



**UNIVERSIDADE FEDERAL DO CEARÁ**  
**CENTRO DE CIÊNCIAS**  
**DEPARTAMENTO DE BIOLOGIA**  
**CURSO DE CIÊNCIAS BIOLÓGICAS**

**VITÓRIA LARA MARTINS MARQUES**

**CONTRAÇÃO E EXPANSÃO DE FAMÍLIAS DE PROTEÍNA DE CHOQUE  
TÉRMICO E A ADAPTAÇÃO AMBIENTAL EM MOLUSCOS**

**FORTALEZA**

**2022**

VITÓRIA LARA MARTINS MARQUES

CONTRAÇÃO E EXPANSÃO DE FAMÍLIAS DE PROTEÍNA DE CHOQUE TÉRMICO E  
A ADAPTAÇÃO AMBIENTAL EM MOLUSCOS

Monografia apresentada ao curso de Bacharelado em Ciências Biológicas do Departamento de Biologia da Universidade Federal do Ceará, como requisito parcial para obtenção do título de bacharel em Ciências Biológicas.

Orientador: Prof. Dr. Nicholas Costa Barroso Lima.

FORTALEZA

2022

Dados Internacionais de Catalogação na Publicação  
Universidade Federal do Ceará  
Biblioteca Universitária  
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

---

M321c Marques, Vitória Lara Martins.  
Contração e expansão de famílias de proteína de choque térmico e a adaptação ambiental em moluscos /  
Vitória Lara Martins Marques. – 2022.  
54 f. : il.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Centro de Ciências,  
Curso de Ciências Biológicas, Fortaleza, 2022.  
Orientação: Prof. Dr. Nicholas Costa Barroso Lima.

1. Moluscos. 2. Proteínas de choque térmico. 3. Adaptação ambiental. I. Título.

CDD 570

---

VITÓRIA LARA MARTINS MARQUES

CONTRAÇÃO E EXPANSÃO DE FAMÍLIAS DE PROTEÍNA DE CHOQUE TÉRMICO E  
A ADAPTAÇÃO AMBIENTAL EM MOLUSCOS

Monografia apresentada ao curso de Bacharelado em Ciências Biológicas do Departamento de Biologia da Universidade Federal do Ceará, como requisito parcial para obtenção do título de bacharel em Ciências Biológicas.

Aprovada em: \_\_/\_\_/\_\_\_\_.

BANCA EXAMINADORA

---

Prof. Dr. Nicholas Costa Barroso Lima (Orientador)  
Universidade Federal do Ceará (UFC)

---

Prof<sup>a</sup>. Dr<sup>a</sup>. Helena Matthews-Cascon  
Universidade Federal do Ceará (UFC)

---

Prof. Dr. Murilo Siqueira Alves  
Universidade Federal do Ceará (UFC)

## AGRADECIMENTOS

Ao curso e todo seu corpo docente, pela apresentação das diversas áreas da Biologia, possibilitando um encantamento particular por cada uma delas.

Em especial ao Prof. Dr. Nicholas, pela orientação e aulas ministradas ao longo do curso, as quais me apaixonaram não por um objeto de estudo, mas por uma forma de estudar.

Aos professores participantes da banca examinadora Prof<sup>ra</sup>. Dr<sup>a</sup>. Helena e Prof. Dr. Murilo pelo tempo, pelas colaborações e sugestões.

Aos diversos colegas de curso, em especial o que me acompanhou do início do mesmo até as cadeiras optativas, compartilhando o interesse por áreas de estudo similares e dividindo perspectivas.

Ao meu companheiro, pela paciência de me ouvir e confortar quantas vezes foram necessárias — muitas — e irritantemente me incentivar nos piores momentos.

A família, em especial minha irmã, pelos momentos de descontração e os vídeos de felinos compartilhados, bem como a constante cobrança pela minha conclusão de curso — aqui está.

“Life is just bytes and bytes and bytes of digital information.” (DAWKINS, 2008, p.19).

## RESUMO

Os moluscos compõem um filo diverso, tanto pela variedade de planos corporais quanto pela amplitude de habitats colonizados. Tal capacidade de adaptação a diferentes ambientes parece facilitar a ocorrência de “espécies invasoras” no filo, as quais impactam negativamente ecossistemas no mundo todo. No entanto, as origens genéticas dessa diversidade são pouco exploradas – é verificado que tal falta de representação se estende para os demais invertebrados, apesar destes somarem cerca de 95% de toda diversidade animal. Assim, o presente estudo propõe a utilização da bioinformática tanto para diminuir essa lacuna de conhecimento quanto para investigar a capacidade adaptativa de moluscos. Para tanto, optou-se por analisar a dinâmica de uma das principais famílias proteicas ligadas a resposta ao estresse – as proteínas de choque térmico. Estando estas relacionadas a resposta a diversos agentes estressores, a expansão de suas famílias poderia implicar no aumento da tolerância a ambientes diversos. Assim, foram obtidas montagens genômicas de moluscos disponíveis no NCBI, bem como as sequências de proteínas de choque térmico do filo. Para identificar porções codificantes destas proteínas em sequências que não apresentavam anotação, um genoma de referência foi selecionado. Além de verificar a abundância de regiões codificantes nos genomas, a formação de clusters entre as proteínas permitiu inferir suas famílias gênicas e seus padrões de expansão. Uma relação direta entre a proporção de proteínas de choque térmico e a capacidade de adaptação ambiental dos moluscos não foi detectada, no entanto, foram preditas 1.586 novas possíveis sequências de proteína e foi verificada uma expansão de HSP70 no filo. Além desta família, agrupamentos menores de HSP10, HSP60 e HSP90 foram detectados. Em bivalves, a expansão é mais expressiva, principalmente nas famílias Ostreidae e Pectinidae. Tais constatações fomentam a necessidade de mais estudos para averiguar a capacidade adaptativa dos moluscos.

**Palavras-chave:** moluscos; proteínas de choque térmico; adaptação ambiental.

## ABSTRACT

The molluscs are unique for the variety of bauplans and for the breadth of colonized habitats. Such ability to adapt to different environments seems to facilitate the occurrence of invasive species among the phylum, which negatively impact environments worldwide. However, the genetic origins of this diversity are little explored – in fact, this lack of representation extends to other invertebrates, despite the fact that they comprise 95% of the animal diversity. Thus, this study proposes the utilization of bioinformatics methods to reduce this knowledge gap and to investigate the adaptability of molluscs. Therefore, the dynamics of one of the main protein families linked to the stress response was investigated – the heat shock proteins (HSP). Since they are related to the response to different stressors, the expansion of their families may imply an increase in tolerance to multiple environments.. Thus, genomic assemblies of molluscs available in the NCBI were obtained, as well as their HSPs sequences. To identify these protein coding sequences in assemblies that were not annotated, a reference genome was selected. In addition to the abundance of protein coding portions in the genomes, the formation of clusters between them were investigated, allowing to infer their families and expansion patterns. A direct relationship between the proportion of HSPs and the environmental adaptability of molluscs was not identified. However, 1,586 new possible HSP sequences were predicted and an expansion of HSP70 in the phylum was verified. In addition to this family, smaller clusters of HSP10, HSP60 and HSP90 were also detected. In bivalves, the expansion is more expressive, mainly in the families Ostreidae and Pectinidae. Such findings encourage the need for further studies to ascertain the abiotic adaptability of molluscs.

**Keywords:** Mollusca; heat shock proteins; environmental adaptation.



## LISTA DE FIGURAS

Figura 1 – Representação das relações taxonômicas de Mollusca	19
Figura 2 – Estrutura geral de uma proteína de choque térmico	21

## LISTA DE TABELAS

Tabela 1 – Lista de sequências de molusco utilizadas no estudo, bem como o número de acesso correspondente do banco de dados utilizado, classificadas por classe .....	23
Tabela 2 – Genomas de moluscos com anotação disponível, classificados conforme o nível da montagem, o índice de completude (N50) e a quantidade de genes de HSP presentes na sequência.....	26
Tabela 3 – Quantidade de sequências de HSP detectadas por predição ou presentes na anotação dos genomas de molusco, classificados por classe .....	29
Tabela 4 – Espécies utilizadas no estudo, classificadas quanto a ocorrência ambiental e a capacidade invasora, seguidas das fontes utilizadas para coleta destas informações .....	32
Tabela 5 – Cluster de sequências de HSP correspondentes a cada organismo e grau de similaridade médio apresentado entre elas .....	37
Tabela 6 – Cluster de HSP10 detectado nos genomas de moluscos, junto ao grau de similaridade médio apresentado entre suas sequências .....	39
Tabela 7 – Cluster de HSP60 detectado nos genomas de moluscos, junto ao grau de similaridade médio apresentado entre suas sequências .....	40
Tabela 8 – Cluster de HSP90 detectado nos genomas de moluscos, junto ao grau de similaridade médio apresentado entre suas sequências .....	41
Tabela 9 – Quantidade de clusters específicos por espécie e de sequências de HSP neles agrupadas .....	44

## LISTA DE ABREVIATURAS E SIGLAS

ATP	Trifosfato de adenosina
DNA	Ácido desoxirribonucleico
GISD	Banco de dados de espécies globalmente invasivas, do inglês “ <i>Global Invasive Species Database</i> ”
HSP	Proteína de choque térmico, do inglês “ <i>Heat shock protein</i> ”
IUCN	União Internacional para a Conservação da Natureza, do inglês “ <i>International Union for Conservation of Nature</i> ”
NCBI	Centro Nacional de Informação Biotecnológica, do inglês “ <i>National Center for Biotechnology Information</i> ”

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> .....	14
<b>2</b>	<b>REFERENCIAL TEÓRICO</b> .....	18
<b>2.1</b>	<b>Filo Mollusca</b> .....	18
<b>2.2</b>	<b>Proteínas de choque térmico</b> .....	20
<b>3</b>	<b>OBJETIVOS</b> .....	22
<b>3.1</b>	<b>Objetivo geral</b> .....	22
<b>3.2</b>	<b>Objetivos específicos</b> .....	22
<b>4</b>	<b>MATERIAIS E MÉTODOS</b> .....	23
<b>4.1</b>	<b>Obtenção das sequências</b> .....	23
<b>4.2</b>	<b>Predição de proteínas nas sequências não anotadas</b> .....	27
<b>4.3</b>	<b>Clusterização das proteínas</b> .....	27
<b>4.4</b>	<b>Listagem de características das espécies</b> .....	27
<b>5</b>	<b>RESULTADOS E DISCUSSÃO</b> .....	29
<b>5.1</b>	<b>Quantidade de sequências de HSP</b> .....	29
<b>5.2</b>	<b>Características ambientais das espécies</b> .....	32
<b>5.3</b>	<b>Clusters de proteínas</b> .....	36
<b>5.3.1</b>	<i>Clusters com maior quantidade de sequências</i> .....	37
<b>5.3.2</b>	<i>Clusters restritos a um gênero</i> .....	42
<b>5.3.2.1</b>	<i>Crassostrea</i> .....	42
<b>5.3.2.2</b>	<i>Argopecten</i> .....	42
<b>5.3.2.3</b>	<i>Octopus</i> .....	43
<b>5.3.2.4</b>	<i>Pomacea, Conus e Haliotis</i> .....	43
<b>5.3.3</b>	<i>Clusters restritos a espécie</i> .....	44
<b>6</b>	<b>CONCLUSÃO</b> .....	47
	<b>REFERÊNCIAS</b> .....	49



## 1 INTRODUÇÃO

Desde a identificação do DNA como a *molécula transformadora*, ou carreadora de informação genética, o interesse por esse componente celular aumentou (AVERY; MACLEOD; MCCARTY, 1995). Isso porque as proteínas já eram conhecidas por seu papel enzimático e a composição repetitiva do código genético, em contraposição, não atraía os pesquisadores (PORTIN, 2014). Com a proposição do papel central deste na síntese proteica (CRICK, 1958), o interesse por decifrar o que seu código expressa é fomentado. O fato de que o mesmo era formado por quatro bases nitrogenadas já era conhecido, mas a forma como ocorria a leitura e a interpretação destas até a formação de cadeias de aminoácidos funcionais, as proteínas, só foi descoberta em 1965 (GRANTHAM et al., 1980).

Assim, a sequência das bases nitrogenadas já podia ser interpretada — restava sequenciar. Foi em 1977 que Sanger e colaboradores apresentaram uma forma de listar as bases que compunham um fragmento de DNA (SANGER; NICKLEN; COULSON, 1977). O conceito e a forma de performar eram relativamente simples, mas foi com a marcação dos didesoxinucleotídeos por fluorescência e outros aprimoramentos que o método pôde ser automatizado e, conseqüentemente, as dificuldades relacionadas ao sequenciamento foram consideravelmente diminuídas (HUNKAPILLER et al., 1991).

Com a automatização, foi possível iniciar o grande projeto de sequenciar o genoma humano, marcado pela colaboração de pesquisadores de todo mundo (MORAES; GÓES, 2016). Este inclui toda informação genética do organismo: tanto de genes quanto de regiões não codificantes do DNA, bem como o código genético de organelas — no caso de animais, somente o DNA mitocondrial (ALBERTS et al., 2002). O projeto demandou um grande investimento de tempo e dinheiro, levando treze anos para ser concluído e cerca de três bilhões de dólares. Hoje, o mesmo tipo de análise pode ser realizado para um único indivíduo, numa abordagem pessoal, de forma rápida e por cerca de mil dólares (GIBBS, 2020).

Foi a partir do desenvolvimento e popularização desses métodos que o número de sequências disponíveis nos bancos de dados biológicos cresceu exponencialmente, evidenciando a necessidade de uma abordagem computacional para lidar com as mesmas. Tal área de estudo interdisciplinar já existia — a bioinformática. Seu enfoque acompanhou a mudança de paradigma proposta pelo dogma central, passando de proteínas para a molécula de DNA. Com o sequenciamento do material genético, esse viés computacional foi fundamental tanto para a automatização dos processos quanto à montagem e interpretação das

sequências (GAUTHIER et al., 2019; VAN DIJK et al., 2014).

Essa alteração na forma de abordar a biologia se estende à interpretação de conceitos biológicos. O “gene”, por exemplo, que era abordado pela sua localização no cromossomo, passa a ser estudado em bancos de dados (SEARLS, 2010). A filogenia, que originalmente se baseava apenas em caracteres morfológicos e anatômicos, por vezes enriquecida pelo registro fóssil, pôde ser complementada pela comparação genômica (PATWARDHAN; RAY; ROY, 2014; SMITH et al., 2011). Em última instância, mudanças na arquitetura da sequência resultam da história evolutiva da espécie e podem refletir seu potencial adaptativo, criando pontes entre o genótipo e fenótipo (DUNN; MUNRO, 2016).

Para tanto, lacunas na representação da biodiversidade precisam ser preenchidas. Apesar da redução de custos, tempo e dificuldade para realizar sequenciamentos, muitos grupos continuam pouco documentados geneticamente. Em 2016, uma análise no banco de dados de sequências de referência do *National Center for Biotechnology Information* (NCBI) contabilizou 170 genomas de vertebrados e 46 de invertebrados. Esta categoria “de referência” se limita a sequências que apresentam anotação e passaram por curadoria. Ou seja, têm identificação estrutural e funcional de suas regiões gênicas, além de uma revisão nesta e em outras etapas de montagem da sequência, reduzindo a ocorrência de erros e de redundância (O’LEARY et al., 2016).

No conjunto geral de genomas sequenciados, uma comparação realizada em 2018 apenas entre invertebrados aponta uma enorme concentração em Arthropoda. Esta seria justificável pela sua posição de filo mais diverso do reino animal, caso os sequenciamentos não fossem compostos quase unicamente por membros de duas de suas classes. Além disso, nos demais filós, o número de sequências não se mostra proporcional à variedade de espécies. Nematoda, com quase uma centena de genomas sequenciados, supera Mollusca, mesmo que este ocupe a segunda posição em diversidade animal (LOPEZ et al., 2019; ZHANG, 2013).

Sendo o grupo dos invertebrados uma convenção, visto que não apresenta origem monofilética, estender a representação genômica de um ou dois filós pela pluralidade que o termo engloba é impossível (GIGA COMMUNITY OF SCIENTISTS et al., 2014). A disparidade fica mais evidente sabendo que este compõe cerca de 95% da diversidade animal, mas que, como apontado anteriormente, não tem o equivalente a um terço da quantidade de genomas de vertebrados sequenciados (LOPEZ et al., 2019; ZHANG, 2013). Em uma realidade de crescente degradação ambiental e perda irreversível de biodiversidade, parece contraditório não se dedicar a conhecer a maior porção dela enquanto há tempo (LEWIN et al., 2018).

Com a maior quantidade de espécies do reino animal, os invertebrados também ocupam posições ecológicas variadas. Auxiliam na produção primária de energia pelos organismos autotróficos por meio de serviços ecológicos de polinização (KNIGHT et al., 2018) e dispersão de sementes (SANTANA et al., 2016). Como consumidores, têm representantes herbívoros, filtradores e predadores e, na reciclagem da matéria orgânica, facilitam a decomposição atuando como detritívoros (HASZPRUNAR; WANNINGER, 2012). Essa quase onipresença dos invertebrados em funções ecológicas e sua elevada diversidade pode ajudar a promover a estabilidade da rede trófica. Ademais, detêm grande influência em seu entorno, promovendo alterações no ambiente e na composição biótica do mesmo (PRATHER et al., 2013).

No que diz respeito aos moluscos, além de estarem presentes em vários níveis tróficos, apresentam morfologia e fisiologia muito diversa, com grande pluralidade de planos corporais e modos de vida (SIGWART, 2017). Seus membros estão presentes em habitats variados, de fontes hidrotermais (SUN et al., 2017) a desertos (SCHMIDT-NIELSEN; TAYLOR; SHKOLNIK, 1971), apresentando múltiplas e independentes transições de ambiente em um mesmo táxon (ROMERO et al., 2016).

Essas circunstâncias exibem a plasticidade do grupo e reforçam a importância de estudá-lo geneticamente, buscando a origem desse potencial. Tal compreensão traz benefícios à medicina, visto que determinados moluscos são hospedeiros intermediários de parasitas cujos hospedeiros definitivos são humanos (MADSEN; FRANSEN, 1989; BAYNE, 2009; THIENGO et al., 2010; MAHULU et al., 2019). Para a agricultura, dado que algumas espécies são consideradas “pestes” em plantações pelo mundo (SIN, 2003; KOZŁOWSKI, 2009; BARONIO et al., 2014). E, por fim, à conservação, especialmente pelas recorrentes bioinvasões protagonizadas pelos moluscos, em outra demonstração de seu potencial adaptativo (KNOP; REUSSER, 2012; ULIANO-SILVA et al., 2015).

Estas impactam todo o ambiente “invadido” e organismos relacionados a ele, com consequências negativas à conservação de várias espécies — inclusive outros moluscos — dado que muitas das extinções ou diminuições populacionais de linhagens nativas têm origem na introdução de espécies exógenas (LYDEARD et al., 2004). Isso ocorre pois, além de não contar com predadores naturais no ambiente exótico e competir com as espécies nativas, as introduzidas frequentemente provocam alterações no meio, sendo denominadas “engenheiras de ecossistema” (SOUSA; GUTIÉRREZ; ALDRIDGE, 2009; BURLAKOVA; KARATAYEV; KARATAYEV, 2012; SIMBERLOFF et al., 2013). Não obstante, esse fenômeno tem forte repercussão econômica. Dentre os moluscos, os impactos dos bivalves



são os mais conhecidos e se estendem do entupimento de canos em usinas de geração de energia, irrigação e filtração de água (CONNELLY et al., 2007) à corrosão e desgaste de estruturas de concreto e metal, como pontes e barcos (BOLTOVSKOY; XU; NAKANO, 2015).

Antes de categorizar os supostos invasores como “vilões”, é importante notar que os moluscos são frequentemente carregados pelos próprios humanos à outros territórios, fenômeno este que ocorre através da construção de canais, da migração e do estabelecimento de rotas comerciais (KARATAYEV et al., 2007). Essa diferença é vital, pois as introduções provocadas pela ação humana apresentam maior frequência, extensão geográfica e, por consequência, probabilidade de impacto ambiental (RICCIARDI, 2007). Ocorrem inclusive introduções intencionais, para fomentar a criação de espécies de interesse econômico em outros territórios ou mesmo para a indústria de *pets* (NUNES et al., 2015; YANAI et al., 2017). Potencializada pela crescente globalização, essa dispersão tende a aumentar. Identificar a capacidade invasiva dos organismos de antemão permite a tomada de decisões mais responsáveis ao invés de seguir na tentativa de mitigar os efeitos dessas introduções (HULME, 2009).

Aqui, a análise bioinformática se apresenta como uma poderosa e pouco onerosa ferramenta que, aplicada a pesquisa de grupos pouco estudados, ajudaria diminuir a desproporção de informação genética que existe entre os filos. Por meio de métodos e abordagens estatísticas, é possível inferir significado biológico e elucidar processos evolutivos (STEVENS, 2013). Para estudar as origens genéticas da adaptação ambiental dos moluscos, o presente estudo se dedica a quantificação de famílias proteicas relacionadas a essa característica, as proteínas de choque térmico (HSPs). Estas foram selecionadas considerando o alto grau de conservação de seus membros, sua atuação como uma resposta celular fisiológica comum para tolerar estímulos estressores e a responsividade das mesmas a vários desses estímulos (HOFMANN; TODGHAM, 2010). Assim, espera-se relacionar a capacidade de adaptação a condições abióticas diversas dos moluscos a processos de expansão e contração dessas proteínas em seus genomas.

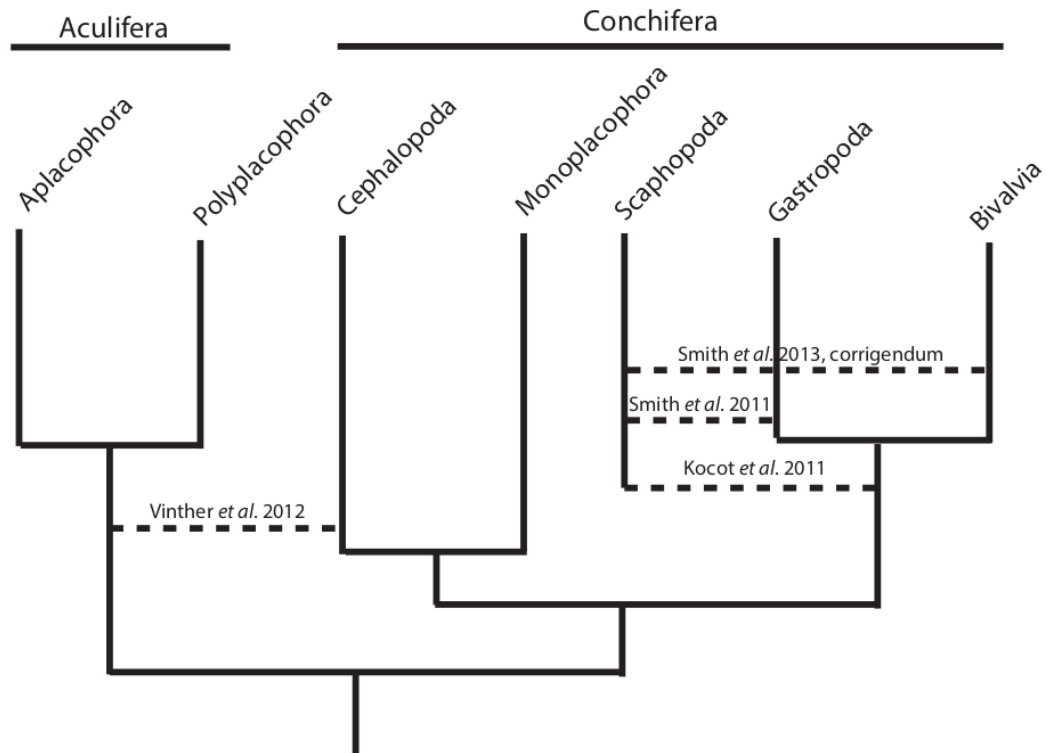
## 2 REFERENCIAL TEÓRICO

### 2.1 Filo Mollusca

O filo é composto por oito classes, das quais são mais conhecidos os representantes de Cephalopoda, Gastropoda e Bivalvia; e menos familiares os de Scaphopoda, Monoplacophora, Polyplacophora, Solenogastres (Neomenimorpha) e Caudofoveata (Chaetodermomorpha) (WANNINGER; WOLLESEN, 2019). Historicamente, a classificação taxonômica era fortemente baseada na concha, quando presente (COLLEY; SIMONE; DE LOYOLA E SILVA, 2012). Assim, estabeleceu-se um padrão de planos corporais representantes das classes, onde os bivalves estão associados à presença de concha em duas valvas, os cefalópodes, à concha internalizada e os gastrópodes à existência de uma única concha ou nenhuma – no caso de lesmas. Em Scaphopoda, a classe menos conhecida do filo, a concha é tubular. Monoplacophora e Polyplacophora apresentam, como os nomes sugerem, uma concha e várias placas, respectivamente. Já nos Aplacophora (grupo que engloba Caudofoveata e Solenogastres) a concha é ausente (PONDER; LINDBERG; PONDER, 2020).

A despeito dessas generalizações, os fenótipos do grupo são extremamente variáveis, como exemplifica Sigwart (2017) ao citar casos de confusão taxonômica provocados pela quantidade de conchas, a ocorrência de gigantismo em organismos contemporâneos, entre outras disparidades. Anteriormente, a filogenia se baseava apenas em dados morfológicos, propondo uma crescente complexidade dos planos corporais. Assim, o ancestral comum do grupo seria desprovido de concha, como um Aplacophora atual. Hoje, entende-se que a perda da concha nesse grupo se deu de forma secundária, com o ancestral do filo já apresentando esse caractere, de forma não mineralizada. Foi a partir da interação de dados moleculares, de desenvolvimento em diferentes estágios de vida e do registro fóssil que se estabeleceram estas relações representadas na Figura 1 (VINTHER, 2015).

**Figura 1** – Representação das relações taxonômicas de Mollusca.



**Fonte:** Vinther (2015).

Equiparando-se à diversidade de planos corporais, os habitats colonizados por moluscos são numerosos, sendo estes encontrados em ambientes de dulcícolas, marinhos e terrestres. A classe Gastropoda é particularmente diversificada, com representantes em todos os habitats citados. Nesta, a transição para o ambiente terrestre é marcante por ter ocorrido múltiplas vezes no táxon de forma independente (KANO et al., 2015). Dentre os bivalves, a maior demonstração de resistência ambiental está nos mexilhões que habitam a zona entremarés, por tal área apresentar grande variação em todos componentes abióticos em um curto período de tempo (TOMANEK; HELMUTH, 2002). Tal classe é constituída tanto por espécies marinhas quanto de água doce.

São também comuns na região entremarés os quítons, representantes de Polyplacophora, com a presença de algumas espécies em águas profundas. Neste habitat estão todos os membros de Monoplacophora – com o grupo já tendo sido considerado extinto, em razão da dificuldade de acesso a esse ambiente. Essas classes, bem como Aplacophora, Scaphopoda e Cephalopoda, são exclusivamente marinhas (PONDER; LINDBERG; PONDER, 2020).

## 2.2 Proteínas de choque térmico

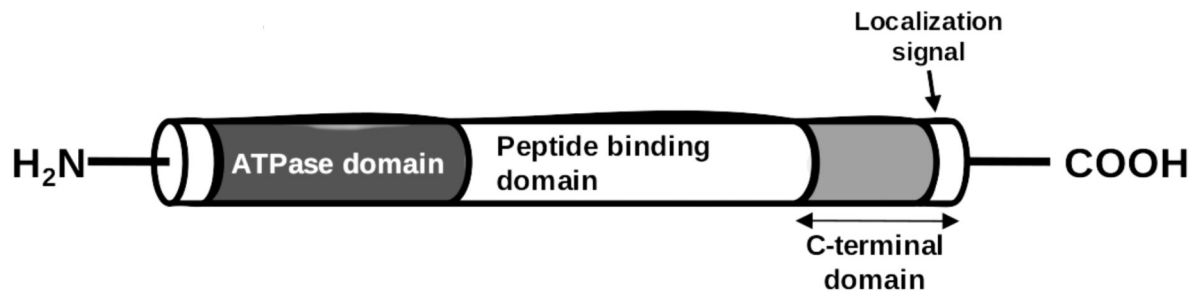
As proteínas de choque térmico são altamente conservadas na natureza e atuam como chaperonas moleculares de outras proteínas, reparando e impedindo dobramentos incorretos ou ligações inapropriadas. Ademais, auxiliam no transporte dessas moléculas entre os compartimentos celulares. Apesar dessas funções serem executadas em condições fisiológicas, a perda de estabilidade proteica sob ação de agentes estressantes aumenta a expressão das HSPs (FEDER; HOFMANN, 1999).

A despeito do nome estar associado a temperatura, vários fatores endógenos e exógenos culminam no aumento da expressão destas proteínas. Como exemplo desses estímulos estão a exposição à radiação ou a metais pesados, privação de nutrientes, hipóxia, infecções e inflamações, dentre outros (LINDQUIST; CRAIG, 1988). Como sintetizam Feder e Hofmann (1999), todas as formas de estresse conhecidas, se suficientemente intensas, induzem a expressão dessas proteínas.

Estas são organizadas em famílias, com base na similaridade de suas sequências codificantes e no peso molecular das proteínas, sendo, em geral, classificadas em HSP10, HSP40, HSP60, HSP70, HSP90, HSP100 e HSP110 (JEE, 2016). Também se realiza uma associação desses grupos a funções ou fenótipos específicos, mas esta é questionada pela dificuldade de determinar o estresse ao qual o aumento de expressão se relaciona, uma vez que raramente organismos enfrentam uma única adversidade por vez (FEDER; HOFMANN, 1999). Além disso, uma mesma família pode atuar na resposta a diversos agentes estressantes (SØRENSEN; KRISTENSEN; LOESCHCKE, 2003).

A família mais estudada é HSP70. Isso ocorre por a mesma ser a mais conservada não só dentre as demais HSPs, mas entre as proteínas no geral. Seus representantes são encontrados em todos os organismos e com tal grau de conservação que as sequências de aminoácido compartilham relevante similaridade, mesmo considerando grandes distâncias filogenéticas (DAUGAARD; ROHDE; JÄÄTTELÄ, 2007). Uma ilustração da sua estrutura geral é indicada abaixo (Figura 2), mas essa generalização é aplicável às demais famílias de HSP (GARRIDO et al., 2012; VOS et al., 2008). Nela, estão representados o domínio de ligação ao ATP, o de ligação ao substrato e o C-terminal, domínio que permite sua ligação a cofatores e a outras HSPs.

**Figura 2** – Estrutura geral de uma proteína de choque térmico.



**Fonte:** adaptado de Fabbri, Valbonesi e Franzelliti (2008).

### **3 OBJETIVOS**

#### **3.1 Objetivo geral**

Identificar se existe relação entre a proporção de proteínas de choque térmico nos genomas de moluscos e a capacidade de adaptação ambiental destes animais.

#### **3.2 Objetivos específicos**

- Identificar proteínas de famílias de HSP nos genomas de moluscos disponíveis;
- Mensurar a presença das mesmas dentre os genomas;
- Descrever características dos organismos incluídos no estudo que estejam relacionadas à adaptação ambiental;
- Avaliar se existe relação entre o histórico invasivo de alguns desses organismos com a proporção dessas famílias em suas sequências.

## 4 MATERIAIS E MÉTODOS

### 4.1 Obtenção das sequências

Os genomas utilizados foram adquiridos no banco de dados NCBI (SAYERS et al., 2022), a partir da busca por montagens de sequência de moluscos, e estão listados abaixo (Tabela 1), junto ao número de acesso. Foram adicionados ao estudo genomas disponibilizados até março de 2020. Estes totalizaram 46 sequências, das quais 21 são de bivalves, 20 de gastrópodes e 5 de cefalópodes, não constando no banco de dados montagens das demais classes.

**Tabela 1** – Lista de sequências de molusco utilizadas no estudo, bem como o número de acesso correspondente do banco de dados utilizado, classificadas por classe.

Classe	Organismo	Número de acesso
<b>Gastropoda</b>	<i>Achatina immaculata</i>	GCA_009760885.1
	<i>Anentome helena</i>	GCA_009936545.1
	<i>Aplysia californica</i>	GCF_000002075.1
	<i>Babylonia areolata</i>	GCA_011634625.1
	<i>Biomphalaria glabrata</i>	GCF_000457365.1
	<i>Colubraria reticulata</i>	GCA_900004695.1
	<i>Conus consors</i>	GCA_004193615.1
	<i>Conus tribblei</i>	GCA_001262575.1
	<i>Elysia chlorotica</i>	GCA_003991915.1
	<i>Haliotis laevigata</i>	GCA_008038995.1
	<i>Haliotis rubra</i>	GCA_003918875.1
	<i>Haliotis rufescens</i>	GCA_003343065.1

	<i>Lanistes nyassanus</i>	GCA_004794575.1
	<i>Lottia gigantea</i>	GCF_000327385.1
	<i>Lymnaea stagnalis</i>	GCA_900036025.1
	<i>Marisa cornuarietis</i>	GCA_004794655.1
	<i>Physella acuta</i>	GCA_004329575.1
	<i>Pomacea canaliculata</i>	GCF_003073045.1
	<i>Pomacea maculata</i>	GCA_004794325.1
	<i>Radix auricularia</i>	GCA_002072015.1
<b>Cephalopoda</b>	<i>Architeuthis dux</i>	GCA_006491835.1
	<i>Euprymna scolopes</i>	GCA_004765925.1
	<i>Octopus bimaculoides</i>	GCF_001194135.1
	<i>Octopus sinensis</i>	GCF_006345805.1
	<i>Octopus vulgaris</i>	GCA_003957725.1
<b>Bivalvia</b>	<i>Argopecten irradians concentricus</i>	GCA_004382765.1
	<i>Argopecten irradians irradians</i>	GCA_004382745.1
	<i>Bathymodiolus platifrons</i>	GCA_002080005.1
	<i>Corbicula fluminea</i>	GCA_001632725.1
	<i>Crassostrea gigas</i>	GCF_902806645.1
	<i>Crassostrea virginica</i>	GCF_002022765.2
	<i>Cyclina sinensis</i>	GCA_012932295.1
	<i>Dreissena polymorpha</i>	GCA_000806325.1
	<i>Dreissena rostriformis</i>	GCA_007657795.1



<i>Limacina bulimoides</i>	GCA_009866985.1
<i>Limnoperna fortunei</i>	GCA_003130415.1
<i>Lutraria rhynchaena</i>	GCA_008271625.1
<i>Mizuhopecten yessoensis</i>	GCF_002113885.1
<i>Modiolus philippinarum</i>	GCA_002080025.1
<i>Mytilus galloprovincialis</i>	GCA_900618805.1
<i>Pecten maximus</i>	GCF_902652985.1
<i>Pinctada imbricata</i>	GCA_002216045.1
<i>Ruditapes philippinarum</i>	GCA_009026015.1
<i>Saccostrea glomerata</i>	GCA_003671525.1
<i>Sinonovacula constricta</i>	GCA_009762815.1
<i>Venustaconcha ellipsiformis</i>	GCA_003401595.1

**Fonte:** elaborada pela autora (2022).

Foram obtidas tanto a sequência de DNA em formato *fasta* quanto as de proteínas HSP de cada organismo, quando presentes. No entanto, com apenas 9 das 46 sequências apresentando anotação, foi necessário estender essa informação para a maior parte dos organismos. Para isso, foi selecionada uma sequência referencial entre as que possuíam anotação. Esta se deu a partir de critérios que priorizam montagens em nível de cromossomo, com boas estatísticas de completude e com a maior quantidade de HSPs. Dentre as disponíveis no grau de montagem desejado, a com maior estatística de contiguidade (N50) e abundância de HSPs disponível foi a do bivalve *Crassostrea virginica*, sendo esta, portanto, a sequência selecionada como referência.

**Tabela 2** – Genomas de moluscos com anotação disponível, classificados conforme o nível da montagem, o índice de completude (N50) e a quantidade de genes de HSP presentes na sequência.

<b>Espécie</b>	<b>Nível de montagem</b>	<b>N50</b>	<b>Quantidade de porções codificantes de HSPs</b>
<i>Aplysia californica</i>	Scaffold	917541	37
<i>Biomphalaria glabrata</i>	Scaffold	48059	84
<i>Crassostrea gigas</i>	Cromossomo	58462999	251
<i>Crassostrea virginica</i>	Cromossomo	75944018	278
<i>Mizuhopecten yessoensis</i>	Scaffold	803631	182
<i>Octopus bimaculoides</i>	Scaffold	475182	44
<i>Octopus sinensis</i>	Cromossomo	105892736	59
<i>Pecten maximus</i>	Cromossomo	44824366	115
<i>Pomacea canaliculata</i>	Cromossomo	31531291	46

**Fonte:** elaborada pela autora (2022).

## 4.2 Predição de proteínas nas sequências não anotadas

As sequências de aminoácido de HSPs de *Crassostrea virginica* foram, então, utilizadas para busca por homologia nos genomas dos moluscos não anotados, utilizando o *software* de alinhamento de sequências Exonerate (SLATER; BIRNEY, 2005). Dessa busca, foram selecionadas apenas sequências com pelo menos 60% de similaridade com as de *Crassostrea*, limiar determinado para evitar alinhamentos espúrios. Tais porções gênicas identificadas como homólogas às proteínas de *Crassostrea* foram posteriormente traduzidas utilizando o mesmo *software*. Assim, foi obtido um conjunto de sequências de HSP dos moluscos que não possuíam anotação. Para reforçar essa predição, as proteínas de *Crassostrea* foram submetidas a análise pelo *software* InterproScan (JONES et al., 2014). Este busca por características da sequência submetida no conjunto de bancos de dados proteicos InterPro a fim de identificar famílias, domínios e sítios funcionais nas mesmas (MULDER; APWEILER, 2007). Essa identificação é estendida às de HSP homólogas a elas presentes nos genomas não anotados, adicionando confiabilidade à predição.

## 4.3 Clusterização das proteínas

Uma comparação geral das sequências de proteínas de todos os organismos foi realizada utilizando o *software* CD-HIT (LI; GODZIK, 2006; FU et al., 2012). Neste, foram efetuados alinhamentos múltiplos entre todas as sequências de HSP para identificar e agrupar proteínas intimamente relacionadas entre si. Pela existência desta relação entre as sequências agrupadas, anotações de famílias de HSP identificadas pela análise anterior, no InterProScan, foram estendidas às demais proteínas quando estas ocorriam em um mesmo grupo.

Para a clusterização, foram admitidas apenas sequências com o mínimo de 60% de similaridade em relação à sequência referencial. Esta é selecionada pelo próprio *software* como centróide do cluster para formar os agrupamentos.

## 4.4 Listagem de características das espécies

Foram levantados os hábitos ambientais dos organismos incluídos na pesquisa, a fim de relacionar estes a presença e variedade das sequências de HSP em seus genomas. Características de distribuição geográfica, tipo de habitat colonizado e presença de histórico invasor foram listadas para cada indivíduo.

Tais informações foram obtidas nos bancos de dados MolluscaBase (<https://www.molluscabase.org/>), *Global Invasive Species Database* (GISD - <http://www.iucngisd.org/gisd/>) e IUCN *Red List of Threatened Species* (<https://www.iucnredlist.org/>). Ademais, algumas das montagens registradas no NCBI possuíam artigo vinculado e, nesses casos, o mesmo foi utilizado para acessar tais informações da respectiva espécie.

## 5 RESULTADOS E DISCUSSÃO

### 5.1 Quantidade de sequências de HSP

A busca por HSPs nas montagens não anotadas detectou um total de 1.586 sequências, com a quantidade de correspondências a proteínas de *Crassostrea* listadas por organismo abaixo (Tabela 3). Entre os nove genomas anotados, foram obtidas no banco de dados 1.136 sequências de HSP. Estas estão também listadas na tabela, sendo destacadas das demais pela cor.

**Tabela 3** – Quantidade de sequências de HSP detectadas por predição ou presentes na anotação dos genomas de molusco, classificados por classe.

<b>Espécie</b>	<b>Quantidade de sequências</b>	<b>Quantidade total de sequências por classe</b>
<i>Achatina immaculata</i>	76	<b>Gastropoda</b> 808
<i>Anentome helena</i>	12	
<i>Aplysia californica</i>	37	
<i>Babylonia areolata</i>	25	
<i>Biomphalaria glabrata</i>	84	
<i>Colubraria reticulata</i>	13	
<i>Conus consors</i>	33	
<i>Conus tribblei</i>	19	
<i>Elysia chlorotica</i>	36	
<i>Haliotis laevigata</i>	43	
<i>Haliotis rubra</i>	39	
<i>Haliotis rufescens</i>	55	
<i>Lanistes nyassanus</i>	22	

<i>Lottia gigantea</i>	40	
<i>Lymnaea stagnalis</i>	51	
<i>Marisa cornuarietis</i>	42	
<i>Physella acuta</i>	26	
<i>Pomacea canaliculata</i>	46	
<i>Pomacea maculata</i>	42	
<i>Radix auricularia</i>	67	
<i>Architeuthis dux</i>	24	<b>Cephalopoda</b> 170
<i>Euprymna scolopes</i>	30	
<i>Octopus bimaculoides</i>	44	
<i>Octopus sinensis</i>	59	
<i>Octopus vulgaris</i>	13	
<i>Argopecten irradians concentricus</i>	65	<b>Bivalvia</b> 1744
<i>Argopecten irradians irradians</i>	58	
<i>Bathymodiolus platifrons</i>	77	
<i>Corbicula fluminea</i>	0	
<i>Crassostrea gigas</i>	251	
<i>Crassostrea virginica</i>	278	
<i>Cyclina sinensis</i>	57	
<i>Dreissena rostriformis</i>	62	
<i>Dreissena polymorpha</i>	0	
<i>Limacina bulimoides</i>	23	

<i>Limnoperna fortunei</i>	43
<i>Lutraria rhynchaena</i>	52
<i>Mizuhopecten yessoensis</i>	182
<i>Modiolus philippinarum</i>	57
<i>Mytilus galloprovincialis</i>	33
<i>Pecten maximus</i>	115
<i>Pinctada imbricata</i>	70
<i>Ruditapes philippinarum</i>	42
<i>Saccostrea glomerata</i>	165
<i>Sinonovacula constricta</i>	50
<i>Venustaconcha ellipsiformis</i>	24

**Fonte:** elaborada pela autora (2022).

Nas sequências de *Corbicula fluminea* e *Dreissena polymorpha* não foi possível detectar HSPs homólogas às do genoma de referência. No entanto, esse resultado decorre da incompletude dessas montagens – são as menores da amostra – e não da inexistência dessas proteínas nesses organismos.

Também é perceptível a maior quantidade de sequências de HSP em bivalves, que apresenta cerca do dobro da quantidade de sequências presentes em gastrópodes. Tal proporção seria justificável pelo fato de que a montagem utilizada como referência para busca nos demais organismos foi de um bivalve, *Crassostrea virginica*, o que poderia favorecer a busca por similaridade entre organismos da mesma classe. No entanto, essa concentração se repete nos genomas já anotados, cujas sequências não foram apuradas por essa mesma homologia. Na literatura, essa expansão de genes de HSP em bivalves já foi observada e relacionada ao modo de vida sésil de muitos desses animais (TAKEUCHI et al. 2016)

## 5.2 Características ambientais das espécies

Informações correspondentes a ocorrência geográfica, aos habitats colonizados e a existência ou não de histórico invasivo de cada espécie estudada estão listadas abaixo (Tabela 4), acompanhadas da fonte que foi utilizada para corroborar essa classificação.

**Tabela 4** – Espécies utilizadas no estudo, classificadas quanto a ocorrência ambiental das mesmas e a capacidade invasora, seguidas das fontes utilizadas para coleta destas informações.

<b>Espécies</b>	<b>Distribuição geográfica</b>	<b>Habitat</b>	<b>Invasora?</b>	<b>Referência</b>
<i>Achatina immaculata</i>	África	Terrestre	Não	1, 2 e Liu et al. (2021)
<i>Anentome helena</i>	Ásia	Água doce	Não	1
<i>Aplysia californica</i>	América do Norte	Marinho	Não	1
<i>Babylonia areolata</i>	Ásia	Marinho	Não	1
<i>Biomphalaria glabrata</i>	África	Água doce	Não	1 e Adema et al. (2017)
<i>Colubraria reticulata</i>	África e Europa	Marinho	Não	1
<i>Conus consors</i>	África, Ásia e Oceania	Marinho	Não	1 e 3
<i>Conus tribblei</i>	Ásia e Oceania	Marinho	Não	1 e 3
<i>Elysia chlorotica</i>	América do Norte	Marinho	Não	1 e Cai et al. (2019)
<i>Haliotis laevigata</i>	Oceania	Marinho	Não	1, 3 e Botwright et al. (2019)



<i>Haliotis rubra</i>	Oceania	Marinho	Não	1, 3 e Masonbrink et al. (2019)
<i>Haliotis rufescens</i>	América do Norte	Marinho	Não	1, 3 e Masonbrink et al. (2019)
<i>Lanistes nyassanus</i>	África	Água doce	Não	1, 3 e Sun et al. (2019)
<i>Lottia gigantea</i>	América do Norte	Marinho	Não	1
<i>Lymnaea stagnalis</i>	Europa e América do Norte	Água doce e salobra	Não	1 e 3
<i>Marisa cornuarietis</i>	América do Sul e África	Água doce	Não	1 e Sun et al. (2019)
<i>Physella acuta</i>	Global	Água doce	Sim	1, 2 e 3
<i>Pomacea canaliculata</i>	América, Ásia e Europa	Água doce	Sim	1 e Sun et al. (2019)
<i>Pomacea maculata</i>	América, Ásia e Europa	Água doce	Sim	1 e Sun et al. (2019)
<i>Radix auricularia</i>	América do Norte, África e Ásia	Água doce	Sim	1, 3 e Feldmeyer et al. (2015)
<i>Architeuthis dux</i>	Global	Marinho	Não	1 e Da Fonseca et al. (2020)
<i>Euprymna scolopes</i>	América do Norte	Marinho	Não	1 e Belcaid et al. (2019)
<i>Octopus</i>	América	Marinho	Não	1 e Albertin

<i>bimaculoides</i>				et al. (2015)
<i>Octopus sinensis</i>	Ásia	Marinho	Não	1
<i>Octopus vulgaris</i>	Global	Marinho	Não	1, 3 e Zarrella et al. (2019)
<hr/>				
<i>Argopecten irradians concentricus</i>	América do Norte	Marinho	Não	1
<i>Argopecten irradians irradians</i>	América do Norte	Marinho	Não	1
<i>Bathymodiolus platifrons</i>	Ásia	Marinho	Não	1 e Sun et al. (2017)
<i>Corbicula fluminea</i>	Global	Água doce	Sim	1, 2 e 3
<i>Crassostrea gigas</i>	Global	Marinho	Sim	1, 2 e Zhang et al. (2012)
<i>Crassostrea virginica</i>	América, África e Europa	Marinho	Não	1
<i>Cyclina sinensis</i>	Ásia	Marinho	Não	1
<i>Dreissena polymorpha</i>	Global	Água doce e salobra	Sim	1, 2, 3 e McCartney et al. ()
<i>Dreissena rostriformis</i>	Europa e Ásia	Água doce e salobra	Sim	1, 3 e Calcino et al. (2019)
<i>Limacina bulimoides</i>	Global	Marinho	Não	1

<i>Limnoperna fortunei</i>	Ásia e América	Água doce	Sim	1, 2, 3 e Uliano-Silva et al. (2018)
<i>Lutraria rhynchaena</i>	África, Ásia e Oceania	Marinho	Não	1 e Thai et al. (2019)
<i>Mizuhopecten yessoensis</i>	África e Ásia	Marinho	Não	1
<i>Modiolus philippinarum</i>	África e Ásia	Marinho	Não	1 e Sun et al. (2017)
<i>Mytilus galloprovincialis</i>	Global	Marinho	Sim	1 e 2
<i>Pecten maximus</i>	Europa	Marinho	Não	1 e Kenny et al. (2020)
<i>Pinctada imbricata</i>	América, África e Ásia	Marinho	Não	1
<i>Ruditapes philippinarum</i>	América do Norte, Ásia e Europa	Marinho	Não	1 e Mun et al. (2017)
<i>Saccostrea glomerata</i>	Europa, Ásia e Oceania	Marinho	Não	1 e Powell et al. (2018)
<i>Sinonovacula constricta</i>	Ásia	Marinho	Não	1 e Ran et al. (2019)
<i>Venustaconcha ellipsiformis</i>	América do Norte	Água doce	Não	1, 3 e Renaut et al. (2018)

**Fonte:** elaborada pela autora (2022).

As referências correspondem a 1- MolluscaDatabase, 2- GISD e 3- IUCN.

Como é perceptível pela frequência nas referências, o banco de dados MolluscaBase foi a fonte de informações mais completa, até pela especificidade do mesmo em relação ao filo. Ainda assim, não conta com uma classificação das espécies em relação à invasibilidade, sendo esta determinada pela catalogação das mesmas no banco de dados de invasores globais (GISD) ou pela indicação nos artigos, quando presentes. Dentre as referências utilizadas, a mais deficiente corresponde a lista de espécies ameaçadas, formulada pela União Nacional para a Conservação da Natureza (IUCN). A ausência de moluscos nesta lista seria apreciada, caso realmente refletisse um estado de conservação positivo destes animais. No entanto, não é o caso – de acordo com a listagem da IUCN em 2016, as extinções de moluscos compunham 40% de todas extinções animais, sendo mais ameaçadas as espécies de água doce e terrestres (RÉGNIER; FONTAINE; BOUCHET, 2009; LOPES-LIMA et al., 2018). O fato de que a lista não reflete a situação de conservação desses animais é preocupante por a mesma ter relevância na tomada de decisões que visam a preservação das espécies (RODRIGUES et al., 2006)

Considerando a amplitude de distribuição geográfica de cada espécie uma demonstração de adaptação ambiental, não foi possível detectar uma relação desta com a quantidade de porções codificantes de HSP em seus genomas.

### **5.3 Clusters de proteínas**

A clusterização das sequências proteicas implica a existência de relação entre elas. Assim, tal análise possibilitou a identificação de famílias de HSP entre as sequências detectadas, bem como a inferência de expansão e diferenciação das mesmas entre os genomas. A análise resultou na formação de 559 grupos, dos quais foram considerados para análise apenas os que apresentavam no mínimo três sequências, restando 238 clusters. Destes, foram destacados para discussão os que se enquadram nas categorias abaixo.

### 5.3.1 Clusters com maior quantidade de sequências

Dois clusters concentraram a maior parte das sequências de proteína, dos quais um apresentou 348 sequências de 29 organismos bem distribuídos entre as classes – 15 gastrópodes, 13 bivalves e um cefalópode – e percentual de similaridade entre as sequências baixo, com uma média geral de 63,17%. O grupo mais representativo está representado abaixo (Tabela 5), contando com 455 sequências de 42 moluscos. Neste, estão ausentes apenas as espécies *Octopus vulgaris* e *Lanistes nyassanus*. Foi observado que as sequências destes organismos ocorreram majoritariamente em grupos exclusivos de suas próprias espécies, o que pode indicar um elevado grau de diferenciação das mesmas em relação às demais HSPs do filo.

**Tabela 5** – Cluster de sequências de HSP correspondentes a cada organismo e grau de similaridade médio apresentado entre elas.

Quantidade de sequências	Espécie	Percentual médio de similaridade (%)
3	<i>Achantina immaculata</i>	78,56
9	<i>Anentome helena</i>	80,64
4	<i>Aplysia californica</i>	74,85
10	<i>Architeuthis dux</i>	74,44
28	<i>Argopecten irradians concentricus</i>	77,6
19	<i>Argopecten irradians irradians</i>	77,08
15	<i>Babylonia areolata</i>	83,19
21	<i>Bathymodiolus platifrons</i>	76,39
9	<i>Biomphalaria glabrata</i>	74,09
5	<i>Colubraria reticula</i>	72,93
15	<i>Conus consors</i>	77,03
10	<i>Conus tribblei</i>	82,01

8	<i>Crassostrea gigas</i>	74,64
7	<i>Crassostrea virginica</i>	71,28
6	<i>Cyclina sinensis</i>	75,98
13	<i>Dreissena rostriformis</i>	78,34
22	<i>Elysia chlorotica</i>	76,66
17	<i>Euprymna scolopes</i>	75
4	<i>Haliotis laevigata</i>	77,29
4	<i>Haliotis rubra</i>	77,41
4	<i>Haliotis rufescens</i>	77,65
11	<i>Limacina bulimoides</i>	77,27
3	<i>Limnoperna fortunei</i>	77,33
13	<i>Lottia gigantea</i>	75,58
5	<i>Lutraria rhynchaena</i>	79,95
19	<i>Lymnaea stagnalis</i>	78,65
17	<i>Marisa cornuarietis</i>	94,32
6	<i>Mizuhopecten yessoensis</i>	74,48
18	<i>Modiolus philippinarum</i>	77,59
5	<i>Mytilus galloprovincialis</i>	69,53
2	<i>Octopus bimaculoides</i>	71,91
6	<i>Octopus sinensis</i>	71,1
6	<i>Pecten maximus</i>	73,13
4	<i>Physella acuta</i>	75,3

24	<i>Pinctada imbricata</i>	73,26
5	<i>Pomacea canaliculata</i>	85,12
15	<i>Pomacea maculata</i>	93,62
22	<i>Radix auricularia</i>	76,97
3	<i>Ruditapes philippinarum</i>	78,88
30	<i>Saccostrea glomerata</i>	73,67
2	<i>Sinonovacula constrIcta</i>	70,96
6	<i>Venustaconcha ellipsiformis</i>	78,16

---

**Fonte:** elaborada pela autora (2022).

Por o cluster apresentar proteínas de *Crassostrea virginica* classificadas pelo InterProScan como membros de HSP70 (com código de acesso IPR013126 no InterPro), foi possível inferir que o mesmo é composto por sequências dessa mesma família. Além desta, foram identificadas as famílias HSP10, HS60 e HSP90, em menor quantidade, nos clusters representados abaixo (Tabela 6, 7 e 8).

**Tabela 6** – Cluster de HSP10 detectado nos genomas de moluscos, junto ao grau de similaridade médio apresentado entre suas sequências.

Quantidade de sequências	Espécie	Percentual médio de similaridade (%)
1	<i>Aplysia californica</i>	80,2
1	<i>Biomphalaria glabrata</i>	78,02
6	<i>Colubraria reticula</i>	69,31
8	<i>Crassostrea gigas</i>	71,29
2	<i>Crassostrea virginica</i>	71,29
2	<i>Dreissena rostriformis</i>	72,28
2	<i>Limacina bulimoides</i>	100

15	<i>Octopus bimaculoides</i>	71,29
4	<i>Physella acuta</i>	72,28

---

**Fonte:** elaborada pela autora (2022).

**Tabela 7** – Cluster de HSP60 detectado nos genomas de moluscos, junto ao grau de similaridade médio apresentado entre suas sequências.

<b>Quantidade de sequências</b>	<b>Organismo</b>	<b>Percentual médio de similaridade (%)</b>
1	<i>Aplysia californica</i>	82,84
2	<i>Biomphalaria glabrata</i>	83,1
2	<i>Colubraria reticula</i>	85,61
4	<i>Crassostrea gigas</i>	74,91
2	<i>Crassostrea virginica</i>	74,39
1	<i>Mizuhopecten yessoensis</i>	76,91
1	<i>Octopus bimaculoides</i>	78,82
1	<i>Octopus sinensis</i>	78,82
1	<i>Pecten maximus</i>	76,44
1	<i>Pomacea canaliculata</i>	100

---

**Fonte:** elaborada pela autora (2022).



Nos clusters de HSP10 e HSP60, as sequências indicadas com um percentual de similaridade de 100% foram as “de referência” dos seus respectivos clusters. Assim, esse percentual é absoluto, pois ele se refere a comparação da sequência com ela mesma. O *software* seleciona uma sequência central tanto para estabelecer um parâmetro para avaliar a semelhança entre as demais, quanto para representar o cluster em análises posteriores, permitindo a utilização dessa única sequência para reduzir a redundância dos dados (LI; GODZIK, 2006). Aqui, essa redundância é desejável, pois representa a quantidade de sequências de HSP de cada indivíduo presente em um mesmo cluster.

**Tabela 8** – Cluster de HSP90 detectado nos genomas de moluscos, junto ao grau de similaridade médio apresentado entre suas sequências.

Quantidade de sequências	Organismo	Percentual médio de similaridade (%)
1	<i>Biomphalaria glabrata</i>	73,68
2	<i>Crassostrea gigas</i>	97,91
1	<i>Crassostrea virginica</i>	96,26
2	<i>Mizuhopecten yessoensis</i>	86,23
4	<i>Octopus sinensis</i>	79,73
1	<i>Pecten maximus</i>	86,9
2	<i>Pomacea canaliculata</i>	85,09

**Fonte:** elaborada pela autora (2022).

Ao contrário de HSP70, as informações disponíveis para as demais famílias dessa proteína em moluscos são escassas. Em bivalves, estudos apontam que membros de HSP90 apresentam expressão similar a de HSP70, quando submetidos aos mesmos estressores. Assim, essas famílias podem apresentar uma sobreposição funcional no grupo. As especificidades de sequência e expressão de HSP60 e HSP10 no filo, no entanto, são pouco aprofundadas (FABBRI; VALBONESI; FRANZELLITTI, 2008).

### 5.3.2 Clusters restritos a um gênero

#### 5.3.2.1 *Crassostrea*

A expansão de HSP70 especificamente em *Crassostrea gigas* já foi proposta durante a anotação dessa sequência (ZHANG et al., 2012). Porém, a partir da comparação com as demais montagens disponíveis, a formação dos clusters aqui indica que essa expansão pode ser muito maior e específica, na verdade, ao gênero. Isso se dá por *C. gigas* e *virginica* apresentarem 31 clusters exclusivos, compostos por 251 sequências – das quais 104 são de *C. gigas* e 147 correspondem a *C. virginica*.

Além dos grupos em que ocorrem associações entre proteínas das duas espécies, 53 sequências de *C. gigas* compõem 15 clusters exclusivos da espécie. *C. virginica*, por sua vez, apresenta 65 sequências proteicas agrupadas em 11 clusters específicos.

Estes resultados apontam uma expansão de HSP70 específica no gênero *Crassostrea*, com diferenciação e contínua amplificação da família entre suas espécies. Tal hipótese é reforçada pela ausência de clusters entre sequências de *Saccostrea glomerata* e *Crassostrea*, apesar da proximidade filogenética, sendo ambos membros da família Ostreidae (LIU et al., 2011). *Saccostrea*, por sua vez, apresenta 59 sequências de HSP em 12 clusters específicos da espécie, o que pode apontar, na verdade, a ocorrência de expansão e diferenciação dessas sequências em toda a família.

#### 5.3.2.2 *Argopecten*

Entre os agrupamentos de *Argopecten irradians irradians* e *Argopecten irradians concentricus* é perceptível uma relação semelhante à que ocorre entre as sequências de *Crassostrea* – apresenta clusters restritos ao gênero e específicos de cada espécie.

Todavia, são concentrações bem menores se comparadas às de *Crassostrea*. Foram 14 sequências de proteína, distribuídas entre dois clusters compostos por ambas espécies e apenas quatro restritas a *A. irradians concentricus*. Já *A. irradians irradians* concentrou 16 sequências em três clusters específicos.

Outros membros da família Pectinidae (FENG et al, 2011), por outro lado, apresentam uma expansão mais expressiva – *Mizuhopecten yessoensis* e *Pecten maximus* possuem 28 clusters limitados às duas espécies, com o total de 159 sequências, das quais 51 são proteínas de *Pecten* e 108 de *Mizuhopecten*. Além desses agrupamentos, ambos apresentam indício de expansão específica de HSP em cada espécie, como ocorre também em *Crassostrea*. Aqui, *Mizuhopecten* apresentou oito clusters exclusivos, de 31 sequências, e *Pecten*, quatro clusters próprios com 17 sequências, no total.

Essas evidências sugerem a ocorrência de expansões independentes na família Pectinidae, visto que sequências dos três gêneros – *Mizuhopecten*, *Pecten* e *Argopecten* – não ocorreram em um cluster comum.

### 5.3.2.3 *Octopus*

As sequências de HSP de cefalópodes, no geral, apresentaram um padrão de clusterização exclusivo à classe. Dentre os *Octopus*, as sequências das três espécies – *O. bimaculoides*, *sinensis* e *vulgaris* – apresentaram 9 grupos comuns entre pelo menos duas delas. Destes, *O. vulgaris* compunha apenas um junto a *O. sinensis*, formado por três sequências e um percentual de similaridade entre as HSPs das espécies de 60,4%. Conforme foi apontado anteriormente, as sequências de *O. vulgaris* aparentam grande diferenciação – não só em relação às do filo, como também entre as de seu próprio gênero.

Assim, foram mais significativos os agrupamentos entre HSPs de *O. bimaculoides* e *sinensis*. Entre estas a relação é íntima, com percentuais de similaridade entre as sequências chegando a 99,88%. Outra evidência dessa relação é o fato de que somente um cluster não foi composto por ambas as espécies – contando com três sequências exclusivas de *O. sinensis*. Nos oito clusters compartilhados por elas, as espécies apresentaram o total de 33 proteínas – das quais 20 são de *O. sinensis* e 13 de *O. bimaculoides*.

### 5.3.2.4 *Pomacea, Conus e Haliotis*

O estudo englobou ainda duas espécies correspondentes a *Pomacea*, duas de *Conus* e três referentes a *Haliotis*. Contudo, nesses gêneros não foram identificados clusters compartilhados entre suas respectivas espécies. Foi observada, na verdade, uma separação bem definida nos grupos de proteínas de cada uma delas.

As espécies *H. rubra*, *H. laevigata* e *H. rufescens* apresentaram apenas dois clusters com mais de um representante do gênero. Estes foram compostos por *H. laevigata* e *rubra* e as sequências apresentaram, considerando a proximidade filogenética, um percentual de similaridade baixo, de 68,67%. No entanto, todas as espécies apresentaram clusters independentes, sendo o de *H. rubra* o mais numeroso – com 12 sequências.

Esta mesma distinção é identificada nos clusters de *Pomacea* e *Conus* que, por sua vez, não apresentaram nenhum cluster particular de suas espécies. Esse padrão de clusterização propõe uma diferenciação específica das sequências de HSP de cada espécie, por as mesmas ocorrerem majoritariamente em clusters restritos a elas e apresentarem um baixo grau de similaridade com as demais sequências do gênero.

### 5.3.3 Clusters restritos a espécie

Foram destacadas na Tabela 9 as espécies que apresentaram agrupamentos de sequências de HSP particulares, a fim de identificar outras expansões restritas a elas.

**Tabela 9** – Quantidade de clusters específicos por espécie e de sequências de HSP neles agrupadas.

<b>Espécie</b>	<b>Quantidade de clusters</b>	<b>Quantidade de sequências agrupadas</b>
<i>Achantina immaculata</i>	3	14
<i>Aplysia californica</i>	3	10
<i>Architeuthis dux</i>	2	19
<i>Argopecten irradians concentricus</i>	1	4
<i>Argopecten irradians irradians</i>	3	16
<i>Babylonia areolata</i>	1	5
<i>Bathymodiolus platifrons</i>	3	17

<i>Biomphalaria glabrata</i>	4	14
<i>Conus consors</i>	2	10
<i>Conus tribblei</i>	1	5
<i>Crassostrea gigas</i>	15	53
<i>Crassostrea virginica</i>	11	65
<i>Cyclina sinensis</i>	3	19
<i>Dreissena rostriformis</i>	6	29
<i>Elysia chlorotica</i>	1	5
<i>Euprymna scolopes</i>	2	10
<i>Haliotis laevigata</i>	1	5
<i>Haliotis rubra</i>	2	12
<i>Haliotis rufescens</i>	1	5
<i>Lanistes nyassanus</i>	2	12
<i>Limacina bulimoides</i>	1	3
<i>Limnoperna fortunei</i>	3	13
<i>Lottia gigantea</i>	1	5
<i>Lutraria rhynchaena</i>	2	12
<i>Lymnaea stagnalis</i>	1	5
<i>Marisa cornuarietis</i>	1	5

<i>Mizuhopecten yessoensis</i>	8	31
<i>Modiolus philippinarum</i>	1	3
<i>Mytilus galloprovincialis</i>	4	19
<i>Octopus sinensis</i>	1	3
<i>Octopus vulgaris</i>	2	9
<i>Pecten maximus</i>	4	17
<i>Physella acuta</i>	1	6
<i>Pinctada imbricata</i>	4	15
<i>Pomacea canaliculata</i>	1	6
<i>Pomacea maculata</i>	1	5
<i>Radix auricularia</i>	3	14
<i>Ruditapes philippinarum</i>	2	10
<i>Saccostrea glomerata</i>	12	59
<i>Sinonovacula constricta</i>	6	40
<i>Venustaconcha ellipsiformis</i>	1	5

---

**Fonte:** elaborada pela autora (2022).

Além dos clusters já destacados, apresentaram concentrações específicas de HSP em seus genomas as espécies *Sinonovacula constricta* e *Dreissena rostriformis*. A última é classificada como uma espécie invasora global, enquanto *Sinonovacula* tem distribuição restrita à Ásia.

## 6 CONCLUSÃO

O presente trabalho comparou a proporção de sequências de HSP em genomas de moluscos e a capacidade de adaptação ambiental dos mesmos, observando que não existe uma relação identificável entre essas variáveis. Tal conclusão foi alcançada ao analisar tanto a quantidade de proteínas nos genomas e quanto as possíveis relações entre elas, quando agrupadas em clusters relacionados. A quantidade de HSPs por organismo não se mostrou proporcional aos seus traços de adaptação ambiental, bem como esses indivíduos mais “adaptados” não apresentaram indício de expansão em suas sequências proteicas – mesmo que filogeneticamente próximos, como foi o caso das *Pomaceas*. A busca por clusters expandidos de forma específica nas espécies tampouco indicou expansão de HSP em indivíduos caracterizados como “invasores”. A ausência dessa relação sugere, portanto, que tais “invasões” têm viés oportunista. Ou seja, não estaria ligada à capacidade prévia dos organismos, mas se dá por meio de introduções dos mesmos em habitats variados.

Ainda assim, foram preditas 1.586 sequências de proteínas em montagens de genomas de moluscos não anotados. Estas, junto às HSPs anotadas disponíveis, foram agrupadas para a identificação de famílias. Nessa abordagem foram identificados agrupamentos de HSP10, HSP60, HSP 90 e HSP70. Esta última se mostrou expandida entre todos os moluscos, formando um grande cluster entre eles. Em bivalves, a expansão de HSPs se mostrou ainda mais expressiva – com a formação de clusters nas famílias Ostreidae e Pectinidae, mais especificamente nos gêneros *Crassostrea*, *Saccostrea*, *Argopecten*, *Mizuhopecten* e *Pecten*. Aqui, são necessários mais estudos para determinar se a expansão ocorreu de forma independente nos gêneros ou se é, de fato, comum à família, com subsequente diferenciação nas sequências de HSP em cada gênero. Em Cephalopoda essas proteínas também aparentam ter passado por grande diferenciação, com as mesmas estando majoritariamente restritas à suas classes e apresentando grande similaridade entre si.

São sugeridos, ainda, estudos para incrementar a predição dessas proteínas, empregando metodologias de busca *ab initio* ou utilizando mais de um genoma como referência para busca por homologia. Isto se dá pois, como a maior parte das proteínas foram identificadas por similaridade, sequências mais diferenciadas podem não ter sido detectadas, bem como organismos mais filogeneticamente próximos podem ter apresentado uma maior identificação.

Ademais, o conceito de capacidade de adaptação ambiental foi baseado em características de amplitude de distribuição geográfica e na caracterização prévia das espécies

como invasivas ou não. Estas informações podem ser influenciadas pela insuficiência e/ou desatualização dos bancos de dados utilizados na busca, o que destaca mais uma lacuna de conhecimento em relação ao filo. Além da necessidade de suprir tais lacunas, para abranger melhor a heterogeneidade das classes podem ser considerados os diferentes modos de vida e hábitos alimentares das espécies.

Considerando essa diversidade, é possível que existam relações identificáveis que expliquem a capacidade de adaptação ambiental destes organismos, mas como as estratégias de adaptação são múltiplas, são necessárias averiguações na dinâmica de outras proteínas relacionadas a essa capacidade. Para isso, é frisada a necessidade de mais genomas de moluscos sequenciados e anotados, bem como a melhora na qualidade de montagens disponíveis. Isto pois, em sequências incompletas, como as de *Corbicula fluminea* e *Dreissena polymorpha*, a identificação de sequências proteicas é inviável.



## REFERÊNCIAS

ADEMA, C. M.; HILLIER, L. W.; JONES, C. S.; et al. Corrigendum: Whole genome analysis of a schistosomiasis-transmitting freshwater snail. **Nature communications**, v. 8, p. 16153, 2017.

ALBERTIN, C. B.; SIMAKOV, O.; MITROS, T.; et al. The octopus genome and the evolution of cephalopod neural and morphological novelties. **Nature**, v. 524, n. 7564, p. 220–224, 2015.

ALBERTS, B.; JOHNSON, A.; LEWIS, J.; et al. **Molecular Biology of the Cell 4th Edition: International Student Edition**. Garland Science, 2002.

**Animal biodiversity: An outline of higher-level classification and survey of taxonomic richness**. Magnolia Press, 2011.

AVERY, O. T.; MACLEOD, C. M.; MCCARTY, M. Studies on the chemical nature of the substance inducing transformation of pneumococcal types. Induction of transformation by a desoxyribonucleic acid fraction isolated from *Pneumococcus* type III. 1944. **Molecular medicine**, v. 1, n. 4, p. 344–365, 1995.

BARONIO, C. A.; BOTTON, M.; GOMES, S. R.; ROBINSON, D. G. First record of qualitative losses caused by *Meghimatium pictum* in vineyards of Southern Brazil and the effects of two molluscicides for its control. **Ciência Rural**, 2014. Disponível em: <<http://dx.doi.org/10.1590/0103-8478cr20130522>>.

BAYNE, C. J. Successful parasitism of vector snail *Biomphalaria glabrata* by the human blood fluke (trematode) *Schistosoma mansoni*: a 2009 assessment. **Molecular and biochemical parasitology**, v. 165, n. 1, p. 8–18, 2009.

BELCAID, M.; CASABURI, G.; MCANULTY, S. J.; et al. Symbiotic organs shaped by distinct modes of genome evolution in cephalopods. **Proceedings of the National Academy of Sciences of the United States of America**, v. 116, n. 8, p. 3030–3035, 2019.

BOLTOVSKOY, D.; XU, M.; NAKANO, D. Impacts of *Limnoperna Fortunei* on Man-Made Structures and Control Strategies: General Overview. **Limnoperna Fortunei**, 2015. Disponível em: <[http://dx.doi.org/10.1007/978-3-319-13494-9\\_21](http://dx.doi.org/10.1007/978-3-319-13494-9_21)>.

BOTWRIGHT, N. A.; ZHAO, M.; WANG, T.; et al. Greenlip Abalone (*Haliotis*) Genome and Protein Analysis Provides Insights into Maturation and Spawning. **G3**, v. 9, n. 10, p. 3067–3078, 2019.

BURLAKOVA, L. E.; KARATAYEV, A. Y.; KARATAYEV, V. A. Invasive mussels induce community changes by increasing habitat complexity. **Hydrobiologia**, 2012. Disponível em: <<http://dx.doi.org/10.1007/s10750-011-0791-4>>.

CAI, H.; LI, Q.; FANG, X.; et al. A draft genome assembly of the solar-powered sea slug *Elysia chlorotica*. **Scientific data**, v. 6, p. 190022, 2019.

CALCINO, A. D.; DE OLIVEIRA, A. L.; SIMAKOV, O.; et al. The quagga mussel genome and the evolution of freshwater tolerance. **DNA research: an international journal for rapid publication of reports on genes and genomes**, v. 26, n. 5, p. 411–422, 2019.

COLLEY, E.; SIMONE, L. R. L.; DE LOYOLA E SILVA, J. Uma viagem pela história da Malacologia. **Estudos de Biologia**, 2012. Disponível em: <<http://dx.doi.org/10.7213/estud.biol.7331>>.

CONNELLY, N. A.; O'NEILL, C. R., Jr; KNUTH, B. A.; BROWN, T. L. Economic impacts of zebra mussels on drinking water treatment and electric power generation facilities. **Environmental management**, v. 40, n. 1, p. 105–112, 2007.

CRICK, F. H. On protein synthesis. **Symposia of the Society for Experimental Biology**, v. 12, p. 138–163, 1958.

DAUGAARD, M.; ROHDE, M.; JÄÄTTELÄ, M. The heat shock protein 70 family: Highly homologous proteins with overlapping and distinct functions. **FEBS letters**, v. 581, n. 19, p. 3702–3710, 2007.

DAWKINS, R. **River Out of Eden: A Darwinian View of Life**. Basic Books, 2008.

VAN DIJK, E. L.; AUGER, H.; JASZCZYSZYN, Y.; THERMES, C. Ten years of next-generation sequencing technology. **Trends in Genetics**, 2014. Disponível em: <<http://dx.doi.org/10.1016/j.tig.2014.07.001>>.

DUNN, C. W.; MUNRO, C. Comparative genomics and the diversity of life. **Zoologica Scripta**, 2016. Disponível em: <<http://dx.doi.org/10.1111/zsc.12211>>.

FABBRI, E.; VALBONESI, P.; FRANZELLITTI, S. HSP expression in bivalves. **Invertebrate Survival Journal**, v. 5, n. 2, p. 135-161, 2008.

FEDER, M. E.; HOFMANN, G. E. Heat-shock proteins, molecular chaperones, and the stress response: evolutionary and ecological physiology. **Annual review of physiology**, v. 61, p. 243–282, 1999.

FENG, Y.; LI, Q.; KONG, L.; ZHENG, X. DNA barcoding and phylogenetic analysis of Pectinidae (Mollusca: Bivalvia) based on mitochondrial COI and 16S rRNA genes. **Molecular biology reports**, v. 38, n. 1, p. 291–299, 2011.

FELDMEYER, Barbara et al. Positive selection in development and growth rate regulation genes involved in species divergence of the genus *Radix*. **BMC evolutionary biology**, v. 15, n. 1, p. 1-11, 2015.

DA FONSECA, R. R.; COUTO, A.; MACHADO, A. M.; et al. A draft genome sequence of the elusive giant squid, *Architeuthis dux*. **GigaScience**, v. 9, n. 1, 2020. Disponível em: <<http://dx.doi.org/10.1093/gigascience/giz152>>.

FU, L.; NIU, B.; ZHU, Z.; WU, S.; LI, W. CD-HIT: accelerated for clustering the next-generation sequencing data. **Bioinformatics**, v. 28, n. 23, p. 3150–3152, 2012.

GARRIDO, C.; PAUL, C.; SEIGNEURIC, R.; KAMPINGA, H. H. The small heat shock proteins family: The long forgotten chaperones. **The International Journal of Biochemistry & Cell Biology**, 2012. Disponível em: <<http://dx.doi.org/10.1016/j.biocel.2012.02.022>>.

GAUTHIER, J.; VINCENT, A. T.; CHARETTE, S. J.; DEROME, N. A brief history of bioinformatics. **Briefings in bioinformatics**, v. 20, n. 6, p. 1981–1996, 2019.

GIBBS, R. A. The Human Genome Project changed everything. **Nature reviews. Genetics**, v. 21, n. 10, p. 575–576, 2020.

GIGA COMMUNITY OF SCIENTISTS; BRACKEN-GRISSOM, H.; COLLINS, A. G.; et al. The Global Invertebrate Genomics Alliance (GIGA): developing community resources to study diverse invertebrate genomes. **The Journal of heredity**, v. 105, n. 1, p. 1–18, 2014.

GRANTHAM, R.; GAUTIER, C.; GOUY, M.; MERCIER, R.; PAVÉ, A. Codon catalog usage and the genome hypothesis. **Nucleic acids research**, v. 8, n. 1, p. r49–r62, 1980.

HASZPRUNAR, G.; WANNINGER, A. Molluscs. **Current biology: CB**, v. 22, n. 13, p. R510–4, 2012.

HOFMANN, G. E.; TODGHAM, A. E. Living in the now: physiological mechanisms to tolerate a rapidly changing environment. **Annual review of physiology**, v. 72, p. 127–145, 2010.

HULME, P. E. Trade, transport and trouble: managing invasive species pathways in an era of globalization. **Journal of Applied Ecology**, 2009. Disponível em: <<http://dx.doi.org/10.1111/j.1365-2664.2008.01600.x>>.

HUNKAPILLER, T.; KAISER, R. J.; KOOP, B. F.; HOOD, L. Large-scale and automated DNA sequence determination. **Science**, v. 254, n. 5028, p. 59–67, 1991.

JEE, H. Size dependent classification of heat shock proteins: a mini-review. **Journal of exercise rehabilitation**, v. 12, n. 4, p. 255–259, 2016.

JONES, P.; BINNS, D.; CHANG, H.-Y.; et al. InterProScan 5: genome-scale protein function classification. **Bioinformatics**, v. 30, n. 9, p. 1236–1240, 2014.

KANO, Y.; NEUSSER, T. P.; FUKUMORI, H.; JÖRGER, K. M.; SCHRÖDL, M. Sea-slug invasion of the land. **Biological Journal of the Linnean Society**, 2015. Disponível em: <<http://dx.doi.org/10.1111/bij.12578>>.

KARATAYEV, A. Y.; PADILLA, D. K.; MINCHIN, D.; BOLTOVSKOY, D.; BURLAKOVA, L. E. Changes in Global Economies and Trade: the Potential Spread of Exotic Freshwater Bivalves. **Biological Invasions**, 2007. Disponível em: <<http://dx.doi.org/10.1007/s10530-006-9013-9>>.

KENNY, N. J.; MCCARTHY, S. A.; DUDCHENKO, O.; et al. The gene-rich genome of the scallop *Pecten maximus*. **GigaScience**, v. 9, n. 5, 2020. Disponível em: <<http://dx.doi.org/10.1093/gigascience/giaa037>>.

- KNIGHT, T. M.; ASHMAN, T.-L.; BENNETT, J. M.; et al. Reflections on, and visions for, the changing field of pollination ecology. **Ecology letters**, v. 21, n. 8, p. 1282–1295, 2018.
- KNOP, E.; REUSSER, N. Jack-of-all-trades: phenotypic plasticity facilitates the invasion of an alien slug species. **Proceedings. Biological sciences / The Royal Society**, v. 279, n. 1747, p. 4668–4676, 2012.
- KOZŁOWSKI, J. On the distribution and food preferences of *Arion subfuscus* (Draparnaud, 1805). **Folia Malacologica**, 2009. Disponível em: <<http://dx.doi.org/10.12657/folmal.016.010>>.
- LEWIN, H. A.; ROBINSON, G. E.; KRESS, W. J.; et al. Earth BioGenome Project: Sequencing life for the future of life. **Proceedings of the National Academy of Sciences of the United States of America**, v. 115, n. 17, p. 4325–4333, 2018.
- LINDQUIST, S.; CRAIG, E. A. THE HEAT-SHOCK PROTEINS. **Annual Review of Genetics**, 1988. Disponível em: <<http://dx.doi.org/10.1146/annurev.ge.22.120188.003215>>.
- LIU, C.; REN, Y.; LI, Z.; et al. Giant African snail genomes provide insights into molluscan whole-genome duplication and aquatic-terrestrial transition. **Molecular ecology resources**, v. 21, n. 2, p. 478–494, 2021.
- LIU, J.; LI, Q.; KONG, L.; YU, H.; ZHENG, X. Identifying the true oysters (*Bivalvia*: *Ostreidae*) with mitochondrial phylogeny and distance-based DNA barcoding. **Molecular ecology resources**, v. 11, n. 5, p. 820–830, 2011.
- LI, W.; GODZIK, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. **Bioinformatics**, 2006. Disponível em: <<http://dx.doi.org/10.1093/bioinformatics/btl158>>.
- LOPES-LIMA, M.; BURLAKOVA, L. E.; KARATAYEV, A. Y.; et al. Conservation of freshwater bivalves at the global scale: diversity, threats and research needs. **Hydrobiologia**, 2018. Disponível em: <<http://dx.doi.org/10.1007/s10750-017-3486-7>>.
- LOPEZ, J. V.; KAMEL, B.; MEDINA, M.; COLLINS, T.; BAUMS, I. B. Multiple Facets of Marine Invertebrate Conservation Genomics. **Annual review of animal biosciences**, v. 7, p. 473–497, 2019.
- LYDEARD, C.; COWIE, R. H.; PONDER, W. F.; et al. The Global Decline of Nonmarine Mollusks. **BioScience**, 2004. Disponível em: <[http://dx.doi.org/10.1641/0006-3568\(2004\)054\[0321:tgdonm\]2.0.co;2](http://dx.doi.org/10.1641/0006-3568(2004)054[0321:tgdonm]2.0.co;2)>.
- MADSEN, H.; FRANDBSEN, F. The spread of freshwater snails including those of medical and veterinary importance. **Acta tropica**, v. 46, n. 3, p. 139–146, 1989.
- MAHULU, A.; CLEWING, C.; STELBRINK, B.; et al. Cryptic intermediate snail host of the liver fluke *Fasciola hepatica* in Africa. **Parasites & vectors**, v. 12, n. 1, p. 573, 2019.
- MASONBRINK, R. E.; PURCELL, C. M.; BOLES, S. E.; et al. An Annotated Genome for *Haliotis rufescens* (Red Abalone) and Resequenced Green, Pink, Pinto, Black, and White Abalone Species. **Genome biology and evolution**, v. 11, n. 2, p. 431–438, 2019.

MCCARTNEY, M. A.; AUCH, B.; KONO, T.; et al. The Genome of the Zebra Mussel, *Dreissena polymorpha*: A Resource for Invasive Species Research. . Disponível em: <<http://dx.doi.org/10.1101/696732>>.

MORAES, F.; GÓES, A. A decade of human genome project conclusion: Scientific diffusion about our genome knowledge. **Biochemistry and molecular biology education: a bimonthly publication of the International Union of Biochemistry and Molecular Biology**, v. 44, n. 3, p. 215–223, 2016.

MULDER, N.; APWEILER, R. InterPro and InterProScan: tools for protein sequence classification and comparison. **Methods in molecular biology** , v. 396, p. 59–70, 2007.

MUN, S.; KIM, Y.-J.; MARKKANDAN, K.; et al. The Whole-Genome and Transcriptome of the Manila Clam (*Ruditapes philippinarum*). **Genome biology and evolution**, v. 9, n. 6, p. 1487–1498, 2017.

NUNES, A.; TRICARICO, E.; PANOV, V.; CARDOSO, A.; KATSANEVAKIS, S. Pathways and gateways of freshwater invasions in Europe. **Aquatic Invasions**, 2015. Disponível em: <<http://dx.doi.org/10.3391/ai.2015.10.4.01>>.

O'LEARY, N. A.; WRIGHT, M. W.; BRISTER, J. R.; et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. **Nucleic acids research**, v. 44, n. D1, p. D733–45, 2016.

PONDER, W. F.; LINDBERG, D. R.; PONDER, J. M. **Biology and Evolution of the Mollusca, Volume 2**. CRC Press, 2020.

PORTIN, P. The birth and development of the DNA theory of inheritance: sixty years since the discovery of the structure of DNA. **Journal of genetics**, v. 93, n. 1, p. 293–302, 2014.

POWELL, D.; SUBRAMANIAN, S.; SUWANSA-ARD, S.; et al. The genome of the oyster *Saccostrea* offers insight into the environmental resilience of bivalves. **DNA research: an international journal for rapid publication of reports on genes and genomes**, v. 25, n. 6, p. 655–665, 2018.

PRATHER, C. M.; PELINI, S. L.; LAWS, A.; et al. Invertebrates, ecosystem services and climate change. **Biological reviews of the Cambridge Philosophical Society**, v. 88, n. 2, p. 327–348, 2013.

RAN, Z.; LI, Z.; YAN, X.; et al. Chromosome-level genome assembly of the razor clam *Sinonovacula constricta* (Lamarck, 1818). **Molecular ecology resources**, v. 19, n. 6, p. 1647–1658, 2019.

RÉGNIER, C.; FONTAINE, B.; BOUCHET, P. Not Knowing, Not Recording, Not Listing: Numerous Unnoticed Mollusk Extinctions. **Conservation Biology**, 2009. Disponível em: <<http://dx.doi.org/10.1111/j.1523-1739.2009.01245.x>>.

RENAUT, S.; GUERRA, D.; HOEH, W. R.; et al. Genome Survey of the Freshwater Mussel *Venustaconcha ellipsiformis* (Bivalvia: Unionida) Using a Hybrid De Novo Assembly Approach. **Genome biology and evolution**, v. 10, n. 7, p. 1637–1646, 2018.

RICCIARDI, A. Are modern biological invasions an unprecedented form of global change? **Conservation biology: the journal of the Society for Conservation Biology**, v. 21, n. 2, p. 329–336, 2007.

RODRIGUES, A.; PILGRIM, J.; LAMOREUX, J.; HOFFMANN, M.; BROOKS, T. The value of the IUCN Red List for conservation. **Trends in Ecology & Evolution**, 2006. Disponível em: <<http://dx.doi.org/10.1016/j.tree.2005.10.010>>.

ROMERO, P. E.; PFENNINGER, M.; KANO, Y.; KLUSSMANN-KOLB, A. Molecular phylogeny of the Ellobiidae (Gastropoda: Panpulmonata) supports independent terrestrial invasions. **Molecular phylogenetics and evolution**, v. 97, p. 43–54, 2016.

SANGER, F.; NICKLEN, S.; COULSON, A. R. DNA sequencing with chain-terminating inhibitors. **Proceedings of the National Academy of Sciences of the United States of America**, v. 74, n. 12, p. 5463–5467, 1977.

SANTANA, F. D.; BACCARO, F. B.; COSTA, F. R. C. Busy Nights: High Seed Dispersal by Crickets in a Neotropical Forest. **The American Naturalist**, 2016. Disponível em: <<http://dx.doi.org/10.1086/688676>>.

SAYERS, E. W.; BOLTON, E. E.; BRISTER, J. R.; et al. Database resources of the national center for biotechnology information. **Nucleic acids research**, v. 50, n. D1, p. D20–D26, 2022.

SCHMIDT-NIELSEN, K.; TAYLOR, C. R.; SHKOLNIK, A. Desert snails: problems of heat, water and food. **The Journal of experimental biology**, v. 55, n. 2, p. 385–398, 1971.

SEARLS, D. B. The roots of bioinformatics. **PLoS computational biology**, v. 6, n. 6, p. e1000809, 2010.

SIGWART, J. D. Zoology: Molluscs All Beneath the Sun, One Shell, Two Shells, More, or None. **Current biology: CB**, 24. jul. 2017.

SIMBERLOFF, D.; MARTIN, J.-L.; GENOVESI, P.; et al. Impacts of biological invasions: what's what and the way forward. **Trends in ecology & evolution**, v. 28, n. 1, p. 58–66, 2013.

SIN, T. S. Damage potential of the golden apple snail *Pomacea canaliculata* (Lamarck) in irrigated rice and its control by cultural approaches. **International Journal of Pest Management**, 2003. Disponível em: <<http://dx.doi.org/10.1080/713867835>>.

SLATER, G. S. C.; BIRNEY, E. Automated generation of heuristics for biological sequence comparison. **BMC bioinformatics**, v. 6, p. 31, 2005.

SMITH, S. A.; WILSON, N. G.; GOETZ, F. E.; et al. Resolving the evolutionary relationships of molluscs with phylogenomic tools. **Nature**, v. 480, n. 7377, p. 364–367, 2011.

SØRENSEN, J. G.; KRISTENSEN, T. N.; LOESCHCKE, V. The evolutionary and ecological role of heat shock proteins. **Ecology Letters**, 2003. Disponível em: <<http://dx.doi.org/10.1046/j.1461-0248.2003.00528.x>>.

SOUSA, R.; GUTIÉRREZ, J. L.; ALDRIDGE, D. C. Non-indigenous invasive bivalves as ecosystem engineers. **Biological Invasions**, 2009. Disponível em: <<http://dx.doi.org/10.1007/s10530-009-9422-7>>.

STEVENS, H. **Life Out of Sequence: A Data-driven History of Bioinformatics**. 2013.

SUN, J.; MU, H.; IP, J. C. H.; et al. Signatures of Divergence, Invasiveness, and Terrestrialization Revealed by Four Apple Snail Genomes. **Molecular biology and evolution**, v. 36, n. 7, p. 1507–1520, 2019.

SUN, J.; ZHANG, Y.; XU, T.; et al. Adaptation to deep-sea chemosynthetic environments as revealed by mussel genomes. **Nature ecology & evolution**, v. 1, n. 5, p. 121, 2017.

TAKEUCHI, T.; KOYANAGI, R.; GYOJA, F.; et al. Bivalve-specific gene expansion in the pearl oyster genome: implications of adaptation to a sessile lifestyle. **Zoological letters**, v. 2, p. 3, 2016.

THAI, B. T.; LEE, Y. P.; GAN, H. M.; et al. Whole Genome Assembly of the Snout Otter Clam, *Lutraria rhynchaena*, Using Nanopore and Illumina Data, Benchmarked Against Bivalve Genome Assemblies. **Frontiers in Genetics**, 2019. Disponível em: <<http://dx.doi.org/10.3389/fgene.2019.01158>>.

THIENGO, S. C.; MALDONADO, A.; MOTA, E. M.; et al. The giant African snail *Achatina fulica* as natural intermediate host of *Angiostrongylus cantonensis* in Pernambuco, northeast Brazil. **Acta tropica**, v. 115, n. 3, p. 194–199, 2010.

TOMANEK, L.; HELMUTH, B. Physiological ecology of rocky intertidal organisms: a synergy of concepts. **Integrative and comparative biology**, v. 42, n. 4, p. 771–775, 2002.

ULIANO-SILVA, M.; AMÉRICO, J. A.; BRINDEIRO, R.; et al. The Genetics of the Golden Mussel (*Limnoperna fortunei*): Are Genes Related to Invasiveness? **Limnoperna Fortunei**, 2015. Disponível em: <[http://dx.doi.org/10.1007/978-3-319-13494-9\\_4](http://dx.doi.org/10.1007/978-3-319-13494-9_4)>.

ULIANO-SILVA, M.; DONDERO, F.; DAN OTTO, T.; et al. A hybrid-hierarchical genome assembly strategy to sequence the invasive golden mussel, *Limnoperna fortunei*. **GigaScience**, v. 7, n. 2, 2018. Disponível em: <<http://dx.doi.org/10.1093/gigascience/gix128>>.

VINTHER, J. The origins of molluscs. **Palaeontology**, 2015. Disponível em: <<http://dx.doi.org/10.1111/pala.12140>>.

VOS, M. J.; HAGEMAN, J.; CARRA, S.; KAMPINGA, H. H. Structural and functional diversities between members of the human HSPB, HSPH, HSPA, and DNAJ chaperone families. **Biochemistry**, v. 47, n. 27, p. 7001–7011, 2008.

WANNINGER, A.; WOLLESEN, T. The evolution of molluscs. **Biological Reviews**, 2019. Disponível em: <<http://dx.doi.org/10.1111/brv.12439>>.

YANAI, Z.; DAYAN, T.; MIENIS, H.; GASITH, A. The pet and horticultural trades as introduction and dispersal agents of non-indigenous freshwater molluscs. **Management of Biological Invasions**, 2017. Disponível em: <<http://dx.doi.org/10.3391/mbi.2017.8.4.07>>.

ZARRELLA, I.; HERTEN, K.; MAES, G. E.; et al. The survey and reference assisted assembly of the *Octopus vulgaris* genome. **Scientific data**, v. 6, n. 1, p. 13, 2019.

ZHANG, G.; FANG, X.; GUO, X.; et al. The oyster genome reveals stress adaptation and complexity of shell formation. **Nature**, v. 490, n. 7418, p. 49–54, 2012.

ZHANG, Z.-Q. Animal biodiversity: An outline of higher-level classification and survey of taxonomic richness (Addenda 2013). **Zootaxa**, v. 3703, p. 1–82, 2013.