenaeus kroyeri) NO LITORAL DO PARANÁ. B. Inst. Pesca. 2007;33(2):237-246.

51. DUNN Daniel C, et al. Spatio-temporal management of fisheries to reduce by-catch and increase fishing selectivity. Fish And Fisheries. 2011;12(1):110-119. DOI: 10.1111/j.1467-2979.2010.00388.x.

52. JANNOT Jason E, et al. Identifying ecological and fishing drivers of bycatch in a U.S. groundfish fishery. Ecological Applications. 2013;23(7):1645-1658. DOI: 10.1890/12-2225.1.

53. Olinto Branco, Joaquim, et al. Bycatch fauna of seabob shrimp trawl fisheries from Santa Catarina State, southern Brazil. Biota Neotropica. 2015;15(2):1-14..DOI: 10.1590/1676-06032015014314.

54. Penchaszadeh, P. E, et al. Bivalve contribution to shallow sandy bottom food web off Mar del Plata (Argentina):. Inference from stomach contents and stable isotope analysis. 2006;25(1): 51-54.DOI: 10.2983/0730-8000(2006)25[51: BCTSSB]2.0.CO;2.

55. Tangerina M. M. P, et al. Bioprospecting from cultivable bacterial communities of marine sediment and invertebrates from the underexplored Ubatuba region of Brazil. Arch Microbiol. 2017;199:155-169. DOI: 10.1007/s00203-016-1290-9.

56. GASU E. N., et al. Peptide extract from olivancillaria hiatula exhibits broad-spectrum antibacterial activity. BioMed Research International. 2018;2018:1-11. DOI: 10.1155/2018/6010572.

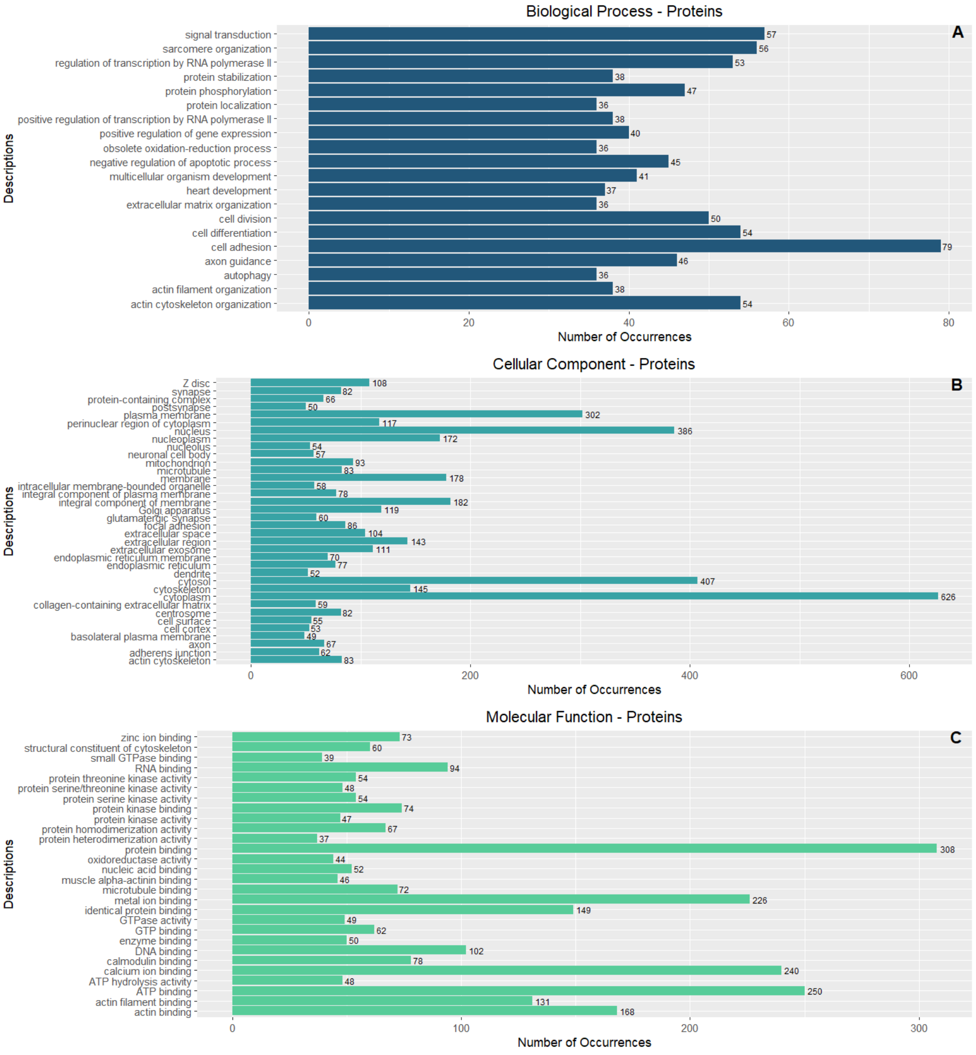
57. GASU E. N., et al. Peptide mix from olivancillaria hiatula interferes with cell-to-cell communication in pseudomonas aeruginosa. BioMed Research International. 2019;2019:1-12. DOI: 10.1155/2019/5313918.

58. TESO VALERIA, et al. A revision of the genus Olivancillaria (Mollusca: Olividae) from the southwestern Atlantic. Zootaxa. 2011;2889:1-34.

59. LIMEIRA JUNIOR Silvio C.M, et al. Characterization of the cross-lamellar structure of Olivancillaria urceus (Gastropoda: olividae) and its dissolution pattern. Micron. 2023;166. DOI: 10.1016/j.micron.2023.103416.

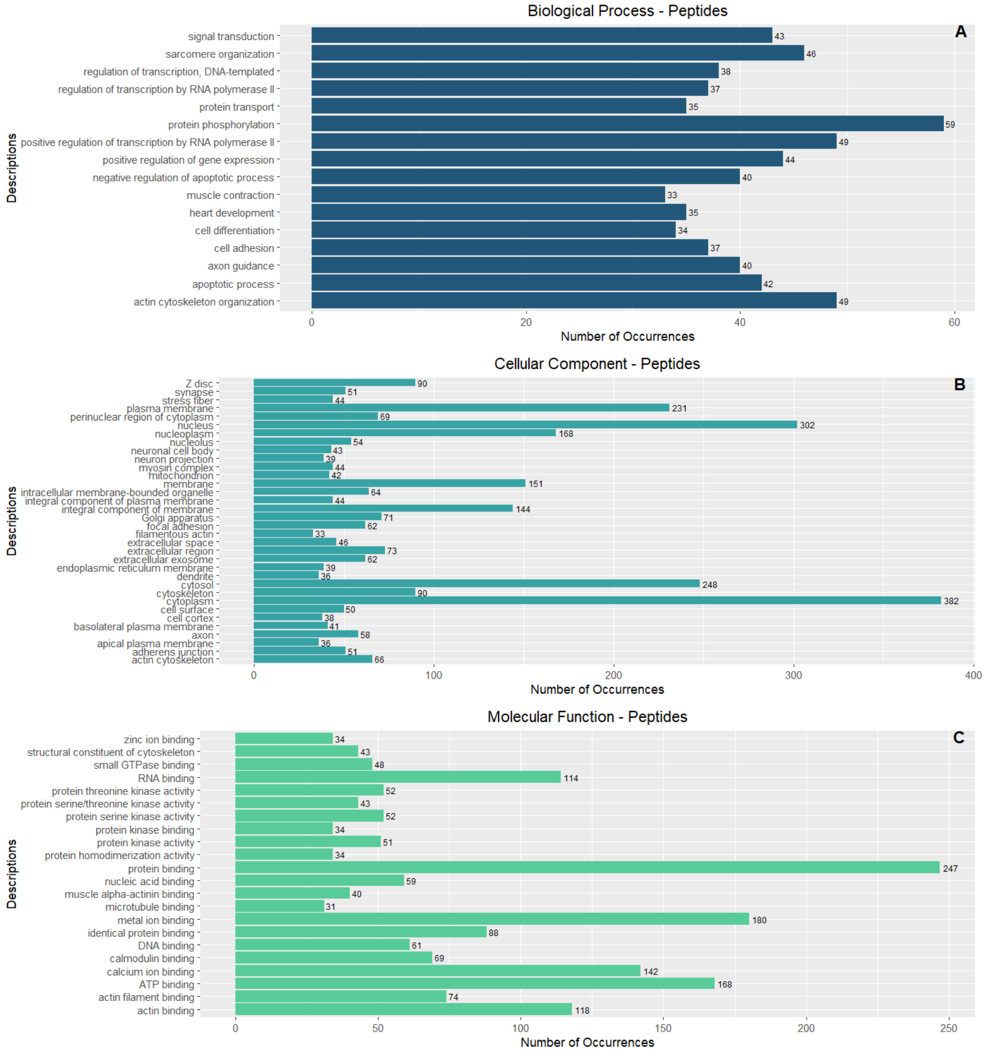
ANEXOS

Figura suplementar 1



Distribuição dos termos de Ontologia Genética (GO) para o proteoma da *Olivancillaria urceus*, classificados nas categorias de Processo Biológico (A), Componente Celular (B) e Função Molecular (C). Os gráficos de barras mostram o número de ocorrências para cada termo GO associado às proteínas identificadas. Na categoria de Processo Biológico, observa-se uma maior frequência de termos relacionados à organização do sarcômero e transdução de sinal. No Componente Celular, destaca-se a prevalência de proteínas localizadas no citoesqueleto e matriz extracelular. Na Função Molecular, as proteínas associadas à ligação ao actina e ao ATP apresentam alta frequência, indicando possíveis papéis em atividades estruturais e catalíticas dentro das células.

Figura suplementar 2



Distribuição dos termos de Ontologia Genética (GO) para o peptidoma de Olivancillaria urceus, classificados nas categorias de Processo Biológico (A), Componente Celular (B) e Função Molecular (C). Os gráficos de barras indicam o número de ocorrências para cada termo GO associado aos peptídeos identificados. Na categoria de Processo Biológico, observa-se uma maior frequência de termos relacionados à organização do sarcômero e transdução de sinal. No Componente Celular, os termos mais comuns incluem regiões do citoesqueleto e membranas plasmáticas. Na Função Molecular, as interações predominantes estão associadas à ligação ao ATP e actina, sugerindo a participação desses peptídeos em processos estruturais e de ligação energética.

Tabelas suplementares

As tabelas suplementares estão disponíveis para consulta no link abaixo:

https://drive.google.com/drive/folders/1J6QFBTDiLe5lljj6czhp1fVKCJR4P3-V?usp=drive_link