U.PORTO

FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

Margaritiferidae: from “pearls” to genomes

André Manuel Gomes dos Santos

Doutoramento em Biologia

Departamento de Biologia, Faculdade de Ciências,
Universidade do Porto
2023

Margaritiferidae: from “pearls” to genomes

André Manuel Gomes dos Santos

Doutoramento em Biologia

Departamento de Biologia, Faculdade de Ciências,
Universidade do Porto
2023

Orientadora
Doutora Elsa Froufe, Investigadora Principal,
Centro Interdisciplinar de Investigação Marinha e Ambiental (CIIMAR)

Coorientador
Doutor Luís Filipe Costa Castro, Professor Auxiliar, Faculdade de Ciências, Universidade do Porto
Centro Interdisciplinar de Investigação Marinha e Ambiental (CIIMAR)

This thesis was supported by the Portuguese Foundation for Science and Technology (FCT) under PhD grants ref: SFRH/BD/137935/2018 and ref: COVID/DB/152933/2022 and developed under the projects: ConBiomics: the missing approach for the Conservation of freshwater Bivalves (Project No:NORTE-01-0145-FEDER-030286), co-financed by COMPETE 2020, Portugal 2020 and the European Union through the ERDF, and by FCT through national funds; and EdgeOmics - Freshwater Bivalves at the edge: Adaptation genomics under climate-change scenarios (PTDC/CTA-AMB/3065/2020) funded by FCT through national funds.

FCT
Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

GOVERNO DA REPÚBLICA
PORTUGUESA

EDGEOmics
PTDC/CTA-AMB/3065/2020

ConBiomics
Conservation and Phylogenomics of Freshwater Bivalves

NORTE2020
PROGRAMA OPERACIONAL REGIONAL DO NORTE

PORTUGAL
2020

UNIÃO EUROPEIA
Fundos Europeus
Estruturais e de Investimento

U.PORTO
FC FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

ciimar
Centro Interdisciplinar
de Investigação
Marinha e Ambiental

Dedicada aos meus pais, Helena e Domingos

AGRADECIMENTOS

Em primeiro lugar quero agradecer aos meus orientadores, por terem sido os principais contribuidores para a realização de todos os trabalhos aqui apresentados.

Quero deixar um agradecimento especial para a Elsa, e por isso vou tentar, o melhor que consigo, expressar o quão agradecido estou por tudo. Foi graças ao incentivo da Elsa que me aventurei no Doutoramento. Hoje vejo o erro que teria sido da minha parte não o ter feito, e ficarei eternamente grato por isso. Agradeço por me introduzir no mundo fantástico da genética e todo o conhecimento científico que pacientemente me passou (e continua a passar). Sei que nem sempre foi fácil lidar com a minha “calma”. Obrigado por me ter acolhido no AEE, pela liberdade de explorar diferentes áreas e, acima de tudo, por me fazer sentir parte da equipa. A somar a tudo isto, nas nossas conivências diárias, aprendi muito para além da ciência e por isso, a Elsa será sempre uma inspiração, não só a nível profissional, mas também a nível pessoal.

Quero também agradecer ao meu coorientador Filipe Castro por todo o conhecimento que me passou, pelo apoio e pragmatismo durante estes anos. Agradeço especialmente por me ensinar a pensar como um evolucionista e apreciar a biodiversidade de uma maneira completamente diferente. Obrigado por me acolher no AGE e por me deixar contribuir para os trabalhos e discussões do grupo.

Apesar das circunstâncias não terem permitido um trabalho mais próximo com os meus coorientadores externos, o Miguel e a Sophie, queria deixar um agradecimento por terem aceitado fazer parte da equipa de orientação. Deixo um agradecimento especial para o Miguel, por ter guiado os meu primeiros passos em Shell scripting, por toda a ajuda na fase inicial e por toda a compreensão e carinho durante todo o percurso.

Por fim, queria agradecer ao Manel, o meu coorientador não oficial, por me passar a paixão pelo estudo de mexilhões de água doce, por todo o conhecimento que partilhou comigo, por abrir as portas às várias colaborações que fiz e, especialmente, por me ter acolhido como membro da “Mussel Team”.

Quero também agradecer a todas as pessoas com quem colaborei nos diferentes trabalhos, especialmente aos membros da “Mussel Team”.

Agradeço aos meus parceiros do AEE e AGE, Giulia, David, Duarte, André, Rui, Manu, Elza, Raquel, Miguel e Marcos, por toda a partilha de conhecimento e por tornarem todos os dias no CIIMAR dias especiais.

Deixo um agradecimento especial para a minha “irmã do Doutoramento”, a Giulia, por ser uma amiga incrível e partilhar todas as minhas alegrias e frustrações como se fossem dela, como uma verdadeira irmã. Grazie! Ao David por ser o meu mago dos peixes e por partilhar comigo um sentido de humor muito especial. Só ele sabe onde fica Sassari. Ao Machado, por me introduzir à Bioinformática e ser diariamente o meu conselheiro da genómica e transcriptómica.

Agradeço a toda a minha grande família, de quem gosto muito, por todo o apoio e compreensão. Agradeço especialmente aos meus pais, por nunca terem questionado as minhas decisões e por me terem apoiado sempre incondicionalmente. Obrigado! Quero também agradecer ao meu irmão, Pedro, por ser um exemplo para mim. Será sempre o meu irmão mais velho. Agradeço também ao Du e espero que se um dia leres esta tese, consigas sentir algum orgulho pelo teu padrinho.

Quero também agradecer a todos os amigos que fiz nos vários grupos do CIIMAR, que me deram a conhecer outros mundos da ciência e que alegraram o dia a dia no CIIMAR.

A todos "Paquitos", por me receberem sempre alegremente e de braços abertos, especialmente a Marta, o Hugo e o Oscar.

Aos meus amigos do "PhD Committee", Ana, Diogo, Fernando, Luana, Rita, Sandra e Tiago, por toda a amizade, loucuras e alegrias partilhadas.

À Inês por todo carinho e pela partilha de conhecimento musical e artes.

E a todos com quem partilhei experiências e conhecimento durante estes anos, Axel, Cláudia, Dimitri, Fredi, Pipa e a Sílvia.

Quero também agradecer a todos os meus amigos da faculdade, Rui, Cib, Mário, Zé e Rita e aos meus amigos da Pousa, Adriana, Di, Bruno, Li, Miguel, Pedro e Vitor, que sempre me apoiaram nesta "busca pelo mexilhão".

Por fim quero agradecer às instituições que tornaram este trabalho possível. À Fundação para a Ciências e a Tecnologia (FCT), que financiou este trabalho, e às instituições que me acolheram, o Centro Interdisciplinar de Investigação Marinha e Ambiental (CIIMAR), a Faculdade de Ciências da Universidade do Porto (FCUP) e o Département de Sciences Biologiques Université de Montréal.

Resumo

Os extraordinários avanços na sequenciação do ADN/ARN ao longo das últimas duas décadas revolucionaram o campo da genómica. Estudos à escala do genoma são agora acessíveis à maioria dos laboratórios de biologia molecular do mundo, representando um passo importante no estudo de todos os ramos da Árvore da Vida (ToL). Apesar dos inegáveis sinais de progresso, esta realidade está ainda na sua infância para muitos grupos do ToL, incluindo grupos altamente diversos e ecologicamente relevantes, tais como os moluscos. O filo Mollusca é o segundo filo animal mais diversificado, compreendendo uma enorme variedade de organismos evolutivamente bem-sucedidos, colonizadores de quase todos os habitats do planeta. Os bivalves estão entre os moluscos mais diversificados, com mais de 20,000 espécies distribuídas em ecossistemas marinhos, salobros e de água doce. Atualmente, várias linhagens independentes de bivalves habitam ecossistemas de água doce. Entre elas, o grupo mais diversificado de bivalves que habitam exclusivamente habitats de água doce é formado por espécies comumente designadas por mexilhões de água doce (Bivalvia: Unionida). Os mexilhões de água doce possuem uma série de traços biológicos que lhes permitem prosperar nos ecossistemas de água doce, incluindo fertilização interna com "cuidados parentais" e larvas parasitárias altamente especializadas (glochidia), que atuam como parasitas obrigatórios de peixes (ou outros vertebrados) e asseguram a dispersão e nutrição até à metamorfose. Além disso, os mexilhões de água doce, tal como muitos outros bivalves, possuem um método muito invulgar de herança de ADN mitocondrial, chamado Dupla Herança Uniparental (DUI), onde duas linhagens de ADN mitocondrial podem ser segregadas diferencialmente entre géneros no decorrer da reprodução. Os mexilhões de água doce estão também entre os taxa mais ameaçados do planeta, com a família Margaritiferidae (margaritiferídeos) a ocupar uma posição particularmente preocupante, como o grupo de mexilhões de água doce mais ameaçado. Os margaritiferídeos incluem uma das espécies mais icónicas e bem estudadas de mexilhões de água doce, a *Margaritifera margaritifera* (Linnaeus, 1758). No entanto, o conhecimento sobre a biologia e ecologia da maioria das espécies de margaritiferídeos é ainda muito limitado. Além disso, a disponibilidade de recursos genómicos, tanto para margaritiferídeos como para mexilhões de água doce, é virtualmente inexistente, o que dificulta ainda mais a compreensão de muitos dos traços biológicos, ecológicos e evolutivos destas espécies.

O trabalho desenvolvido nesta tese visa avançar o estudo da biologia e evolução dos margaritiferídeos, através da aplicação de estratégias de sequenciação de nova geração (NGS), a fim de gerar novos recursos com aplicações em vários campos emergentes de genómica, tais como a filogenómica, a genómica populacional, a genómica de conservação e a genómica adaptativa.

Dada a recente história de estudos genómicos dos moluscos, esta tese inclui duas grandes revisões sobre o tema, primeiro, com foco na genómica num sentido mais amplo e, segundo, na genómica mitocondrial dos moluscos. O estudo dos moluscos está lentamente a entrar na era genómica, contudo está ainda muito atrás do estudo de outras linhagens de metazoários. Os recursos genómicos produzidos para moluscos são quase inteiramente representados por RAD-seq, transcriptómica, mitogenómica e sequenciação de genomas, e em grande parte focados em espécies conhecidas e comercialmente relevantes (ou seja, gastrópodes, bivalves marinhos e cefalópodes). Apesar disso, os poucos estudos já disponíveis começam a desvendar os mecanismos moleculares que estão subjacentes a muitas das características biológicas e adaptativas que diferenciam os moluscos. Os recursos mais abundantes para moluscos são mitogenomas, que representam, em vários taxa, a primeira fronteira na era da genómica. O catálogo já disponível (e em rápido crescimento) de mitogenomas de moluscos revelou a existência de frequentes exceções à "descrição universal" de um mitogenoma. Estas exceções, incluem variações significativas no tamanho e disposição dos genes, duplicações/perdas de genes, novos genes e ainda o DUI. Os mitogenomas de moluscos representam, portanto, modelos ideais para o estudo da evolução mitocondrial e funções adaptativas, bem como para a filogenómica e a genómica populacional.

A aplicação de estratégias moleculares alterou fundamentalmente os estudos de filogenética e a sistemática de mexilhões de água doce. Contudo, a maioria destes estudos ainda se baseavam num pequeno número selecionado de marcadores moleculares. Nesta tese, são gerados e aplicados vários recursos genómicos inovadores em dois estudos de filogenómica de margaritiferídeos. Estes estudos geraram os mais completos conjuntos de mitogenomas de margaritiferídeos até à data, bem como o primeiro conjunto de 813 marcadores nucleares obtidos usando a estratégia Anchor Hybrid Enrichment loci (AHE). Estas duas estratégias são utilizadas independentemente e comparadas para testar a sua utilidade em estudos de filogenética. Além disso, foi desenvolvida uma nova *pipeline* informática para a captura e montagem das regiões-alvo de AHE utilizando resultados de sequenciação de

genomas. Todos estes novos recursos servirão como ferramentas complementares para estudos filogenómicos, não só para margaritiferídeos, mas também para mexilhões de água doce como um todo.

O transcriptoma e genoma são talvez os recursos genómicos mais determinantes e difíceis de gerar para uma espécie. O genoma é especialmente desafiante, sobretudo em espécies com genomas de grande complexidade e tamanho, tais como os mexilhões de água doce. Contudo gerar estes recursos permite-nos disponibilizar “catálogos” nos quais os mecanismos moleculares da biologia das espécies são impressos e, portanto, abrir novos caminhos para o estudo das mesmas. Aqui, são produzidos os transcriptomas das brânquias de cinco espécies de mexilhões de água doce europeus, ou seja, *Margaritifera margaritifera*, *Unio crassus*, *Unio pictorum*, *Unio mancus* e *Unio delphinus*. Além disso, os dois primeiros genomas de margaritiferídeos foram produzidos para *M. margaritifera*, utilizando diferentes abordagens de NGS. Estes representam o quarto, e quinto, genomas de mexilhões de água doce disponíveis até à data. Os transcriptomas e genomas aqui produzidos são recursos chave para começar a explorar os mecanismos moleculares que governam muitas das intrigantes características biológicas, ecológicas e evolutivas dos mexilhões de água doce.

Palavras-chave: Sequenciação de ADN/ARN, Sequenciação de Nova Geração (NGS), Mitogenoma, Transcriptoma, Genoma, Mollusca, Bivalvia, Unionida, Mexilhões de água doce, Margaritiferidae, *Margaritifera margaritifera*

Abstract

The exceptional advancements in DNA/RNA sequencing undertaken over the last two decades revolutionized the field of genomics. Genome-scale studies are now accessible to most molecular biology laboratories in the world, representing a fundamental step towards reaching every branch of the Tree of Life (ToL). Despite the unarguable signs of progress, this reality is still in its infancy for many groups of the ToL, even for highly diverse, widespread, and ecologically relevant groups, such as molluscs. Mollusca is the second most diverse animal phylum, comprising many familiar organisms (e.g., mussels, snails, and octopuses), which over their long evolutionary history have attained outstanding adaptive success and colonised almost all habitats. Bivalves are among the most diversified molluscs, with over 20,000 species distributed throughout most marine, brackish and freshwater ecosystems. In freshwater, several independent lineages of bivalves exist today with the most diverse group of strictly freshwater bivalves being the freshwater mussels (FMs) (Bivalvia: Unionida). Freshwater mussels possess a series of characteristic biological features allowing them to thrive under the challenging conditions posed by freshwater ecosystems. This includes internal fertilization with 'parental care' and, most remarkably, highly specialized parasitic larvae (glochidia or lasidia), which act as obligatory parasites of fish (or other vertebrates) and ensures dispersion and nutrition during metamorphosis. Moreover, FMs, along with many other bivalves, possess a highly unusual method of mitochondrial DNA inheritance, called Doubly Uniparental Inheritance (DUI), in which two lineages of mitochondrial DNA can be differentially segregated between genders during reproduction. Freshwater mussels are also among the most imperilled taxa, with the family Margaritiferidae (margaritiferids or freshwater pearl mussels) occupying a particularly concerning position as the most threatened group of FMs. Margaritiferids include one of the most emblematic and well-studied species of FMs, the freshwater pearl mussel *Margaritifera margaritifera* (Linnaeus, 1758). However, the knowledge of the biology and ecology of most species of the group is still limited. Moreover, the availability of genomic resources for Margaritiferidae (and FMs) is practically inexistent, which further hampers a deeper understanding of many of the species' biological, ecological and evolutionary traits.

The work developed in this thesis aims to advance the study of the biology and evolution of margaritiferids through the application of several next-generation sequencing (NGS) approaches, to generate resources with applications in a myriad of emerging "omics"

fields, such as phylogenomics, population genomics, conservation genomics, and adaptative genomics.

Given the recent, but flourishing, history of the field of molluscan genomics, this thesis provides two broad revisions of the topic, firstly, focusing on genomics in a broader sense and, secondly, on molluscan mitochondrial genomics. The study of molluscs is slowly entering the genomic era but still lags far behind many other metazoan lineages. Genomic resources produced for molluscs are almost entirely represented by RAD-seq, transcriptome, mitogenome and whole-genome sequencing, and largely biased towards the most well-known and commercially relevant species (i.e., gastropods, marine bivalves and cephalopods). Despite this, the few studies already available are starting to unravel the molecular mechanisms underscoring many differentiating biological and adaptative novelties of molluscs. The most abundant resources for molluscs are mitogenome assemblies, which for many taxa represent the first frontier of the genomic era. The already available (and rapidly increasing) catalogue of molluscan mitogenomes has revealed frequent deviations and exceptions to the ‘textbook description’ of a mitogenome. This includes significant variations in size and gene arrangements, gene duplications/losses, the emergence of novel genes and DUI. Consequently, molluscan mitogenomes represent ideal models for studying mitochondrial evolution and adaptation roles, as well as for phylogenomic and population genomics.

The application of molecular approaches has fundamentally changed the phylogenetics and systematics of FMs. However, most of these studies, including the most comprehensive phylogenetic study available for margaritiferids, still relied on a small number of selected molecular markers. In this thesis, a series of novel genomic resources are generated and applied in two comprehensive phylogenomic studies of margaritiferids. This includes the most comprehensive dataset of margaritiferids whole mitogenomes assemblies, as well as the first family-wide dataset of 813 Anchor Hybrid Enrichment loci (AHE), which are independently used and compared for their phylogenetic applications. Moreover, a new highly efficient pipeline for capturing and *de novo* assembly of the AHE targeted regions, using whole genome re-sequencing reads, is developed and a catalogue of well-curated functional annotations of the targeted regions is provided. All these new resources will serve as complementary tools for phylogenomic studies, not only within margaritiferids but also within FMs as a whole.

Perhaps the two most defining and challenging genomic resources to generate for a species are its transcriptome and whole-genome assemblies. The latter is particularly changeling for species with highly complex and large genomes, such FMs, which have

among the largest, within molluscs. However, undertaking these endeavours, allow us to generate catalogued frameworks, in which the molecular mechanisms of species' biology are imprinted, opening new ways to study them. Here, the gill transcriptomes of five threatened European FMs are produced, i.e., for *Margaritifera margaritifera*, *Unio crassus*, *Unio pictorum*, *Unio mancus* and *Unio delphinus*. Moreover, the first two margaritiferids' whole genome assemblies are produced for *M. margaritifera,* using distinct NGS sequencing approaches. These represent the fourth and fifth FMs genome assemblies available to date, with one of them being the most contiguous and complete FMs genome assembly. The transcriptomes and genomes here produced are key resources to start exploring the molecular mechanisms that govern many of the FMs’ intriguing biological, ecological, and evolutionary features.

Overall, the present thesis provides a series of novel, but timely needed, multiscale genomics resources for margaritiferids (and FMs), that will propel the study of these organisms in the genomics era. The important advances here presented will have multiple applications in several emerging fields, such as phylogenomics, population genomics, conservation genomics, and adaptative genomics, which will help to better comprehend the biology of this fascinating group of organisms and ultimately promote their conservation.

Keywords: DNA/RNA sequencing, Next Generation Sequencing (NGS), Mitogenome, Transcriptome, Genome, Mollusca, Bivalvia, Unionida, Freshwater Mussels, Margaritiferidae, *Margaritifera margaritifera*.

Table of Contents

List of Tables xii

List of Figures xv

List of Abbreviations xxi

Chapter 1 – General Introduction 24

1 – *Beginnings*: nucleic acid sequencing 24

2 – “*Sequences, sequences, and sequences*” 26

3 – The Human genome sequence: the start of a new revolution 28

4 – Next Generation Sequencing (NGS).... 29

5 – Genomic revolution and the path to the Tree of Life (ToL) 32

6 – Phylum Mollusca 35

7 – Biology and conservation of freshwater mussels (Bivalvia: Unionida).... 39

8 – The Family Margaritiferidae 42

9 – Objectives and thesis structure.... 45

Chapter 2 – The study of Mollusca in the Genomic era 49

1.1. Molluscan genomics: the road so far and the way forward.... 49

1.2. Molluscan mitochondrial genomes break the rules.... 74

Chapter 3 – Freshwater mussels (Bivalvia: Unionida) 105

2.1. The male and female complete mitochondrial genomes of the threatened freshwater pearl mussel *Margaritifera margaritifera* (Linnaeus, 1758) (Bivalvia: Margaritiferidae) 105

2.2. A novel assembly pipeline and functional annotations for targeted sequencing: A case study on the globally threatened Margaritiferidae (Bivalvia: Unionida) 112

2.3. The gill transcriptome of threatened European freshwater mussels 137

2.4. The Crown Pearl: a draft genome assembly of the European freshwater pearl mussel *Margaritifera margaritifera* (Linnaeus, 1758).... 153

2.5. Manuscript 1 – The Crown Pearl V2: an improved genome assembly of the European freshwater pearl mussel *Margaritifera margaritifera* (Linnaeus, 1758) ... 174

Chapter 4 – General Discussion 185

3.1 Phylogenomic applications 185

3.2 Mitogenomic of Freshwater Mussels 191

3.3 From Genomes to Phenotypes 197

Conclusion 207

References 209

List of Tables

Chapter 1 – General Introduction

Table I. 1 - List of some global Consortiums aiming to increase genomics data on different organisms through the Tree of Life (ToL). 34

Chapter 2 - Molluscan genomics: the road so far and the way forward

Table P1. 1 - Number and respective percentages of invertebrate species mitogenomes, transcriptome shotgun assemblies (TSA), whole-genome shotgun (WGS) and total number of sequence read archive (SRA). The percentages were calculated based on the estimated number of species reported in Brusca et al. (2003)............................. 60

Table P1. 2 - Main characteristics of molluscan genomes projects and genomes statistics, as reported in their respective references. The codes used for the sequencing technologies refer to: Illumina Paired-end (PE) + Illumina Mate Paired (MP), Pacific Biosciences (PacBio) + Dovetail Genomics Hi-C or Chicago libraries (CD/Hi-C), Pyrosequencing Roche 454 (454), EST Sanger sequencing (EST), Bacterial Artificial Chromosome (BAC), 10× Chromium Genomics (10x Chr), Fosmid Clones (FC), Linkage maps 2b-RAD (2b-RAD), Linkage maps RAD-seq (RAD-seq) and Synthetic long-read DNA (TSLR). For the Benchmarking Universal Single-Copy Orthologs (BUSCO), the analysis reported with Metazoa (M) and Eukaryotic (Eu) database is shown. For Core Eukaryotic Genes Mapping Approach (CEGMA), the Total (T), Complete (C) and Partial (P) results are shown. Scaffold N50 values marked with * refer to assemblies that reached chromosome level. Heterozygosity values marked with × were reported to be high, in the reference, but the values were not expressed. Heterozygosity values marked with +* refer to individuals artificially inbreeded prior to the assembly (**To see the expanded Figure open the link** https://link.springer.com/article/10.1007/s10750-019-04111-1/tables/2)... 62

Chapter 2 - Molluscan mitochondrial genomes break the rules

Table P2. 1 - Number of molluscan mitogenome sequences in GenBank over time. The GenBank (GB) search was structured as follows: ((‘Mollusca’[Organism]) AND (biomol_genomic[PROP] AND mitochondrion[filter] AND (‘8000’[SLEN]:

‘100000’[SLEN]) AND (‘1900/01/01’[PDAT]] : ‘1999/12/31’[PDAT])). The term ‘Mollusca’ was replaced for family-level searches with ‘Gastropoda; Bivalvia; Scaphopoda; Cephalopoda; Polyplacophora; Monoplacophora; Aplacophora’ and the years were adjusted for specific time intervals. 78

Chapter 3 - A novel assembly pipeline and functional annotations for targeted sequencing: A case study on the globally threatened Margaritiferidae (Bivalvia: Unionida)

Table P4. 1 - List of samples used in each dataset and respective NCBI accession codes. * Only used in Figure P4.S1. 118

Table P4. 2 - General statistics of AHE assemblies and probe capturing count. * The pipeline to generate matrices alignments followed (Breinholt et al., 2018). 125

Table P4. 3 - General characteristics of the alignment matrices used for phylogenetic inferences using AHE and mtDNA. + Includes a *Margaritifera margaritifera* sample originating from RNA-seq probe assembly (RNAassembly). * TPM2u+F+I+G4: 1atp6 1nad3, GTR+F+I+G4: 2atp6 2nad4 2nad4l 2nad5, TPM3+F+R3: 3atp6 3nad3, TIM2+F+I+G4: 1cob 1nad1 rrnL rrnS, GTR+F+I+G4: 2cob 2nad1, TIM3+F+R3: 3cob, TN+F+I+G4: 1cox1 1cox2 1cox3 2nad3, TVM+F+R2: 2cox1 2cox2 2cox3, TPM2u+F+R3: 3cox1, K3Pu+F+I+G4: 3cox2 3cox3, TN+F+G4: 3nad1 3nad2 3nad6, TIM2+F+I+G4: 1nad2 1nad6, TPM3u+F+G4: 2nad2 2nad6, TIM2+F+I+G4: 1nad4 1nad4l 1nad5, K3Pu+F+G4: 3nad4, K3Pu+F+I+G4: 3nad4l, TPM2u+F+R2: 3nad5.. 126

Chapter 3 - The gill transcriptome of threatened European freshwater mussels

Table P5. 1 - MixS descriptors for the five freshwater mussel species. 142

Table P5. 2 - Basic statistics of raw sequencing datasets and percentages of removed reads at each step of the preassembly processing strategy. 148

Table P5. 3 - Transrate and Busco scores of redundant and non-redundant gill transcriptome assemblies for each species. *euk/met. Euk: Dataset with 303 genes of Eukaryota library profile. Met: Dataset with 978 genes of Metazoa library profile. 149

Table P5. 4 - Structural and functional annotation statistics for the final gill transcriptome assemblies for each species. 151

Chapter 3 - The Crown Pearl: a draft genome assembly of the European freshwater pearl mussel *Margaritifera margaritifera* (Linnaeus, 1758)

Table P6. 1 - *Margaritifera margaritifera* genome assembly, read alignment, gene prediction, and annotation general statistics. *All statistics are based on contigs/scaffolds of size ≥1,000bp; # Euk: From a total of 303 genes of Eukaryota library profile; # Met: From a total of 978 genes of Metazoa library profile; + Euk: From a total of 255 genes of Eukaryota library profile; + Met: From a total of 954 genes of Metazoa library profile; #,+ C: Complete; S: Single; D: Duplicated; F: Fragmented; ± All statistics are based on contigs/scaffolds of size ≥2,500bp. 164

Table P6. 2 - Statistics of the content of repetitive elements in the *M. margaritifera* genome assembly............ 167

Chapter 3 - The Crown Pearl V2: an improved genome assembly of the European freshwater pearl mussel *Margaritifera margaritifera* (Linnaeus, 1758)

Table M1. 1 - MixS descriptors for the freshwater pearl mussel *Margaritifera margaritifera* specimen used for whole genome sequencing............ 177

Table M1. 2 - List of proteomes used for BRAKER2 gene prediction pipeline. 181

Table M1. 3 - General statistics of the *Margaritifera margaritifera* genome assemblies (V1 and V2) and other published freshwater mussel’s genome assemblies. * Genome V2 is at solely contig level, i.e., has no scaffolds; # Euk: From a total of 303 genes of Eukaryota library profile; # Met: From a total of 978 genes of Metazoa library profile; + Euk: From a total of 255 genes of Eukaryota library profile; + Met: From a total of 954 genes of Metazoa library profile; #,+ C: Complete; S: Single; D: Duplicated; F: Fragmented. 183

List of Figures

Chapter 1 – General Introduction

Figure I. 1 - A timeline of major discoveries and releases in DNA/RNA sequencing since the 1950s. 25

Figure I. 2 - Species diversity in the Sequence Read Archive (SRA). The amount of human data exceeds that of the next top 10 species, measured as (A) terabases and (B) individuals sequenced. (C) The human proportion increased between 2010 and 2020, and (D) the proportion from species without known commercial/medical relevance (“other”) dropped. (E) A tiny proportion of IUCN-recognized (80) species have a reference genome (red) or are otherwise represented in the SRA (dark grey). Retrieved November 14, 2020. Adapted from Stephan et al., 2022........ 33

Figure I. 3 - Phylogenetic inference of phylum Mollusca (outgroup taxa not shown). (A) RAxML Maximum Likelihood (ML) tree with bootstrap support values below 100 shown. (B) IQ-TREE ML tree with bootstrap support values below 100 shown. (C) PhyloBayes Bayesian Inference (BI) tree with posterior probabilities below 1.0 shown. (D) ASTRAL tree with local posterior probabilities below 1.0 shown. Adapted from Kocot et al., 2020. 37

Figure I. 4 - Graphical representations of the mitochondrial DNA inheritance system called Doubly Uniparental Inheritance (DUI). 41

Figure I. 5 - Freshwater Mussels global conservation categories accessed by the Red List of the IUCN. 42

Figure I. 6 - a-c) Freshwater pearl mussel *Margaritifera margaritifera* in its natural environment. Pictures in panels b and c were taken within 10min apart and show two specimens burrowing themselves in the sediment using their muscular foot. Photos by André Gomes-dos-Santos (August 2022). 44

Chapter 2 - Molluscan genomics: the road so far and the way forward

Figure P1. 1 - Number of publications per year, since 2009, in peer-reviewed journals by querying the Web of Knowledge using the following terms: “RAD”, “RAD-seq”, “mitogenome” and “transcriptome” associated with Mollusca, mollusk, mollusc, mollusks, gastropod, Gastropoda, bivalve, Bivalvia, Cephalopoda, cephalopods, Caudofoveata, Aplacophora, Polyplacophora, Monoplacophora and Scaphopoda. Coloured lines refer to the four main genomic approaches applied to molluscs, i.e. restriction site-associated DNA, mitogenome, transcriptome and whole-genome sequencing, as well as a total of other less frequently used genomic resources. References for these publications are detailed in Supplementary Table P1.1. 57

Figure P1. 2 Number of bivalves, cephalopods and gastropods’ whole-genome shotgun (left), transcriptome shotgun assemblies (right) and mitogenome assemblies (right) available on NCBI at 29 June 2019. Black and yellow lines represent the effective number of submits per year. 58

Figure P1. 3 - Compiled data set comprising 304 publications (2009 to June 2019) in peer-reviewed journals by querying the Web of Knowledge using the terms described in Figure P1.1 legend. References for these publications are detailed in Supplementary Table P1.1. The aims were divided into 13 main categories with others referring to the total of less frequently observed aims. Top columns: relative proportion of each of the categories regarding the four main genomic approaches applied to molluscs (restriction site-associated DNA, mitogenome, transcriptomics and whole-genome sequencing, and other for less frequently used genomic resources). Bottom columns: relative proportion of the molluscan taxa (cephalopods, gastropods, bivalves and other) represented by each of the four main genomic approaches applied. 59

Chapter 2 - Molluscan mitochondrial genomes break the rules

Figure P2. 1 - Relationship between the length of (a) non-coding and (b) coding regions on total mtDNA length in molluscan classes. Variation in non-coding length explains a greater proportion of variation in total mtDNA length compared to variation in coding length. Each circle represents a single species. When multiple mtDNAs were available for a single species, the mean across all individual records was taken as the species value. Colours represent different molluscan classes and are indicated by the key in (a). 80

Figure P2. 2 - In mitogenomes of planobid gastropods, the atp8 gene is bracketed by trnN(aac) and trnL2(tta). Shaded boxes, tRNA genes; white boxes, protein-coding genes; arrowheads indicate directionality; asterisk, stop codon. ORF analyses of the mitogenome sequences that ignore the concept of tRNA gene excision from polycistronic mitogenomic transcripts frequently yield incorrect prediction of protein-coding sequence intervals. Whereas the start codon is correctly indicated, the ORF for atp8 from *Biomphalaria glabrata* (underlined in both nucleotide and predicted amino acid sequences, NC_005439) falls short, despite an effort to accommodate an incomplete stop codon (T--). Another issue impacts the ORF selected from the *Planorbella duryi* mitogenome (KY514384). It comprises a (correct) start codon and TAA stop codon but overlaps with trnL2 and yields an unusually long protein sequence. For both snail species, considering the boundaries of the (MITOS predicted) tRNA genes, the ATA is the first possible start codon downstream from trnN. At the 3′ end, a single T nucleotide remains after excision of trnL2, completed by polyadenylation to a TAA (underlined) stop codon. Such peculiarities challenge prediction of multiple genes from molluscan mitochondrial sequences, as is evidenced in several GenBank entries, despite the purported curation of submissions by this NCBI database. Re-evaluation and, if appropriate, updates by contributors of previous GenBank accessions will greatly benefit correct annotation. 90

Figure P2. 3 - Top: Graphic showing the number of complete (dark colours) and partial (light colours: min. size 10 000 bp) mitogenomes available in GenBank; middle: mean, minimum and maximum size (bp) of complete mitogenomes per Mollusca class; bottom: graphic showing the percentage of total species with complete mitogenomes published in GenBank. Asterisk superscripts refer to unverified size values, due to assembly challenges; critical evaluation of these publicly available mitogenome sizes and sequence content is highly recommended. 101

Chapter 3 - The male and female complete mitochondrial genomes of the threatened freshwater pearl mussel *Margaritifera margaritifera* (Linnaeus, 1758) (Bivalvia: Margaritiferidae)

Figure P3. 1 - Margaritiferidae and Unionidae Bayesian phylogenetic tree of Male and Female mitogenomes sequences based on concatenated nucleotide sequences of 13 mitochondrial protein-coding genes and the two rRNA genes. GenBank accession numbers are behind species names, numbers at the nodes indicate the percentage posterior probabilities and bootstrap support values. The * above the branches indicate posterior probabilities and bootstrap support values > 95%. 110

Chapter 3 - A novel assembly pipeline and functional annotations for targeted sequencing: A case study on the globally threatened Margaritiferidae (Bivalvia: Unionida).

Figure P4. 1 - Maximum Likelihood phylogenetic trees of family Margaritiferidae based on: a) concatenated alignments of 514 Unioverse loci (head+probe+tail; n=1541); b) concatenated alignments of the mitochondrial DNA from 12 PCG and 2 rRNA; *Above the nodes refer to bootstrap with maximum support. 127

Figure P4. 2 - Frequency distribution of Unioverse probe regions per *Bathymodiolus platifrons* whole transcripts. Top right corner is a schematic representation of several probe regions originating from the same gene. 128

Figure P4. 3 - Percentages distributions of the Gene Ontology (GO) Terms main categories count for: a) All Unioverse probes; b) The Unioverse loci used for the final AHE Margaritiferidae samples alignments. 130

Figure P4. 4 - Percentage of total Unioverse loci assigned to three BUSCO databases. Euk - Dataset with 255 genes of Eukaryota library profile; Met - Dataset with 954 genes of Metazoa library profile; Mus - Dataset with 5295 genes of Mollusca library profile; No BUSCO – No match found. 131

Chapter 3 - The gill transcriptome of threatened European freshwater mussels

Figure P5. 1 - Maps of the five species’ potential distributions produced by overlapping points of recent presence records (obtained from Lopes-Lima et al., 2017c) with the Hydrobasin level 5 polygons (Lehner and Grill, 2013). Overlapping distribution polygons between *Unio mancus* and *Unio crassus* are represented by a light purple shade, in the left panel. Overlapping distribution polygons between *Unio pictorum* and *Margaritifera margaritifera* are represented by an orange shade, in the right panel. 141

Figure P5. 2 - FastQC quality report of the trimmed and decontaminated RNA-seq reads (after Centrifuge for each species. (a) *Margaritifera margaritifera*; (b) *Unio crassus*; (c) *Unio pictorum*; (d) *Unio mancus*; and (e) *Unio delphinus*. 144

Figure P5. 3 - Bioinformatics pipeline applied for the transcriptome assembly and annotation. Auxiliary representative figures were created with BioRender.com. 147

Chapter 3 - The Crown Pearl: a draft genome assembly of the European freshwater pearl mussel *Margaritifera margaritifera* (Linnaeus, 1758)

Figure P6. 1 - *Margaritifera margaritifera* specimens in their natural habitat. Red arrows point to two individuals. 158

Figure P6. 2 - (A) GenomeScope2 k-mer (25 and 31) distribution displaying estimation of genome size (len), homozygosity (aa), heterozygosity (ab), mean k-mer coverage for heterozygous bases (kcov), read error rate (err), the average rate of read duplications (dup), k-mer size used on the run (k:), and ploidy (p:). (B) *Margaritifera margaritifera* genome assembly assessment using KAT comp tool to compare the Illumina PE k-mer content within the genome assembly. Different colours represent the read k-mer frequency in the assembly. 163

Figure P6. 3 - Maximum Likelihood phylogenetic tree based on concatenated alignments of 118 single-copy orthologous amino acid sequences retrieved by OrthoFinder. *Above the nodes refer to bootstrap and posterior probabilities support values above 99%. .. 168

Figure P6. 4 - Hox and ParaHox Maximum Likelihood gene tree constructed using Mollusca homeodomain amino acid sequences. Bootstrap values are presented above the nodes. Red squares highlight the position of *M. margaritifera* Hox and ParaHox. 170

Chapter 3 - The Crown Pearl V2: an improved genome assembly of the European freshwater pearl mussel *Margaritifera margaritifera* (Linnaeus, 1758)

Figure M1. 1 - Bioinformatics pipeline applied for the genome assembly and annotation. 178

Figure M1. 2 - (a) GenomeScope2 k-mer (21) distribution displaying the estimation of genome size (len), homozygosity (aa), heterozygosity (ab), mean coverage of k-mer for heterozygous bases (kcov), read error rate (err), average rate of read duplications (dup), size of the k-mer used on the run (k:), and ploidy (p:). (b) *Margaritifera margaritifera* genome assembly assessment using KAT comp tool to compare the Illumina paired-end k-mer content within the genome assembly. Different colors represent the read k-mer frequency in the assembly. 182

General Discussion

Figure D. 1 - a) Maximum Likelihood reconstruction from a nucleotide supermatrix of 569 AHE loci from Bivalvia. Colour blocks highlight the six different Unionida families, as well as the marine bivalve family Trigoniidae. The clade including all representatives of the family Unionidae is collapsed for clarity. Node support values are not displayed as they were 100 for both bootstrap in Maximum Likelihood and posterior probabilities in Bayesian Inferences analyses. Figure adapted from Pfeiffer et al., 2019; b) Maximum Likelihood reconstruction from concatenated nucleotide alignment of genes COI [3 codons], 16S, 18S, 28S and H3 [3 codons] from Paleoheterodonta. Support values above the branches are posterior probabilities and below the branches are bootstrap supports. Figure adapted from Lopes-Lima et al., 2018a. 189

Figure D. 2 - a) Linear representation of all the currently known mitochondrial gene orders for FMs and Trigoniidae family; b) Circular representation of the typical F, M and H-type mitogenomes of margaritiferids. Adapted from Guerra et al., 2019 193

Figure D. 3 - Potencial application of the whole genome assembly of the freshwater pearl mussel *Margaritifera margaritifera*.. 206

List of Abbreviations

AHE	ANCHORED HYBRID ENRICHMENT
ANTP	ANTENNAPEDIA HOMEOBOX GENE CLASS
BLAST	BASIC LOCAL ALIGNMENT SEARCH TOOL
BI	BAYESIAN INFERENCE
BC	BEFORE CHRIST
BUSCO	BENCHMARKING UNIVERSAL SINGLE-COPY ORTHOLOGS
BWA	BURROWS-WHEELER ALIGNER
CCS	CIRCULAR CONSENSUS SEQUENCES
CDS	CODING REGION
CLR	CONTINUOUS LONG READ
ddNTP	2,3-DIDEOXYNUCLEOSIDE TRIPHOSPHATE
DNA	DEOXYRIBONUCLEIC ACID
DUI	DOUBLY UNIPARENTAL INHERITANCE
DOGMA	DUAL ORGANELLAR GENOME ANNOTATOR
ESTs	EXPRESSED SEQUENCE TAGS
FCUP	FACULTY OF SCIENCES OF THE UNIVERSITY OF PORTO
F-TYPE	FEMALE INHERITED MITOCHONDRIAL DNA
F-*orf*	FEMALE MITOGENOMES SPECIFIC UNIQUE PUTATIVE GENES WITH UNKNOWN HOMOLOGY OR FUNCTION
GO	GENE ONTOLOGY
ORFAN	GENES WITH UNKNOWN HOMOLOGY OR FUNCTION
gDNA	GENOMIC DNA
GIGA	GLOBAL INVERTEBRATE GENOMICS ALLIANCE
AGAT	GTF/GFF ANALYSIS TOOLKIT
H-TYPE	HERMAPHRODITIC INHERITED MITOCHONDRIAL DNA
H-*orf*	HERMAPHRODITIC MITOGENOMES SPECIFIC UNIQUE PUTATIVE GENES WITH UNKNOWN HOMOLOGY OR FUNCTION
HMW	HIGH MOLECULAR WEIGHT
HTS	HIGH-THROUGHPUT SEQUENCING
HGP	HUMAN GENOME PROJECT
IUCN	INTERNATIONAL UNION FOR CONSERVATION OF NATURE
KEGG	KYOTO ENCYCLOPEDIA OF GENES AND GENOMES

LINES	LONG INTERSPERSED NUCLEAR ELEMENTS
LTR	LONG TERMINAL REPEATS
M-*orf*	MALE MITOGENOMES SPECIFIC UNIQUE PUTATIVE GENES WITH UNKNOWN HOMOLOGY OR FUNCTION
M-TYPE	MALES INHERITED MITOCHONDRIAL DNA
MP	MATE PAIR
ML	MAXIMUM LIKELIHOOD
MBP	MEGA BASE PAIRS
MYA	MILLION YEARS AGO
MITOBIM	MITOCHONDRIAL BAITING AND ITERATIVE MAPPING
MEGA	MOLECULAR EVOLUTIONARY GENETICS ANALYSIS
MAFFT	MULTIPLE ALIGNMENT USING FAST FOURIER TRANSFORM
MUSCLE	MULTIPLE SEQUENCE COMPARISON BY LOG-EXPECTATION
NCBI	NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION
NGS	NEXT GENERATION SEQUENCING
NUMTs	NON-FUNCTIONAL NUCLEAR COPIES OF MITOCHONDRIAL SEQUENCES
ORF	OPEN READING FRAME
ONT	OXFORD NANOPORE TECHNOLOGIES
OXPHOS	OXIDATIVE PHOSPHORYLATION
PACBIO	PACIFIC BIOSCIENCES
PE	PAIRED-END
PCR	POLYMERASE CHAIN REACTION
PCG	PROTEIN-CODING GENES
QUAST	QUALITY ASSESSMENT TOOL FOR GENOME ASSEMBLIES
RAD-SEQ	RESTRICTION SITE ASSOCIATED DNA SEQUENCING
RNA	RIBONUCLEIC ACID
rRNA	RIBOSOMAL RNA GENES
SBL	SEQUENCING BY LIGATION
SBS	SEQUENCING BY SYNTHESIS
SINES	SHORT INTERSPERSED NUCLEAR ELEMENTS
SMRT	SINGLE MOLECULE REAL-TIME

SNPs	SINGLE NUCLEOTIDE POLYMORPHISMS
SMLS	SINGLE-MOLECULE LONG-READ SEQUENCING
SMS	SINGLE-MOLECULE SEQUENCING
SNCRNAs	SMALL NON-CODING RNAS
SLR	SYNTHETIC LONG-READS
KAT	THE K-MER ANALYSIS TOOLKIT
SOLID	THERMO FISHER SEQUENCING BY OLIGONUCLEOTIDE LIGATION AND DETECTION
tRNA	TRANSFER RNA
TE	TRANSPOSABLE ELEMENTS
TOL	TREE OF LIFE
UCE	ULTRA CONVERSED ELEMENT
UP	UNIVERSITY OF PORTO
WGS	WHOLE GENOME SEQUENCING

Chapter 1 – General Introduction

1 – *Beginnings*: nucleic acid sequencing

Two decades separate the announcements of the first human genome sequence (Craig Venter et al., 2001; Lander et al., 2001) and the recently described gapless human genome assembly (Nurk et al., 2022). In between, a critical revolution occurred in genome biology, providing a fundamental shift in the process and the scale at which DNA/RNA are studied (Goodwin et al., 2016; L. Koch et al., 2021; Sedlazeck et al., 2018; Stephan et al., 2022). Today, nucleic acid sequencing has become a mundane, cheap and easy task accessible to most molecular biology laboratories in the world. The history that led to this stage started almost 60 years ago and was possible due to a series of landmark discoveries and projects throughout the second half of the 20th century and the beginning of the 21st century (Giani et al., 2020; Heather and Chain, 2016; Hutchison, 2007).

In 1953, soon after the Watson and Crick publication that unveiled the three-dimensional structure of DNA (critically helped by crystallographic studies of Rosalind Franklin and Maurice Wilkins) (Watson and Crick, 1953; Zallen, 2003), it became clear that the next step would be to develop "*read*" approaches for nucleic acid sequences. Although protein chains were already being sequenced at the time, nucleic acid molecules were inherently very different, and distinct sequencing approaches were required, thus the "rush" for DNA sequencing began (Heather and Chain, 2016; Hutchison, 2007). Initial progress was slow, as DNA shares several properties that made the process difficult, such as the large chain lengths with only four very similar monomeric unities, which was further complicated by the lack of known base-specific DNAases at the time (Giani et al., 2020; Heather and Chain, 2016; Hutchison, 2007). For that reason, the first efforts focused on some RNA molecules (microbial ribosomal or transfer RNA) that lack some of these shortcomings (Giani et al., 2020; Heather and Chain, 2016; Hutchison, 2007) and, in 1965, Robert Holley and colleagues sequenced the first ever nucleic acid molecule, a 76 base long (bp) alanine tRNA of *Saccharomyces cerevisiae* (Holley et al., 1965) (Figure I.1). Between 1968 and 1971, Wu and colleagues, using DNA polymerase for the incorporation of radiolabel nucleotides at the cohesive ends of the purified linear molecule of viral phage lambda DNA, reported the first DNA sequence. i.e., a 12 bp complementary fragment of the cohesive end (Kaiser and Wu, 1968; Wu and Kaiser, 1968; Wu and Taylor, 1971) (Figure I.1). Ray Wu also proposed a generalised method to apply this sequencing approach based on synthetic location-specific oligonucleotide primers during DNA sequencing reactions, which to some degree is still the fundamental principle at the base of many sequencing approaches that came afterwards (Padmanabhan et al., 1972; Ray et al., 1973). However, RNA sequencing was still head and, in parallel, Sanger and colleagues developed a related approach that detected radiolabelled partial-digestion fragments after two-dimensional fractionation (2-D fractionation) (Sanger et al., 1965) (Figure I.1). This approach allowed the sequencing of several ribosomal and transfer RNA fragments, including the first complete protein-coding gene of a coat protein of bacteriophage MS2, generated in 1972 by Walter Fiers team (Jou et al., 1972) and soon after the first-ever complete RNA genome, the 3,569 bp long RNA bacteriophage MS2 genome (Fiers et al., 1976) (Figure I.1).

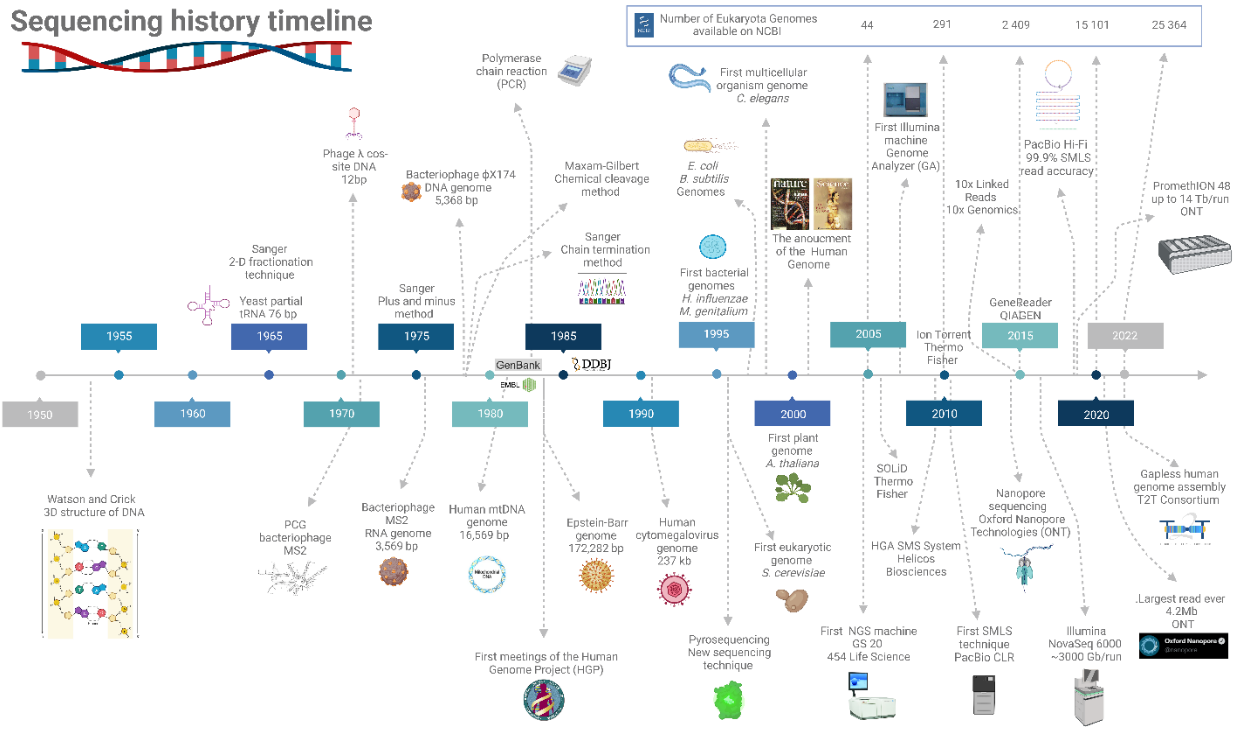

Sequencing history timeline
Number of Eukaryota Genomes available on NCBI
44
291
2 409
15 101
25 364
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020
2022
Watson and Crick 3D structure of DNA
Sanger 2-D fractionation technique
Yeast partial tRNA 76 bp
Phage λ cos-site DNA 12bp
PCG bacteriophage MS2
Bacteriophage ϕX174 DNA genome 5,368 bp
Sanger Plus and minus method
Bacteriophage MS2 RNA genome 3,569 bp
Polymerase chain reaction (PCR)
Maxam-Gilbert Chemical cleavage method
Sanger Chain termination method
GenBank
EMBL
DDBJ
Human mtDNA genome 16,569 bp
Epstein-Barr genome 172,282 bp
First meetings of the Human Genome Project (HGP)
Human cytomegalovirus genome 237 kb
First multicellular organism genome C. elegans
E. coli B. subtilis Genomes
First bacterial genomes H. influenzae M. genitalium
Pyrosequencing New sequencing technique
First eukaryotic genome S. cerevisiae
First plant genome A. thaliana
The anoucment of the Human Genome
First Illumina machine Genome Analyzer (GA)
First NGS machine GS 20 454 Life Science
SOLiD Thermo Fisher
HGA SMS System Helicos Biosciences
Ion Torrent Thermo Fisher
10x Linked Reads 10x Genomics
First SMLS technique PacBio CLR
GeneReader QIAGEN
Nanopore sequencing Oxford Nanopore Technologies (ONT)
PacBio Hi-Fi 99.9% SMLS read accuracy
Illumina NovaSeq 6000 ~3000 Gb/run
PromethION 48 up to 14 Tb/run ONT
Gapless human genome assembly T2T Consortium
.Largest read ever 4.2Mb ONT
Oxford Nanopore

Figure I. 1 - A timeline of major discoveries and releases in DNA/RNA sequencing since the 1950s.

The flourishing of DNA sequencing began in the second half of the 1970s, initially with the development of the plus and minus method for DNA sequencing by Sanger and Coulson in 1975 (Sanger and Coulson, 1975), which was used for the sequencing of the first DNA genome from the 5,368bp long bacteriophage ϕX174 (Sanger et al., 1977). In the same year, two new DNA sequencing methods were presented, which many regarded as the “first generation sequencing” approaches, the Maxam and Gilbert chemical base-specific cleavage approach (Maxam and Gilbert, 1977) and Sanger’s polymerase reaction “chain termination method” or dideoxy technique (Sanger et al., 1977) (Figure I.1). Gilbert’s approach, simply known as Sanger sequencing, was the first technique to be widely implemented, later overthrown after improvements that further simplify the technical application and reduce the cost of the dideoxy technique (Giani et al., 2020; Heather and Chain, 2016; Hutchison, 2007; Shendure et al., 2017). The original Sanger method relied on the addition, at a reduced concentration, of four different radiolabel 2,3-dideoxynucleoside triphosphate (ddNTP) in separate polymerization reactions, which when incorporated by DNA polymerase resulted in the termination of elongation (due to the lack of a 3′-hydroxyl group), thus producing fragments with different lengths (Sanger et al., 1977). The results of the four individual reactions were run on a polyacrylamide gel and through autoradiography, the nucleotide sequence composition was inferred at each corresponding reaction termination length (Sanger et al., 1977). Several improvements allowed standardization and automatization of Sanger sequencing, establishing its potential and opening the way for the first commercial sequencing machines (Giani et al., 2020; Heather and Chain, 2016; Hutchison, 2007; Shendure et al., 2017). The two major advances were the replacement of radiolabelling with fluorometric detection (allowing a single polymerization reaction) and detection through capillary-based electrophoresis (Ansorge et al., 1987, 1986; Kambara et al., 1988; Luckey et al., 1990; Prober et al., 1987; Smith et al., 1985; Swerdlow and Gesteland, 1990). In the end, despite using many of the same principles of previous techniques (i.e., labelling the last nucleotide of distinct DNA fragments), Sanger sequencing offered both accuracy, robustness, and technical simplicity that made it the ruler of the sequencing world for the next 40 years (Giani et al., 2020; Heather and Chain, 2016; Hutchison, 2007; Shendure et al., 2017) (Figure I.1).

2 – “*Sequences, sequences, and sequences*”

The first revolution in DNA/RNA sequencing had just started, which is testified by the number/complexity of sequencing projects, commercially available sequencing platforms, data processing approaches and genomic repositories that emerged during the 1980s and 1990s (Giani et al., 2020; Heather and Chain, 2016; Hutchison, 2007) (Figure I.1). In 1981 the first 16,569 bp human mitochondrial genome was sequenced (Anderson et al., 1981). The next year, Sanger and colleagues applied the then recently developed "shotgun sequencing" (i.e., sequence several random overlapping fragments of a long sequence for posterior *in silico* assemblage)(Messing et al., 1981; Staden, 1979) to construct the 48,502 bp complete phage lambda genome (Sanger et al., 1982). In 1984, Bart Barrell and colleagues produced the 172,282 bp long Epstein–Barr virus (Baer et al., 1984), and six years later the 237 kb human cytomegalovirus genome (Bankier et al., 1991). Throughout this period the constant increase in sequencing studies both aided and was boosted by several methodologic innovations, such as recombinant DNA (Cohen et al., 1973; Jackson et al., 1972) and polymerase chain reaction (PCR) (Saiki et al., 1988, 1985). Reflecting on the excitement of this period, in 1988 Sanger publishes a review entitled "Sequences, sequences, and sequences" (Sanger, 1988).

The constant increase in sequencing outputs demanded new ways for data delivery/storage, thus many sequencing repositories also emerged during the 1980s, including the three main sequencing repositories today, i.e., GenBank, the US National Institute of Health (NIH) sequence database (https://www.ncbi.nlm.nih.gov/genbank/statistics/), EMBL Nucleotide Sequence Data Library (https://www.ebi.ac.uk/history) and the DNA Data Bank of Japan (DDBJ) (https://www.ddbj.nig.ac.jp/about/index-e.html) (Figure I.1). On the other hand, biotechnology companies realized the potential of sequencing and several commercial sequencing machines started to be released with increasing accuracy, speed, and sequencing outputs, with Applied Biosystems dominating the market (Giani et al., 2020; Heather and Chain, 2016; Hutchison, 2007).

All these rapid advancements, further leveraged and/or funded by the Human Genome Project (HGP) initiative (https://www.genome.gov/human-genome-project/timeline), allowing for the constant increase in the number and complexity of genome sequencing throughout the 1990s and early 2000s solely based on Sanger sequencings, such as the first bacterial genomes (hence the first living organisms) of *Haemophilus influenzae* (Fleischmann et al., 1995) and *Mycoplasma genitalium* (Fraser et al., 1995); the first eukaryotic genome assembly, the yeast *Saccharomyces cerevisiae* (Goffeau et al., 1996); the bacteria *Escherichia coli* and *Bacillus subtilis* genomes (Blattner et al., 1997; Kunst et al., 1997); the first multicellular organism, the worm *Caenorhabditis elegans* (*C. elegans* Sequencing Consortium, 1998); the first plant *Arabidopsis thaliana* (The Arabidopsis Genome Initiative, 2000); and the fruit fly *Drosophila melanogaster* (Adams et al., 2000; Myers et al., 2000). The ultimate goal, the assembly of the human genome, was achieved in 2001 nearly 50 years after Watson and Crick's publication (Craig Venter et al., 2001; Lander et al., 2001).

3 – The Human genome sequence: the start of a new revolution

The first steps toward the HGP initiative were taken at the end of the 1980s and its completion took almost 15 years, costing ~$3,000 million (https://www.genome.gov/human-genome-project/Completion-FAQ). Although the two human genomes presented in 2001 were sequenced using automated Sanger sequencing, they leveraged the begging of the next sequencing revolution, i.e., “the Next Generation Sequencing” or “the genomic” era. Sanger sequencing was already being used routinely in many research groups throughout the world. However, large sequencing projects, such as the above-mentioned, were extremely demanding both in labour and cost, hence were only possible through massive collaborative engagements between major genetic groups. This is testified by the number of authors and institutions involved, with some publications even being authored by the consortiums, e.g., (*C. elegans* Sequencing Consortium, 1998; The Arabidopsis Genome Initiative, 2000). By then it was clear that new more efficient and affordable sequencing technologies were needed, which became the focus of several funding agencies and private biotechnology companies, e.g., the NHGRI started the “Revolutionary DNA Sequencing Technologies program – The $1000 Genome” to support grants for developing new sequencing approaches for affordable genomes sequencing (https://grants.nih.gov/grants/guide/rfa-files/RFA-HG-10-014.html). Alternative sequencing approaches (non-electrophoretic dependent) start to appear during the 1980s and 1990s (Giani et al., 2020; Heather and Chain, 2016; Hutchison, 2007; Shendure et al., 2017) (Figure I.1). In the late 1980s, Pål Nyrén and colleagues started exploring the potentiality of the then recently discovered measuring method of pyrophosphate synthesis based on a proportional production of light (luminescent) produced by luciferase (Nyrén and Lundin, 1985). Upon perfecting their approach, they eventually proposed a new sequencing approach, known as pyrosequencing, which is regarded as the first next-generation sequencing method (Nyrén, 1987; Ronaghi et al., 1998, 1996) (Figure I.1). Pyrosequencing possesses several advantages over dideoxy sequencing as it allows the use of natural nucleotides (instead of heavily modified), register sequencing in real-time (instead of a post-electrophoreses process) and most importantly multiplexing (instead of one tube/reaction) (Ronaghi et al., 1998). Pyrosequencing was later licensed to 454 Life Sciences (after being acquired by Roche). Around the same time another biotechnology company emerged, Solexa (later acquired by Illumina), founded by Balasubramanian and Klenerman. Solexa started exploring other massively parallel sequencing approaches that relied on fluorescent sequencing on polymerase colonies (bridge amplification), where tightly clustered individual molecules copies are generated over a surface from an immobilized template library (Adessi et al., 2000; Mayer et al., 1998; Mitra et al., 2003; Mitra and Church, 1999). Similar to Sanger sequencing, Solexa relied on fluorescent label chain-terminating nucleotides, with the difference that the chain-terminating nucleotides can be reversible, thus allowing for a subsequential nucleotide addition and reading over each cluster (Bennett, 2004; Bennett et al., 2005).

Not surprisingly, soon after the conclusion of HGP, 454 Life Science released the first ever commercially available NGS machine (Figure I.1), the GS 20 capable of producing 400–500 bp 99% accurate reads with an output of 25 Mbp/run at one-sixth the cost of other methods. Three years later Solexa releases its first machine (Genome Analyzer) (Figure I.1), which allowed for a higher throughput of 1 Gbp/run at the cost of shorter reads of 35 bp. The success of these early NGS approaches was demonstrated by resequencing previously obtained genomes, such as *Escherichia coli* (Shendure et al., 2005), *Mycoplasma genitalium* (Margulies et al., 2005), and the human genome (Bentley et al., 2008), triggered a chain reaction of emerging sequencing technologies that determined the entering in a new era of sequencing.

4 – Next Generation Sequencing (NGS)

Over the next decade, several new sequencing methods and machines appeared, benefiting from a series of developments in high-resolution imaging, microfabrication, and the exponential increase in computational power (Giani et al., 2020; Goodwin et al., 2016; Heather and Chain, 2016; Hutchison, 2007; Shendure et al., 2017) (Figure I.1). In general terms almost every NGS approach that initially emerged shared the same series of conceptual steps: library preparation, where DNA (native or amplified) is fragmented and/or size selected, followed by adapter ligation at the ends of the fragments; DNA amplification, where millions of individual reaction centres (e.g., beads, solid plates, or DNA nanoballs) with clonal copies of DNA template are created and; Sequencing where the library is loaded on a flow cell and massively parallel sequencing reactions are performed (Giani et al., 2020; Goodwin et al., 2016; Heather and Chain, 2016; Hutchison, 2007; Shendure et al., 2017).

Today, there are four main sequencing categories; three that include almost all short-read sequencing approaches, i.e., sequencing by synthesis (SBS), sequencing by ligation (SBL) and synthetic long-reads (SLR), and single-molecule sequencing (SMS) (Goodwin et al., 2016). Moreover, the sequencing outputs of these four sequencing categories can be divided into two types, short-read sequencing and long-read sequencing approaches. The latter type is represented by only two competing SMS strategies and is known as single-molecule long-read sequencing (SMLS) (Goodwin et al., 2016).

Sequencing by synthesis (SBS), the most successful sequencing strategy, encompasses all (but one) DNA-polymerase-dependent methods including both NGS methods (e.g., 454 GS pyrosequencing, Qiagen GeneReader, Illumina, and Ion Torrent) and Sanger sequencing (Goodwin et al., 2016; Heather and Chain, 2016) (Figure I.1). Although many of the SBS methods were promising at the beginning, as of today, only Illumina and Sanger sequencing maintain a significant role, with Illumina suppressing Sanger as the dominant sequencing provider worldwide. In fact, since the Solexa Genome Analyzer, Illumina has constantly released new machines offering constant improvement in accuracy, read length, prices, and throughput as well as diversification of library preparation methods with a variety of genomic applications, thus solidifying its place as the leader of the sequencing market (Goodwin et al., 2016). In 2017, Illumina launches NovaSeq 6000 (S4 flow cell) platforms capable of generating ~3000 Gb/run, and today the human genome can be sequenced for less than 500$, a 6-million-fold decrease in the cost of the HGP.

Sequencing by ligation (SBL) approaches rely on DNA ligase for the hybridization and ligation of ladled probes and anchored sequences (Goodwin et al., 2016). This includes two sequence strategies, Thermo Fisher Sequencing by Oligonucleotide Ligation and Detection (SOLiD) and BGI Complete Genomics, both with a very residual representation in sequencing projects today (Goodwin et al., 2016) (Figure I.1).

Although short-read sequencing approaches quickly demonstrated their potential, it also became clear that the small fragments generated were not ideal when working with complex genomes, with large sizes, high repetitive content, and structural variability (Goodwin et al., 2016; Sedlazeck et al., 2018). To overcome these caveats, two types of sequencing strategies have emerged, SLR and SMLS.

Synthetic long-reads (SLRs) still rely on short read sequencers and thus will output short reads, however, during library preparation long DNA fragments are tagged/reordered/barcoded to keep long-range bridging information within each short read. Among others, SLS includes the Illumina synthetic long-read sequencing and the 10X Genomics emulsion-based system (Figure I.1). At first, these strategies offered a very promising alternative to SMLS, especially because of the comparatively cost and higher base call accuracy (Goodwin et al., 2016; Sedlazeck et al., 2018). However, in the last decade, SMLS prices have continually dropped, their throughput increased, and the error rates decreased. Consequently, SLR plays a very residual role, with 10X Genomics discontinuing its whole genome sequencing branch.

Unlike short-read approaches, SMS does not produce a clonal cluster of DNA fragments nor requires stepwise dNTP chemical cycling. The first available SMS approach was the Helicos Genetic Analysis System (Helicos BioSciences), which quickly lost relevance as it produced very short reads (35 bp), at a high cost and slow pace (Giani et al., 2020; Harris et al., 2008) (Figure I.1). The true revolution emerged soon after with two SMLS distinct approaches: first, the Single Molecule Real-Time (SMRT) sequencing (Levene et al., 2003) commercialized by Pacific Biosciences (PacBio); and second, the Nanopore sequencing, a nearly 30 years hypothetical concept (Deamer et al., 2016) applied and commercialized by Oxford Nanopore Technologies (ONT) (Giani et al., 2020; Goodwin et al., 2016; Heather and Chain, 2016; Shendure et al., 2017) (Figure I.1). While the underlying sequencing process of these approaches is very divergent, both offer long reads, which have become a necessary resource for ensuring the contiguity of genome projects as well as RNA isoform determination. PacBio offers two distinct sequencing outputs: Continuous Long Read (CLR), the first strategy offered by the company, which produces an average read length of 20 kbp (but can generate reads longer than 50 kbp) and an error rate of around 13% (Goodwin et al., 2016). Consequently, projects need to complement this approach with lower error sequencing strategies for error correction; Circular Consensus Sequences (CCSs), introduced in 2019, offer a lower error rate (similar to Illumina short reads), while slightly reducing the read size (up to 20 kbp) and at higher per base sequencing cost (Wenger et al., 2019). The most recent PacBio sequencing machine, Sequel II, can generate 160 Gb per SMRT cell (Giani et al., 2020). Nanopore sequencing has a similar error rate to PacBio CLR, thus requiring third-party sequencing methods for correction. However, it offers unique advantages that make it a competing force, such as an Ultra-Long Read Sequencing Kit capable of routinely producing reads between 50kbp-4Mbp (Jain et al., 2018), including the largest DNA sequence to date, a 4.2Mbp read in 2021 (Figure I.1). Ultra-long reads have been fundamental in sequencing complex genomes (e.g., the 37 Gbp giant lungfish genome Meyer et al., 2021) and/or highly complex genomic regions (e.g., centromeric regions Jain et al., 2018). Nanopore also offers sequencing control and portability to the user by providing very low-cost sequencing devices (MinION), as well as ultra-high-throughput platforms (PromethION) capable of sequencing up to 14Tbp/72h/run matching the potential of the latest with Illumina machines (https://nanoporetech.com). Finally, given its low cost, rapid and easy use, nanopore has also revealed a high potential for metabarcoding and metagenomics studies (Egeter et al., 2022).

5 – Genomic revolution and the path to the Tree of Life (ToL)

Retrospectively, although many sequencing methods have emerged and perished during the last 20 years. The most significant advancement to come out of this continuous frantic competition for the “best sequencing technology” was the progressive democratization of sequencing power. Sanger sequencing democratized “Genetics” while NGS democratized “Genomics”. These advancements had a profound impact on the culture, range, and structuring of the field of genomics. Genomic scale studies are no longer restricted to large institutions and large consortiums. Importantly, this fundamental approach has expanded from exclusive “model” species. The “genome era” is progressively reaching the complete ToL (Figure I.2), although the ultimate goal, i.e., whole genome sequencing (WGS), is still far for some groups (Stephan et al., 2022). However, many other NGS applications are already at reach today, such as transcriptomics, metagenomics or the many available genomic partitioning strategies, e.g. Anchored Hybrid Enrichment (AHE), Restriction Site Associated DNA Sequencing (Rad-seq), Ultra Conversed Element (UCE) (Andrews et al., 2016; Lemmon and Lemmon, 2013; McGettigan, 2013; Quince et al., 2017). The broadening taxonomic scope of the application of genomics approaches has opened the way for comparative genomics and with it a new view of the evolutionary processes within the ToL, allowing to link DNA variation to diversification, adaptation, and survival at an unparalleled scale (Stephan et al., 2022). Not surprisingly, a diversification of fields has evolved within the genomic era, such as phylogenomics (Kapli et al., 2020), population genomics (Hohenlohe et al., 2021), adaptive genomics (Barghi et al., 2020), conservation genomics (Formenti et al., 2022), epigenomics (Mehrmohamadi et al., 2021), metagenomics (Quince et al., 2017) and mitochondrial genomics (DeSalle and Hadrys, 2017). However, despite the unprecedented and exciting proliferation of genomic resources available today, there are still many underrepresented groups within the ToL (Stephan et al., 2022) (Figure I.2). To level the biased representation of genomics resources (especially whole genome assemblies), over the last decades, countless global consortiums and initiatives directed towards non-model organisms have emerged (Table I.1), many of which are affiliated with the Earth Biogenome Project (Lewin et al., 2018)(https://www.earthbiogenome.org/affiliated-project-networks). These initiatives will be fundamental to generate a deeper understanding of non-model organism biology, identifying the evolutionary history and phenotypic expression of genomic features and using this information as a proxy to understand organism phylogeography, demography, adaptation patterns, and conservation traits (Arumugam et al., 2019; Dunn and Ryan, 2015; Lopez et al., 2019; Richards, 2015; Savolainen et al., 2013; Stephan et al., 2022).

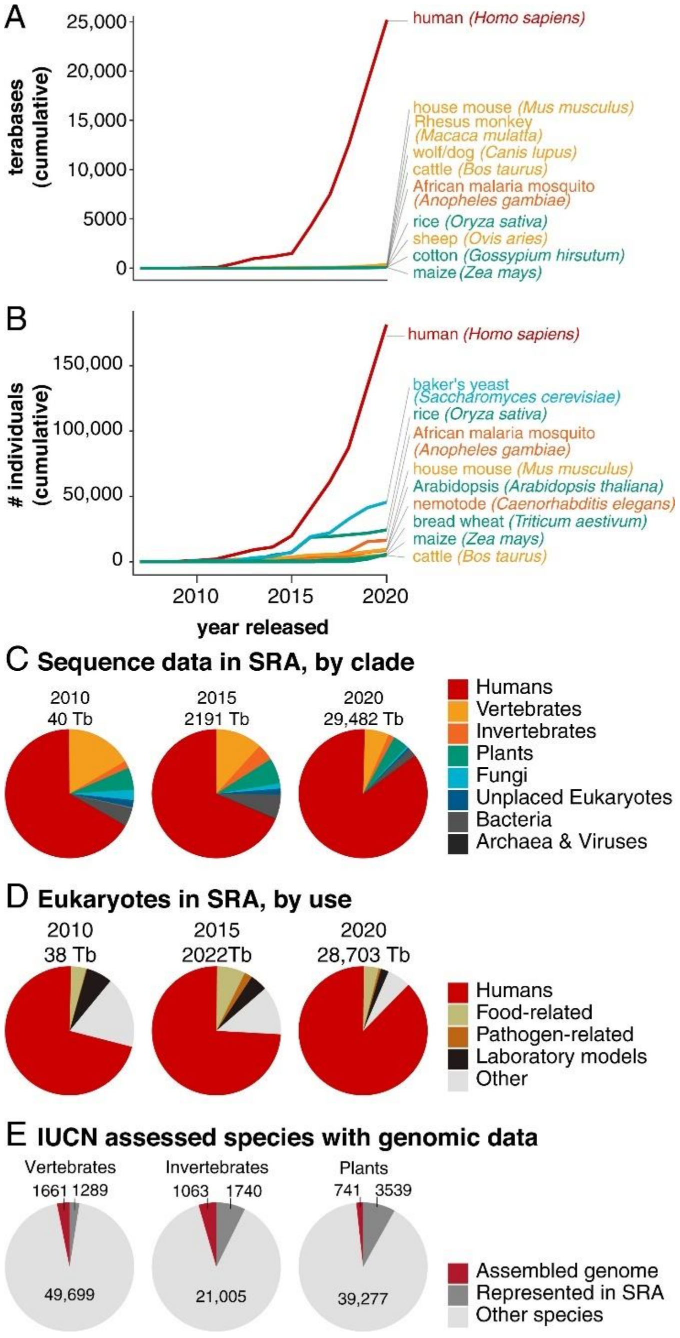
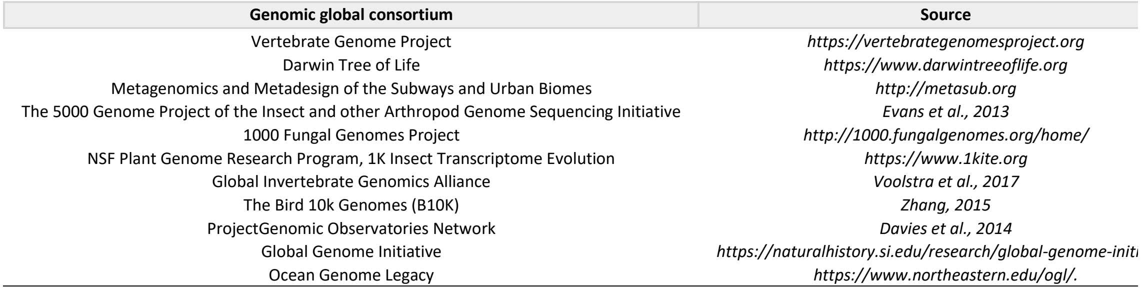


Figure I. 2 - Species diversity in the Sequence Read Archive (SRA). The amount of human data exceeds that of the next top 10 species, measured as (A) terabases and (B) individuals sequenced. (C) The human proportion increased between 2010 and 2020, and (D) the proportion from species without known commercial/medical relevance (“other”) dropped. (E) A tiny proportion of IUCN-recognized (80) species have a reference genome (red) or are otherwise represented in the SRA (dark grey). Retrieved November 14, 2020. Adapted from Stephan et al., 2022.

Table I. 1 - List of some global Consortiums aiming to increase genomics data on different organisms through the Tree of Life (ToL).



Genomic global consortium	Source
Vertebrate Genome Project	*https://vertebrategenomesproject.org*
Darwin Tree of Life	*https://www.darwintreeoflife.org*
Metagenomics and Metadesign of the Subways and Urban Biomes	*http://metasub.org*
The 5000 Genome Project of the Insect and other Arthropod Genome Sequencing Initiative	*Evans et al., 2013*
1000 Fungal Genomes Project	*http://1000.fungalgenomes.org/home/*
NSF Plant Genome Research Program, 1K Insect Transcriptome Evolution	*https://www.1kite.org*
Global Invertebrate Genomics Alliance	*Voolstra et al., 2017*
The Bird 10k Genomes (B10K)	*Zhang, 2015*
ProjectGenomic Observatories Network	*Davies et al., 2014*
Global Genome Initiative	*https://naturalhistory.si.edu/research/global-genome-initi*
Ocean Genome Legacy	*https://www.northeastern.edu/ogl/.*

Despite the increasing attention toward new groups of organisms, there is still a noticeable imbalance of genomic data produced from non-model species. Within the Metazoa, vertebrates receive much more attention than invertebrates, and except for arthropods and nematodes, most invertebrates have comparatively fewer data, with some groups without any representation presently (David et al., 2019; Dunn and Ryan, 2015; Hotaling et al., 2021; Lopez et al., 2019, 2014; Stephan et al., 2022; Voolstra et al., 2017). The advancements in sequencing offer the opportunity to fill these gaps, not only for the sake of cataloguing the genomic diversity but to create the means for comparative analysis. Evolution represents a ~4,000 million years experimental trial that incorporates all life on Earth. Therefore, a comprehensive sampling of diversity (within and between groups) is fundamental to understanding how genomic variation moulded organism functional traits and adaptations, as well as their synergetic interactions with the ecosystems (Paez et al., 2022; Paps, 2018; Stephan et al., 2022). Genomes, by coding the information that sustains cellular machinery, may be considered the most informative tool of a species’ biology. However, a reference genome by itself is not always enough to answer all the countless and highly complex aspects of biology (such as development, differentiation, and biotic and environmental interactions). Consequently, apart from a comprehensive sampling of diversity, a multi-informative application of genomic resources is also essential to disentangle the mechanism underlying many evolutionary processes.

The genomic revolution has democratized sequencing, opening the way for genomic exploration of the many branches of life. Therefore, it is nowadays essential to start sampling the underrepresented organisms and balance the availability of genomic data. This will not only help to understand the biology of those often overlooked organisms but also help to untangle and understand complex evolutionary processes among many other taxa.

Among the many underrepresented taxa, Mollusca occupies a prominent position. This group is immensely diverse with more than 200,000 species, with a significant ecosystem relevance and fundamental cultural legacy. Yet, the scarce availability of genomics resources to study them reflects little on the phylum diversity. Moreover, many of the resources are biased towards specific groups, especially highly abundant and/or economically relevant species. Consequently, increasing the availability of genomic resources and diversifying its targets within molluscs is fundamental to fully place the phylum in the genomic era.

6 – Phylum Mollusca

Mollusca is a highly diverse Metazoa phylum that comprises many familiar organisms, such as mussels, scallops, cockles, limpets, snails, slugs, octopuses, and squids (Brusca et al., 2003; Haszprunar and Wanninger, 2012; Ponder et al., 2019; Winson F and David R, 2008). Commonly known for their beautifully diverse shell-bearing species, this widely diverse group of organisms includes an astonishing availability of fossil records dating back to the early Cambrian Period (~543 million years ago [Mya]) (Ponder et al., 2019; Winson F and David R, 2008). Throughout their long evolutionary history, molluscs developed a highly diverse phenotypic repertoire (physiological, behavioural and ecological) resulting in outstanding adaptive success. Molluscs have conquered almost all types of habitats, from the deep high-pressure oceans to the high altitude of mountain tops, from the high saline oceans (~35‰ salinity) to oligosaline and oligotrophic mountain streams (< 0.5‰) or the contrasting environments of tropical rain forests and arid deserts (Haszprunar and Wanninger, 2012; Ponder et al., 2019; Winson F and David R, 2008). They occupy multiple levels of trophic webs, from grazers, decomposers, predators and suspension-feeders to parasites, assuming many shapes and sizes, with highly specialized group-specific adaptations and structures, including biomineralized protective shells, scraping radula, symbiotic organs, hematophagy, venom production, memory and cognitive learning (Belcaid et al., 2019; Boyle and Rodhouse, 2008; Haszprunar and Wanninger, 2012; Marin, 2020; Modica et al., 2015; Ponder et al., 2019; Sigwart and Sumner-Rooney, 2015; Winson F and David R, 2008; Yarra et al., 2021).

The diversity and "cosmopolitan" biology of the phylum have resulted in long-lasting interaction with humans, resulting in several direct and indirect impacts on human lifestyle and well-being, either by providing important services such as food resources (mussels, clams, abalones, limpets, whelks, land, octopus, and squids), and cultural objects (shells, opercula, pearls, and mother-of-pearl and natural pigmentation), or by negatively affecting humans both as pests (e.g., *Dreissena polymorpha* or *Pomacea canaliculata*) and as hosts for human and cattle parasites (e.g., *Schistosomiasis*, *Eosinophilic meningitis*), (Adema et al., 2017; Boyle and Rodhouse, 2008; From et al., 2000; Furuhashi et al., 2009; Haszprunar and Wanninger, 2012; Karatayev et al., 2015; Ponder et al., 2019; Strack, 2015; Strayer, 2009; Takeuchi, 2017; Winson F and David R, 2008). From a scientific point of view, molluscs have also served humans as biological markers and models, such as for monitoring and studying pollution, biomineralization, ocean acidification, climate changes adaptation, cell biology, neurobiology, and pharmacology (Bailey et al., 1983; Gazeau et al., 2013; Glanzman, 2009; Li et al., 2021; Powell et al., 2018; Sigwart and Sumner-Rooney, 2015; Suleria et al., 2017; Talmage and Gobler, 2010; Walters and Moroz, 2009; Yarra et al., 2021). In almost every environment, molluscs, along with other invertebrates, generally dominate in terms of species richness, abundance, and biomass (Cardoso et al., 2011; Cuttelod et al., 2011; Kay, 1995; Winson F and David R, 2008). Consequently, many species are essential to the ecosystems, working as keystone species and/or ecosystem engineers and providing a wide range of ecological functions and services (Atkinson et al., 2013; Spooner et al., 2013; Strong et al., 2007; van der Schatte Olivier et al., 2020; Vaughn and Hakenkamp, 2001; Vaughn and Hoellein, 2018). Humans have had a major impact on molluscs' diversity worldwide, particularly in nonmarine environments, where overexploitation, habitat degradation and loss, invasive species, and climate change have caused accentuated population declines and several extinctions (Böhm et al., 2021; Cuttelod et al., 2011; Dudgeon et al., 2006; Graf, 2013; Lopes-Lima et al., 2018b; Lydeard et al., 2006). There are more extinctions assessed within nonmarine molluscs than in all tetrapod vertebrates combined (Lydeard et al., 2006).

Mollusca is the second most speciose animal phylum (only second to Arthropoda), with an estimated 200,000 species, of which less than half have been described (Brusca et al., 2003). For comparison, the Chordata (probably the best-studied Metazoa group) has only a third of the species (i.e., ~ 60.000) (Brusca et al., 2003). Nowadays, eight (often disputed) classes are described within the phylum. The more diverse are Gastropoda (snails, limpets, slugs, whelks) and Bivalvia (mussels, clams, scallops, shipworms and oysters), including 96% of the species. The remaining taxa are distributed by the less represented classes Cephalopoda (squid, octopuses, cuttlefish and nautilus), Polyplacophora (chitons), Scaphopoda (tusk shells), Monoplacophora (deep-sea limpets) and the Aplacophora group (spicule worms) with two classes, the Solenogastres (or Neomeniomorpha) (solenogasters) and the Caudofoveata (or Chaetodermomorpha) (caudofoveates) (Brusca et al., 2003; Haszprunar and Wanninger, 2012; Ponder et al., 2019; Winson F and David R, 2008) (Figure I.3).

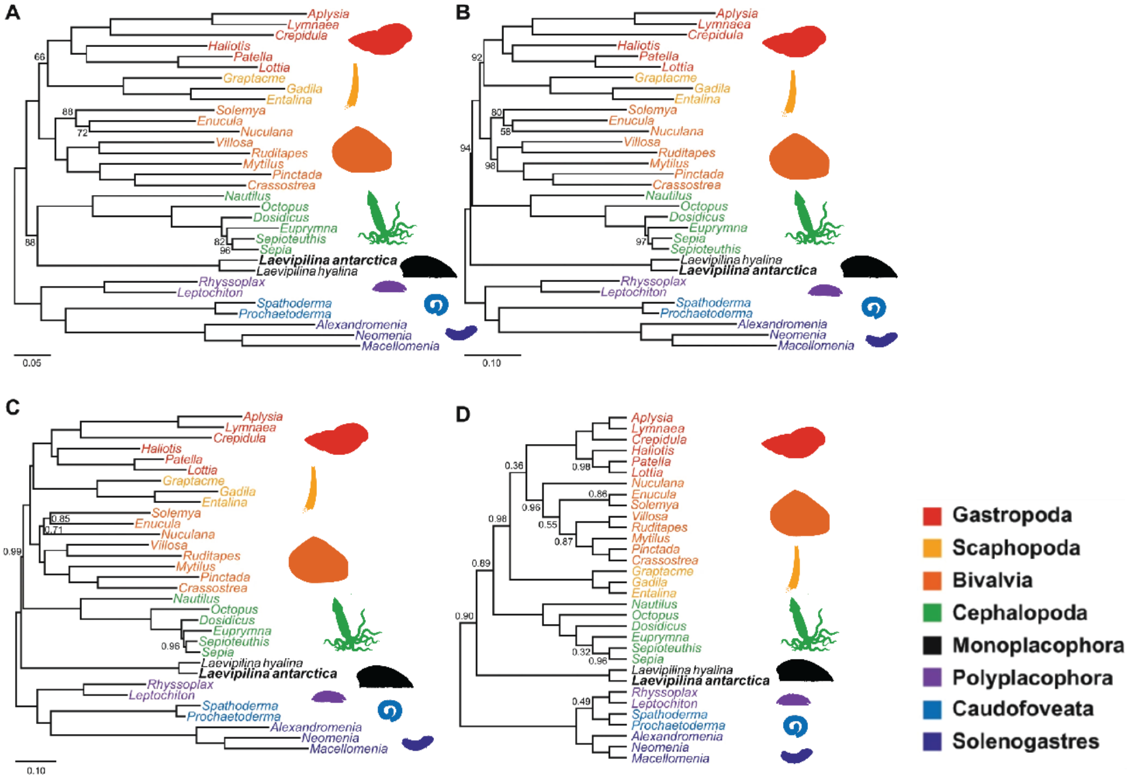


Figure I. 3 - Phylogenetic inference of phylum Mollusca (outgroup taxa not shown). (A) RAxML Maximum Likelihood (ML) tree with bootstrap support values below 100 shown. (B) IQ-TREE ML tree with bootstrap support values below 100 shown. (C) PhyloBayes Bayesian Inference (BI) tree with posterior probabilities below 1.0 shown. (D) ASTRAL tree with local posterior probabilities below 1.0 shown. Adapted from Kocot et al., 2020.

Despite their significant representation of Metazoa diversity, the study of molluscs is often neglected. Even the phylogenetic relationships among the main classes have been constantly disputed and reassessed during the last decades. Only recently, helped by phylogenomic approaches, consistent monophyletic classes started to be recovered (Kocot et al., 2020, 2011; Smith et al., 2011; Vinther et al., 2012; Wanninger and Wollesen, 2019) (Figure I.3). Moreover, these studies often contradicted accepted morphocladistic hypotheses. For example, phylogenomic studies unambiguously contradict the Testaria hypothesis (worm-like Aplacophora as a paraphyletic basal group of Mollusca) and instead unambiguously support a basal dichotomy that splits Mollusca into, the Aculifera (including the Polyplacophora and the reciprocally monophyletic Aplacophora) and the Conchifera (including the Monoplacophora, Cephalopoda, Scaphopoda, Gastropoda, and Bivalvia) (Kocot et al., 2020, 2011; Smith et al., 2011; Vinther et al., 2012; Wanninger and Wollesen, 2019) (Figure I.3). However, relationships within Conchifera are still controversial, with conflicting results emerging every often over the last few years, highlighting the importance of increasing the genomic resources within Mollusca, particularly in less sampled groups (Kocot et al., 2020, 2011; Smith et al., 2011; Vinther et al., 2012; Wanninger and Wollesen, 2019) (Figure I.3).

Increasing the availability of genomic resources within Mollusca will have a major impact, not only in phylogenetic studies but in many other fields, and ultimately benefit a wide range of ecological, economic and scientific purposes, such as: improving the global-scale market of molluscan fisheries and aquaculture, by helping to understand adaptation to changing conditions, immunological response to diseases and threats, as well as to develop efficient management of breeding programmes to increase productivity and value (Agriculture Organization of the United Nations. Fisheries Department, 2000; Boyle and Rodhouse, 2008; Clark et al., 2020; Guo, 2009; Houston et al., 2020; Mun et al., 2017; Murgarella et al., 2016; Powell et al., 2018; Takeuchi, 2017); generate new insights in medical and pharmaceutical applications, such as in neuroscience by untangling the complex cephalopods neurologic systems (Albertin et al., 2015; Boyle and Rodhouse, 2008; Kim et al., 2018), in cancer research by exploring underlying genetic processes of abalones tumour-suppressing features (Nam et al., 2017; Suleria et al., 2017), in drug research by exploring the venom specific peptide genes present in some species (Andreson et al., 2019; Barghi et al., 2016), human parasites research, by exploring the molecular mechanisms of molluscs that act as intermediary host (e.g. developments of molluscicides and gene drivers) (Adema et al., 2017; Liu et al., 2018; Sun et al., 2019); surveys and control of invasive species by studying the molecular mechanism behind ecological plasticity (Calcino et al., 2019; Liu et al., 2018; McCartney et al., 2022; Uliano-Silva et al., 2018); understanding biomineralization and pearl formation (Aguilera et al., 2017; Clark et al., 2020; Du et al., 2017; Kocot et al., 2016a; Marin, 2020; Takeuchi et al., 2012; Yarra et al., 2021); explore the molecular mechanism that control many of taxa-specific novelties such as: doubly uniparental inheritance of mitochondrial DNA in many bivalves (Breton et al., 2018; Zouros, 2013), kleptoplasty of the sacoglossan sea slugs (Cai et al., 2019), symbiotic relationships with microorganisms with specialized symbiotic organs in bobtailed squid (Belcaid et al., 2019); and comparative analysis with other taxa to study higher lever relationships and evolutionary processes within Metazoan (Kocot et al., 2020; Regan et al., 2021; Roberts and Kocot, 2021; Simakov et al., 2012; Sun et al., 2019; Takeuchi et al., 2016; S. Wang et al., 2017).

Numerous relevant questions substantiate the importance of bringing molluscs to the genomic era. Although many studies are paving the way for this new era, there is still a long road ahead. Increasing resources will be fundamental to addressing long-lasting questions of diversity, evolution, and molecular signatures of phenotypic adaptations within the group to explore the many known and unknown features of this fascinating and widely diverse group of organisms.

7 – Biology and conservation of freshwater mussels (Bivalvia: Unionida)

Within Molluscs, Bivalvia is the second most diverse group, with more than 20,000 species divided into two main groups, Protobranchia (exclusively marine species) and Autobranchia (marine and freshwater species) (Brusca et al., 2003; Ponder et al., 2019). The vast majority of commonly known bivalve species are comprised within Autobranchia, further divided into Pteriomorphia (e.g., mussels, scallops, and oysters) and Heteroconchia (e.g., freshwater mussels, shipworms, clams) (Ponder et al., 2019). Although bivalves have a marine origin, several independent lineages of freshwater bivalves have emerged throughout their evolutionary history, resulting from at least 11 distinct events (Bieler et al., 2014; Calcino et al., 2019; Combosch et al., 2017; Graf, 2013). These lineages differ in the extent of radiation and expansion from their original marine habitats, with 97% of the species belonging to either family Cyrenidae (order Venerida) or orders Sphaeriida and Unionida (Bieler et al., 2014; Calcino et al., 2019; Combosch et al., 2017; Graf, 2013; Lemer et al., 2019). The latter two represent the only bivalve orders of strictly freshwater organisms, with Unionida showing the highest species diversity of all freshwater bivalves with nearly 1,000 species (Graf and Cummings, 2007, 2022; Haag, 2012; Strayer, 2008; J. D. Williams et al., 2017).

Order Unionida is a >200 Mya monophyletic group often referred to as Freshwater Mussels (hereafter referred to as FMs), freshwater clams, or naiads (Graf and Cummings, 2007; Haag, 2012; Strayer, 2008a). Freshwater mussels are found in freshwater ecosystems on all continents (except Antarctica) and have acquired a series of remarkable adaptations to survive under constant water flow (Haag, 2012; Strayer, 2008). This adaptative repertoire includes internal fertilization with 'parental care' and, the most remarkable, a highly specialized larvae stage called glochidia, which acts as an obligatory parasite of fish (and occasionally other vertebrates) and ensures dispersion and nutrition during metamorphosis (Barnhart et al., 2015; Modesto et al., 2018). Furthermore, their ability to maintain osmolarity in freshwater is likely due to FMs’ specific gene expansion of transmembrane passive water transporters (Calcino et al., 2019). All these adaptations have likely played a fundamental role in the FMs' successful dispersion and colonization of the world’s freshwater habitats. Additionally, some FMs share with several other groups of bivalves an interesting biological feature, the unusual mitochondrial DNA inheritance system, called Doubly Uniparental Inheritance (DUI) (Breton et al., 2007, 2011b) (Figure I.4). Under DUI, male and female individuals inherit mitochondrial DNA from their mothers (F-type), while males also inherit a male-specific mtDNA lineage from their fathers (M-type). Moreover, M and F-type mitogenomes of unionids have two reciprocally unique putative genes with unknown homology or function (ORFan genes), referred to as M-*orf* and F-*orf*, respectively (Breton et al., 2011b; Guerra et al., 2019) (Figure I.4). Alternatively, in strictly hermaphroditic species the M-type mitogenome is absent and the only retained mitogenome lineage is a derived F-type lineage, that possesses a modified ORFan referred to as H-*orf* (Breton et al., 2011b, 2018) (Figure I.4). This strong correlation between DUI and the sexual reproductive mechanism of bivalves has raised the hypothesis that DUI might play a role in sex determination and/or sexual development (Breton et al., 2011b, 2018; Zouros, 2013), although no solid evidence of such role has been generated so far.

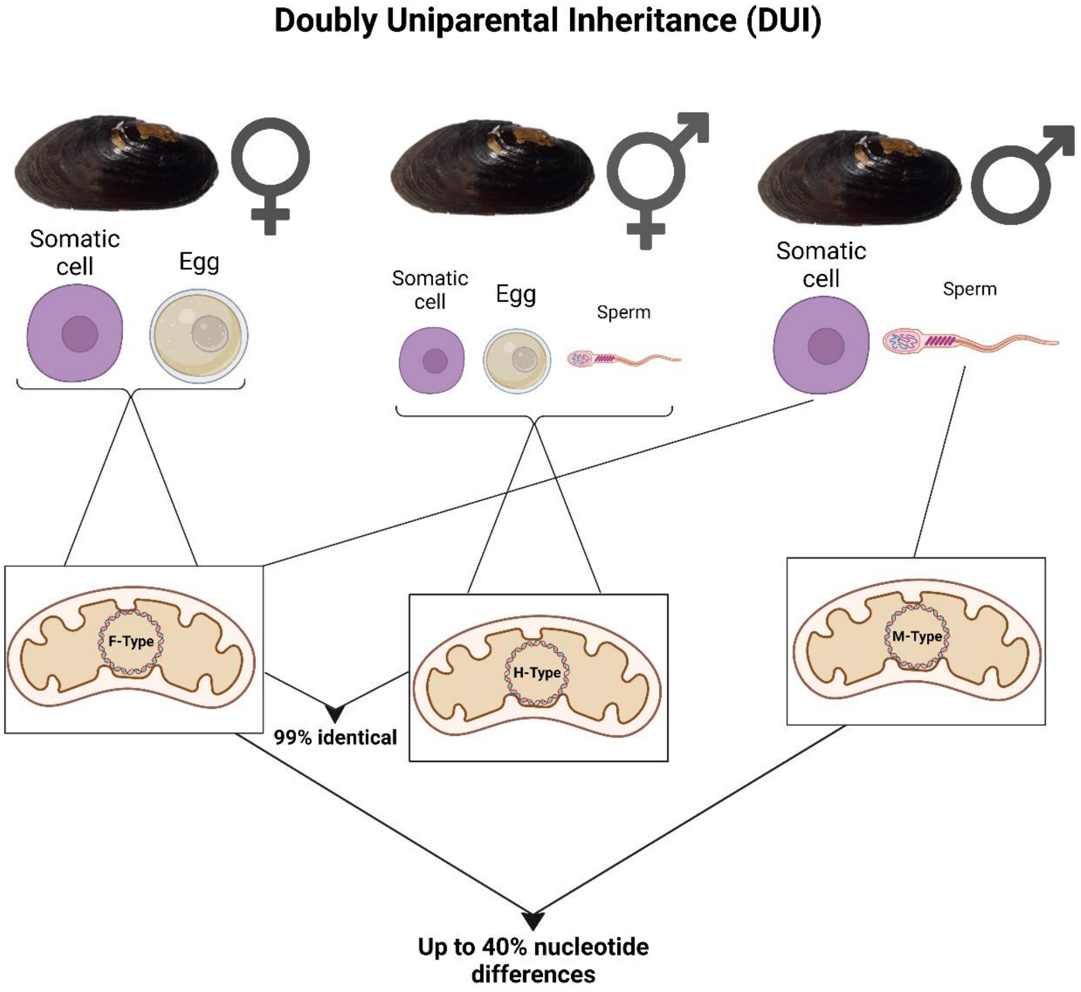
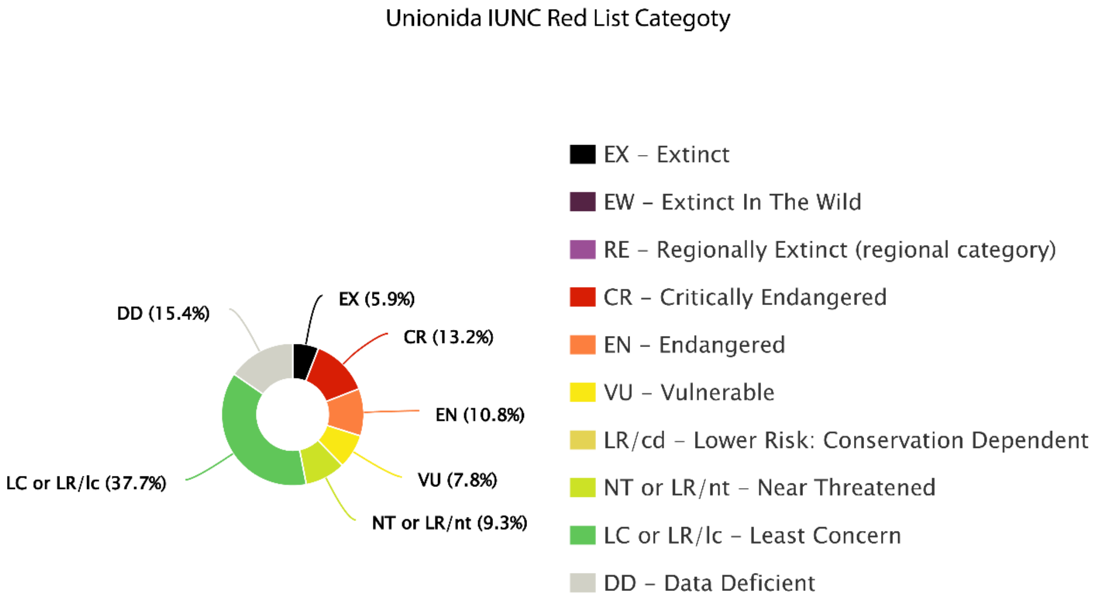


Figure I. 4 - Graphical representations of the mitochondrial DNA inheritance system called Doubly Uniparental Inheritance (DUI).

Freshwater mussel species have often a dominant role in the benthic biomass of freshwater systems, where they take part in key ecological functions, including biofiltration, nutrient and energy cycling, habitat structuring, and sediment mixing (also known as bioturbation), as well as provide valuable ecosystem services, including increased water clarity, raw material sources (pearls and shell) and as a food source in some human cultures (Howard and Cuffey, 2006; Strayer, 2008; Vaughn, 2017; Vaughn and Hakenkamp, 2001). Despite the over-mentioned roles, similarly to other freshwater taxa, FMs have experienced massive defaunation on a global scale, being one of the most imperilled Metazoan groups in the world (Ferreira-Rodríguez et al., 2019; Lopes-Lima et al., 2018b; Lopes-Lima et al., 2021; Lydeard et al., 2006). At a global scale, the major threats to FM are habitat loss, degradation and fragmentation, loss of the glochidia hosts, introduction of non-native species, climate change and overexploitation, with other localized known and unknown factors contributing to the declines (Geist, 2011; Lopes-

Lima et al., 2018b; Modesto et al., 2018; Strayer and Dudgeon, 2010). Based on the IUCN Red List, 41.1% of the 539 accessed FMs species are either Near Threatened or Threatened (Vulnerable/Endangered/Critically Endangered), 5.9% are already extinct (the highest of any Metazoan) and 15.4% are data deficient (IUCN 2022) (Díaz et al., 2019; Lopes-Lima et al., 2021) (Figure I.5). Concerningly, these values regard little more than half of the number of estimated FMs species and are uneven with some regions being better covered by the IUCN Red List than others (Ferreira-Rodríguez et al., 2019; Graf and Cummings, 2022; Lopes-Lima et al., 2014a, 2021). This overall concerning scenario has leveraged research and conservation action devoted to FMs, which, however, have also been largely focused on a very small number of charismatic species in Europe and North America (Ferreira-Rodríguez et al., 2019; Lopes-Lima et al., 2014a).



https://www.iucnredlist.org/search/stats

Figure I. 5 - Freshwater Mussels global conservation categories accessed by the Red List of the IUCN.

8 – The Family Margaritiferidae

Currently, six families are recognized within Unionida based both on morphological characters and molecular markers (Bogan, 2008; Graf and Cummings, 2007; Pfeiffer et al., 2019). Two families have a wide northern hemisphere distribution, i.e., the Unionidae (the most speciose family with ~600 species) and the Margaritiferidae. Another two families can be found mainly in the southern hemisphere, i.e., the Mycetopodidae which occurs in South America and the Iridinidae which occurs in Africa. Of the remaining two families, one is restricted to Africa, i.e., the Etheriidae, and the other is found in both South America and Australia, i.e., the Hyriidae (Bogan, 2008; Graf and Cummings, 2007; Pfeiffer et al., 2019). Of these six families, Margaritiferidae (hereafter referred to as margaritiferids) has the highest percentage of species at risk of extinction, 66.7% (IUCN 2022), and includes one of the 100 most threatened species on earth, *Pseudunio marocanus* (Pallary, 1928) (Baillie and Butcher, 2012). Currently, 15 species, distributed throughout many freshwater systems of the Holarctic region, are recognized within the family Margaritiferidae (Lopes-Lima et al., 2018a). Although widely distributed as a whole, Margaritiferidae species have a sparse and intermittent distribution with almost no overlap (Bolotov et al., 2016; Lopes-Lima et al., 2018a; Takeuchi et al., 2015). The family includes one of the most emblematic and widespread Unionida species, the freshwater pearl mussel *Margaritifera margaritifera* (Linnaeus, 1758) (Figure I.6), which due to its centuries-long cultural relevancy in Europe for pearl harvesting has led to margaritiferids colloquially being referred as pearl mussels (Bauer, 2001; Strack, 2015). In general terms, it is assumed that margaritiferids species reach large sizes (typically around 100 mm in length), have a large life expectancy (from 50 to up 200 years), are confined to streams and rivers, have a muscular and extensible foot (for digging and anchoring) (Figure I.6) and show a highly restrictive use of host fishes (often highly vagile or migratory fishes such as salmonids) (Bauer, 2001; Benaissa et al., 2022; Haag and Rypel, 2011; Johnson and Brown, 2011; Kondo and Kobayashi, 2005; Lopes-Lima et al., 2018a, 2017e; Parmalee and Bogan, 1998; Stone et al., 2011; Vikhrev et al., 2019).

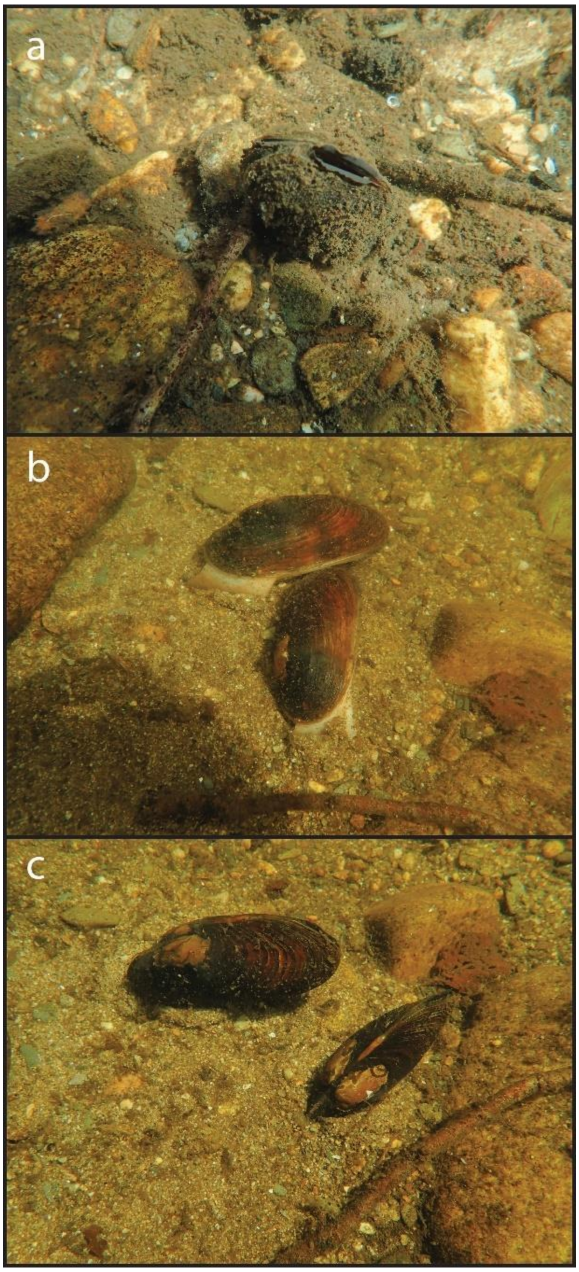


Figure I. 6 - a-c) Freshwater pearl mussel *Margaritifera margaritifera* in its natural environment. Pictures in panels b and c were taken within 10min apart and show two specimens burrowing themselves in the sediment using their muscular foot. Photos by André Gomes-dos-Santos (August 2022).

However, like other Unionida species, margaritiferids from North America and Europe received much more attention than the remaining species for which ecology, life history, and even distribution are still not well known (Lopes-Lima et al., 2018a). Even the systematics of the family has been unstable, with constant species, genus, and family reassignments (Lopes-Lima et al., 2018a). This is largely due to highly variable or homoplastic morphological characters, with molecular phylogenetics representing a fundamental complement to producing improved knowledge about the classification of the Margaritiferidae (Araujo et al., 2017; Bolotov et al., 2016; X.C. Huang et al., 2018; Lopes-Lima et al., 2018a). Added to the generalized lack of information about the biology, ecology, and systematics of the group, the availability of genomic resources within Margaritiferidae is equally limited (but see Bertucci et al., 2017; Breton et al., 2011b; Farrington et al., 2020; Garrison et al., 2021; Guerra et al., 2017, 2019; X.C. Huang et al., 2018; Lopes-Lima et al., 2018a, 2017a; S. Yang et al., 2014). This is not only true for margaritiferids but for Unionida as well, for which the only available genomic resources are three whole genomes (0.3% of species) (Renaut et al., 2018; Rogers et al., 2021; Smith, 2021), less than 20 transcriptomes (2% of species) (Bertucci et al., 2017; Capt et al., 2018, 2019; Chen et al., 2019; Cornman et al., 2014; Ganser et al., 2015; D. Huang et al., 2019; Luo et al., 2014; Patnaik et al., 2016; Robertson et al., 2017; Roznere et al., 2018; R. Wang et al., 2015; X. Wang et al., 2017; Q. Yang et al., 2021), a couple of RAD-seq studies (Farrington et al., 2020; Garrison et al., 2021) and three target-capturing Anchored Hybrid Enrichment (AHE) studies (Pfeiffer et al., 2021, 2019; Smith et al., 2020).

Although many of the biological and ecological characteristics of the Unionida have long captivated scientists worldwide, the fact is that the molecular mechanisms underlying the regulation and functioning of many of these features are poorly studied and practically unknown. Thus, increasing the availability and range of genomic resources is critical to improving our knowledge of such mechanisms, which will have applications in multiple fields. This knowledge will provide the depiction of genetic features, identification of genomic novelties as well as a comprehensive and accurate framework enhancing the characterization of genetic variation, population structure and dynamics, selective pressures, and adaptative traits, fundamental to launch both basic and applied research and conservation efforts on this emblematic group of organisms.

9 – Objectives and thesis structure

The overall goal of this thesis is to advance the study of margaritiferids' biology and evolution by applying several next-generation sequencing approaches to obtain a series of novel mitogenomic, transcriptomic and genomic resources with practical applications in several emerging fields, such as phylogenomics, population genomics, conservation genomics, and adaptative genomics.

The specific goals of this thesis are:

1. To provide a review of the history and availability of genomics resources of the second most diverse phylum of Metazoans, i.e., the Mollusca.

2. To provide a review of the current knowledge on variation in architecture, molecular functioning, and intergenerational transmission of molluscan mitochondrial genomes.

3. To provide the whole mitogenomes of male, female, and hermaphroditic specimens of the freshwater pearl mussel *Margaritifera margaritifera;* to determine and compare the gene order and content of those mitogenomes; to produce phylogenetic analyses using all available F-type and M-type mitogenomes of the Margaritiferidae family.

4. To generate the first family-wide set of Anchor Hybrid Enrichment loci (AHE) for Margaritiferidae, as well as produce several new whole mitogenomes assemblies within the family. Both resources will serve as complementary tools for phylogenomic studies, not only within the family but also within the Unionida order.

5. To provide the functional annotation of the 813 AHE targeted regions recently developed for order Unionida. This well-curated functional catalogue represents a complementary tool for scrutinizing phylogenetic inferences while opening the way for future applications of this set of target sequences.

6. To develop and provide a new pipeline for *de novo* assembly of the targeted AHE probe regions using whole genome re-sequencing outputs. This tool allows us to easily combine the AHE target sequencing approach with the rapidly emerging whole genome sequencing outputs.

7. To generate the novel transcriptomes for European freshwater mussels with conservation concern: *Margaritifera margaritifera*, *Unio crassus*, *Unio pictorum*, *Unio mancus* and *Unio delphinus*. These transcriptomes represent a valuable resource to study these organisms’ genetic repertoire and thus serve as baseline tools to search for optimum environmental conditions and adaptation to stressful conditions.

8. To generate the first reference genome from a Margaritiferidae species, the freshwater pearl mussel *Margaritifera margaritifera* (Linnaeus, 1758). This genome, produced using Illumina paired-end and mate-pair short-read sequencing, represents an essential resource for the characterization of genetic features and identification of genomic novelties, such as single genes or gene families, genomic pathways and single-nucleotide polymorphism. Such findings will help to understand how genomic variation

shaped diversity and functions, which ultimately will allow the unravelling of the molecular mechanisms that govern many of the species’ fascinating biological features.

9. To provide an improved reference genome for freshwater pearl mussel *Margaritifera margaritifera* (Linnaeus, 1758). These results, accomplished by combining PacBio CLR long reads and Illumina paired-end short-read sequencing, translate into a highly enhanced genome, both in contiguity and completeness, that will facilitate future genomic studies.

This thesis is organised into four chapters:

Chapter 1 provides a general introduction to the topics of the thesis. Chapter 2 focuses on the phylum Mollusca and includes two scientific papers published in international peer-reviewed journals (Papers 1 and 2). Chapter 3 focuses on freshwater mussels of order Unionida with particular attention to the family Margaritiferidae and includes three scientific papers published/submitted in international peer-reviewed journals (Papers 3-5) and a genome report that complements the results of Paper 6 (Manuscript 1). Finally, Chapter 4 provides a general discussion that reflects on the overall outputs of this thesis, as well as future research applications.

Paper 1 - Gomes-dos-Santos, A., Lopes-Lima, M., Castro, L.F.C., Froufe, E., 2020. Molluscan genomics: the road so far and the way forward. Hydrobiologia 847, 1705–1726. https://doi.org/10.1007/s10750-019-04111-1

Paper 2 - Ghiselli, F., Gomes-Dos-Santos, A., Adema, C.M., Lopes-Lima, M., Sharbrough, J., Boore, J.L., 2021. Molluscan mitochondrial genomes break the rules. Philosophical Transactions of the Royal Society B: Biological Sciences B376,20200159-20200159. https://doi.org/10.1098/rstb.2020.0159

Paper 3 - Gomes-dos-Santos, A., Froufe, E., Amaro, R., Ondina, P., Breton, S., Guerra, D., Aldridge, D.C., Bolotov, I.N., Vikhrev, I. v., Gan, H.M., Gonçalves, D. V., Bogan, A.E., Sousa, R., Stewart, D., Teixeira, A., Varandas, S., Zanatta, D., Lopes-Lima, M., 2019. The male and female complete mitochondrial genomes of the threatened freshwater pearl mussel *Margaritifera margaritifera* (Linnaeus, 1758) (Bivalvia: Margaritiferidae). Mitochondrial DNA B Resour 4, 1417–1420. https://doi.org/10.1080/23802359.2019.1598794

Paper 4 - Gomes-dos-Santos, A., Froufe, E., Pfeiffer, J., Smith, C., Machado, A., Castro, L.F.C., Do, V., Hattori, A., Garrison, N., Whelan, N., others, 2022. A novel assembly pipeline and functional annotations for targeted sequencing: A case study on the globally threatened Margaritiferidae (Bivalvia: Unionida). Authorea Preprints. https://doi.org/10.22541/au.166799900.04572038/v1

Paper 5 - Gomes-dos-Santos, A., Machado, A.M., Castro, L.F.C., Prié, V., Teixeira, A., Lopes-Lima, M., Froufe, E., 2022. The gill transcriptome of threatened European freshwater mussels. Scientific Data 9, 494, 1–10. https://doi.org/10.1038/s41597-022-01613-x

Paper 6 - Gomes-dos-Santos, A., Lopes-Lima, M., Machado, A.M., Marcos Ramos, A., Usié, A., Bolotov, I.N., Vikhrev, I. v, Breton, S., Castro, L.F.C., da Fonseca, R.R., Geist, J., Österling, M.E., Prié, V., Teixeira, A., Gan, H.M., Simakov, O., Froufe, E., 2021. The Crown Pearl: a draft genome assembly of the European freshwater pearl mussel *Margaritifera margaritifera* (Linnaeus, 1758). DNA Research 28:2. https://doi.org/10.1093/dnares/dsab002

Chapter 2 – The study of Mollusca in the Genomic era

1.1. Paper 1 – Molluscan genomics: the road so far and the way forward

Gomes-dos-Santos, A., Lopes-Lima, M., Castro, L.F.C., Froufe, E., 2020. Molluscan genomics: the road so far and the way forward. Hydrobiologia 847, 1705–1726. https://doi.org/10.1007/s10750-019-04111-1

Molluscan Genomics: The road so far and the way forward

André Gomes-dos-Santos[1,2]*, Manuel Lopes-Lima[1,3,4]*, L. Filipe C. Castro[1,2], Elsa Froufe[1]

[1] CIIMAR/CIMAR — Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Terminal de Cruzeiros do Porto de Leixões, Avenida General Norton de Matos, S/N, P 4450-208 Matosinhos, Portugal; [2] Department of Biology, Faculty of Sciences, University of Porto, Rua do Campo Alegre 1021/1055, Porto, Portugal; [3] CIBIO/InBIO - Research Center in Biodiversity and Genetic Resources, Universidade do Porto, Campus Agrário de Vairão, Rua Padre Armando Quintas, 4485-661 Vairão, Portugal; [4] IUCN SSC Mollusc Specialist Group, c/o IUCN, David Attenborough Building, Pembroke St., Cambridge, England

* Corresponding authors.

DOI: https://doi.org/10.1007/s10750-019-04111-1

Abstract

Mollusca is the second most species-rich phylum within the metazoans, displaying critical economic, ecological and scientific importance. Yet, they are still largely underrepresented with respect to genomic resources. The emergence of next-generation sequencing technologies has revolutionized deep-scale genomic characterization of non-model organisms and molluscs are slowly entering this transformative era. Here, we provide an historical contextualization of the Genome Revolution in molluscs with a *tour de force* revision of key research trends observed over the past decade. *Omic* approaches such as Rad-seq, Transcriptome, Mitogenome and Whole Genome sequencing represent the most significant resources produced for this phylum. Importantly, the molecular mechanisms underscoring multiple biological novelties and adaptations observed in molluscs are starting to be unravelled. In contrast, compared to other metazoan lineages the genomic resources currently available for this lineage still lag far behind. We put forward that to fully grasp the evolutionary and adaptive roads of this tantalizing group of organisms, crucially depends on the full embracement of High-Throughput Sequencing (HTS) technologies in the near future.

Keywords

Review; High-Throughput Sequencing; Genomic Resources; Genome, Transcriptome; Mitogenome.

"***Love's feeling is more soft and sensible than are the horns of cockled snails***"

Shakespeare, *Love's Labor's Lost*

1. Phylum Mollusca

Molluscs are among the most ancient, widely diverse and ecologically successful group of Metazoans. The fossil record of molluscs can be traced back to the earliest Cambrian Period (~543 Mya) (Winson F and David R, 2008). This remarkably diverse phylum is the second most species-rich within the Metazoa, comprising an estimated 200,000 species distributed across marine, freshwater and terrestrial ecosystems (Winson F and David R, 2008). Mollusca is composed of eight classes: Cephalopoda, Gastropoda, Bivalvia, Polyplacophora, Scaphopoda, Monoplacophora and the Aplacophora (Solenogastres and Caudofoveata). Gastropoda and Bivalvia are responsible for around 96% of the species within this phylum and along with Cephalopod represent the three best-known classes (Brusca et al., 2003). Nevertheless, Scaphopoda and Polyplacophora match Cephalopod in terms of species richness (Brusca et al., 2003). Molluscs have successfully colonized almost every type of habitat (e.g. the high-pressure environments of the deep ocean, mountain tops, tropical rain forests, and deserts) and alongside with other invertebrates generally dominate in terms of species richness, abundance and biomass (Cardoso et al., 2011; Cuttelod et al., 2011; Kay, 1995; Sun et al., 2017). Representatives of the phylum reveal a large number of body plans and sizes (from laterally flattened microscopic bivalves to elongated and complex bodies of giant cephalopods), diversified life-histories specializations (e.g. parasitic life stages) with a wide life-span spectrum (including the longest-lived non-colonial metazoan, *Arctica islandica* with 507 years,) and growth rate, specialized body structures (e.g. hard protective shells and radula), defensive and predatory mechanisms (e.g. hematophagy and venom production) as well as an extensive repertoire of behavioural adaptations (e.g. memory and learning traits of cephalopods and the controlled swimming of scallop bivalves) (Butler et al., 2013; Kim et al., 2018; Y. Li et al., 2017; Modica et al., 2015; Schell et al., 2017; Simakov et al., 2012; Takeuchi et al., 2012; S. Wang et al., 2017). This incredibly diverse and phenotypic abundance confers molluscs an overall adaptive success that allows them to thrive in marine, freshwater, and terrestrial environments. All these factors have inevitably resulted in a close relationship with humans. Molluscs are commonly used as a source of protein throughout the world, i.e. bivalves (e.g., mussels and clams), cephalopods (e.g., octopus and squids) and gastropods (e.g. abalones, limpets, whelks, land snails) (Haszprunar and Wanninger, 2012). They display an elevated economic importance given the high number of species that are pursued for fishery (e.g. cephalopod) and for aquaculture (e.g. bivalves) (Haszprunar and Wanninger, 2012). Global aquaculture production of mussels, oysters and clams reaches over 17 million tons annually (Figueras et al., 2019; Hollenbeck and Johnston, 2018; Takeuchi, 2017), with massive worldwide landings values (e.g. North and South America (Gómez-Chiarri et al., 2015; Li et al., 2018), Europe (Figueras et al., 2019; Murgarella et al., 2016), Oceania (Nguyen et al., 2014; Powell et al., 2018) and Asia (Figueras et al., 2019; Mun et al., 2017). Furthermore, the aquaculture industry has important social relevance due to local job creation (Figueras et al., 2019; Murgarella et al., 2016). On the other hand, raw materials of economic and cultural value can be directly obtained from their multiplicity of forms and materials, such as shells, opercula, pearls, and mother-of-pearl that are used as decoration, jewels, or even currency (Du et al., 2017; Haszprunar and Wanninger, 2012; Strack, 2015). In fact, molluscs have been valuable sources of raw-materials for centuries and while some are still relevant today, others, such as natural pigmentation from secretions and sea-shells, had a high economic and cultural relevance in many ancient cultures (e.g. Phoenician and Roman) (Haszprunar and Wanninger, 2012; Strack, 2015). Molluscs are also used as important pollution monitoring "tools" (Farrington et al., 2016; Figueras et al., 2019) and as biological models to study biomineralization, ocean acidification and coastal ecosystem adaptation to climate change (Gazeau et al., 2013; Talmage and Gobler, 2010; Zhang et al., 2012). Furthermore, the highly complex cephalopod nervous system (the largest among invertebrates) has long been a model in cell biology and neurobiology and the marine snail, *Aplysia californica*, a model system to study memory and learning (e.g. Albertin et al., 2015; Bailey et al., 1983; Glanzman, 2009; Sigwart and Sumner-Rooney, 2015; Walters and Moroz, 2009).

Some molluscs are also very important from a medical and public health point of view. For example, many freshwater gastropod species are hosts of human pathogens causing relevant diseases (e.g. Schistosomiasis, Eosinophilic meningitis). The study of their ecology and biology is therefore fundamental for controlling and understanding the dispersion of these pathogens (Haszprunar and Wanninger, 2012; Liu et al., 2018; Raghavan and Knight, 2006; Strong et al., 2007). Furthermore, some species are intermediate hosts of pathogens that affect economically important livestock (Haszprunar and Wanninger, 2012; Strong et al., 2007). On the other hand, unique features such as hematophagy, venom production and tumour suppression found in some mollusc species suggest promising pharmacological applications for the identification of bactericidal and fungicidal molecules (Haszprunar and Wanninger, 2012) and for studying biological activities such cell signalling, immunological response and inhibition of tumour growth (Barghi et al., 2016; Modica et al., 2015; Nam et al., 2017).

Besides the direct human-related resources, molluscs play vital roles in their habitats, which sustain ecosystem functioning, promote ecological services and produce economic benefits. Molluscs occupy different positions within food chains, from basal decomposers to top-predators, with a high diversity of biotic (e.g. parasitic behaviours) and abiotic (e.g. keystone species and engineers) interactions that help moulding the ecosystems. Many molluscs are considered keystone species and/or ecosystem engineers promoting nutrient recycling, soil-generation and water filtration (Atkinson et al., 2013; Gómez-Chiarri et al., 2015; Powell et al., 2018; Renaut et al., 2018; Schell et al., 2017; Sun et al., 2017).

Some molluscs are also highly threatened, especially those living in freshwater or of restricted habitats, such as islands. Overexploitation, habitat degradation and loss, pollution, spreading of disease and invasive species, and climate changes are among the most harmful threats affecting molluscs worldwide (Gómez-Chiarri et al., 2015; Lydeard et al., 2006; Powell et al., 2018; Renaut et al., 2018; Zhang et al., 2012).

On the other hand, molluscs high diversity and adaptive success in combination with anthropogenic factors have resulted in some species becoming devastating invaders of many ecosystems worldwide (Haszprunar and Wanninger, 2012; Strong et al., 2007). Although, the International Union for Conservation of Nature - Invasive Species Specialist Group (IUCN—ISSG) has listed six molluscs in the 100 of the world's worst invasive alien species (Boudjelas et al., 2000), two terrestrial (i.e. *Achatina fulica* and *Euglandina rosea*) and four from aquatic ecosystems (i.e. *Pomacea canaliculata*, *Potamocorbula amurensis*, *Mytilus galloprovincialis*, and *Dreissena polymorpha*) many other species are harmful invaders worldwide (e.g. Sousa et al., 2008). The biological and ecological features of some species, such as fast growth, short life span, high fertility and high dispersal ability coupled with strong tolerance to pollutants and climatic fluctuations, as well as aggressive and diverse feeding behaviour confer to these species outstanding adaptive successes (Liu et al., 2018; Peñarrubia et al., 2015a, 2015b; Uliano-Silva et al., 2018). Once these species are successfully introduced outside their native geographic range, they can be responsible for damages at multiple levels: ecologically, they may promote local biodiversity loss by direct trophic interaction, surpassing resources consumption or promoting ecosystem modification (Calcino et al., 2019; Cowie, 2009; Liu et al., 2018; Peñarrubia et al., 2015a; Sun et al., 2019; Uliano-Silva et al., 2018); culturally by causing loss of endemic diversity but are also degrading systems with recreational interest (Karatayev et al., 2015; Peñarrubia et al., 2015a, 2015b); economically by promoting degradation of industrial infrastructures (Karatayev et al., 2015; Peñarrubia et al., 2015a, 2015b), stock improvement of aquaculturally important species (Li et al., 2018) and extirpation of economical relevant species and/or destruction of economical relevant crops cultivations (Cowie, 2009; Liu et al., 2018; Sun et al., 2019).

2. High-Throughput Sequencing (HTS) revolution and molluscs entrance into the genomics era

Molecular techniques for sequencing DNA and assess its variation had an extraordinary impact in biology, revolutionizing the fields of systematics, physiology, biochemistry, evolutionary biology and ecology (Andrew et al., 2013; Geist, 2011; Pauls et al., 2014). The main propel for this revolution started in 1975 with the development of the first widely used DNA sequencing technology that became known as Sanger's sequencing (Sanger et al., 1977; Sanger and Coulson, 1975). Sanger's sequencing is highly accurate and remained the leading form of DNA sequencing for decades. Despite still being widely used today, its overall high cost and time requirements limit its usage to produce small-scale information, especially for non-model organisms (Arumugam et al., 2019; Ghosh et al., 2018). Interestingly enough, although the Human Genome Project, which took over 10 years and cost approximately 3 billion USD, relied on Sanger sequencing, it impelled the beginning of a new era of sequencing technologies due to the increasing demand for fast, low-cost and High-Throughput Sequencing (HTS) outcomes (Arumugam et al., 2019; Ghosh et al., 2018). Since then, massively parallel sequencing or HTS techniques, have rapidly gained popularity over Sanger and have become increasingly more efficient, less time consuming and less expensive to the point of genome and transcriptome sequencing projects costing no more than a few thousand US dollars and being 200 times faster than Sanger sequencing alternatives (Arumugam et al., 2019; Calisi and MacManes, 2015; Goodwin et al., 2016). HTS techniques themselves have suffered an expanding evolution during the last decade and with it, new distinct terminologies such second, third and fourth generation sequencing also emerged (Arumugam et al., 2019; Calisi and MacManes, 2015; Goodwin et al., 2016). HTS sequencing methods use distinct biochemistry approaches. Although conceptually similar, four major broad categories can be defined, i.e. Sequencing by ligation, Sequencing by synthesis, Single-molecule real-time long reads, and Synthetic long reads, all with advantages and weaknesses regarding the applicability, accuracy and cost (reviewed in Goodwin et al., 2016). Additionally, in order to deal with the large quantities and diversity of data generated by HTS technologies, new bioinformatic tools and analysis methods have also started to emerge in parallel (e.g. Bradnam et al., 2013; Geniza and Jaiswal, 2017; Sedlazeck et al., 2018; Zhang et al., 2011).

At first, HTS approaches were mainly applied to human and/or model organism, especially in the biomedicine field (Arumugam et al., 2019). However, these tools rapidly took the stage and are nowadays commonly used for targeting transcriptome, metagenome, epigenome, exome, mitogenome and more importantly whole genome (re)-sequencing (see Dunisławska et al., 2017) and references within). In fact, one of the most remarkable outcomes of the HTS revolution was the democratization of whole genome analysis, critically for non-model organisms.

Consequentially, countless global consortium-based projects that intend to generate and catalogue whole-genomes for specific groups of non-model organisms have emerged in the last decade: Vertebrate Genome Project (https://vertebrategenomesproject.org), The 5000 Insect Genome Project of the Insect and other Arthropod Genome Sequencing Initiative (Evans et al., 2013), 1000 Fungal Genomes Project (http://1000.fungalgenomes.org/home/), NSF Plant Genome Research Program, 1K Insect Transcriptome Evolution (https://www.1kite.org), Global Invertebrate Genomics Alliance (Voolstra et al., 2017), Genomic Observatories Network (Davies et al., 2014), Global Genome Initiative (https://naturalhistory.si.edu/research/global-genome-initiative), Ocean Genome Legacy (https://www.northeastern.edu/ogl/). Many of these are now converging into the Earth Biogenome Project (EBP, Lewin et al., 2018). Although these initiatives aim to sequence a broad genome representation of each organism of interest, producing the genomes *per se* is not the main goal. Instead, it is expected to generate a deeper knowledge of non-model organism biology, identifying the evolutionary history and phenotypic expression of genomic features and use this as a proxy to understand organisms´ phylogeography, demography, adaptation patterns and conservation traits (Arumugam et al., 2019; Dunn and Ryan, 2015; Lopez et al., 2019; Richards, 2015; Savolainen et al., 2013).

In spite of the increasing attention towards new groups of organisms, the fact is that there is still a noticeable imbalance of genomic data produced from non-model species. Notably, vertebrates receive much more attention than invertebrates, and with the exception of arthropods and nematodes, most invertebrates have comparatively fewer data, with some groups without any representation (David et al., 2019; Dunn and Ryan, 2015; Lopez et al., 2019; Voolstra et al., 2017).

3. Molluscan Genomic resources

There is an increasing trend in the number of published papers that include genomic resources applied to molluscs. A search in peer-reviewed journals by querying the Web of Knowledge revealed that there are four main published genomic resources being applied: Restriction Site Associated DNA, Mitogenomes, Transcriptomes, and Whole Genomes (Figure P1.1, Supplementary Table P1.1). Moreover, in the last ten years Mitogenomes and Transcriptomes were the two genomic resources with the greatest increase with over 100 outputs each, followed by Whole Genomes (Figure P1.1, Supplementary Table P1.1). However, their use in the three major classes within Mollusca has not been homogenous. In fact, the vast majority of genomic resources has been applied to gastropods and bivalves, with very few being available for cephalopods, and even less or none for other mollusc lineages (Figure P1.2, Supplementary Table P1.1).

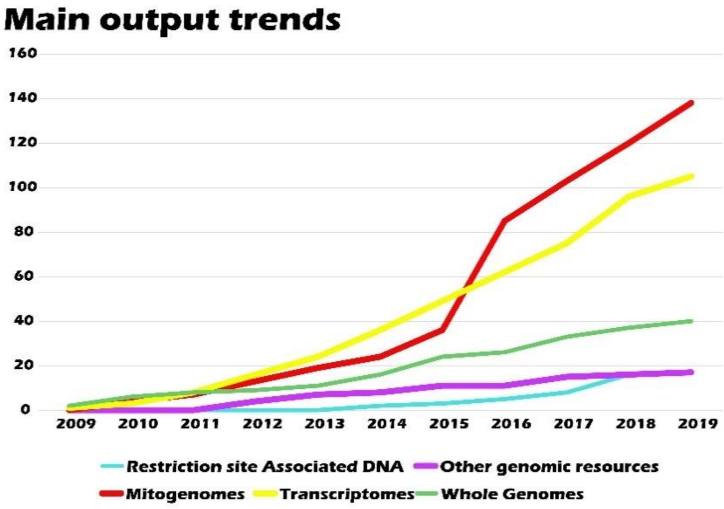


Figure P1. 1 - Number of publications per year, since 2009, in peer-reviewed journals by querying the Web of Knowledge using the following terms: "RAD", "RAD-seq", "mitogenome" and "transcriptome" associated with Mollusca, mollusk, mollusc, mollusks, gastropod, Gastropoda, bivalve, Bivalvia, Cephalopoda, cephalopods, Caudofoveata, Aplacophora, Polyplacophora, Monoplacophora and Scaphopoda. Coloured lines refer to the four main genomic approaches applied to molluscs, i.e. restriction site-associated DNA, mitogenome, transcriptome and whole-genome sequencing, as well as a total of other less frequently used genomic resources. References for these publications are detailed in Supplementary Table P1.1.

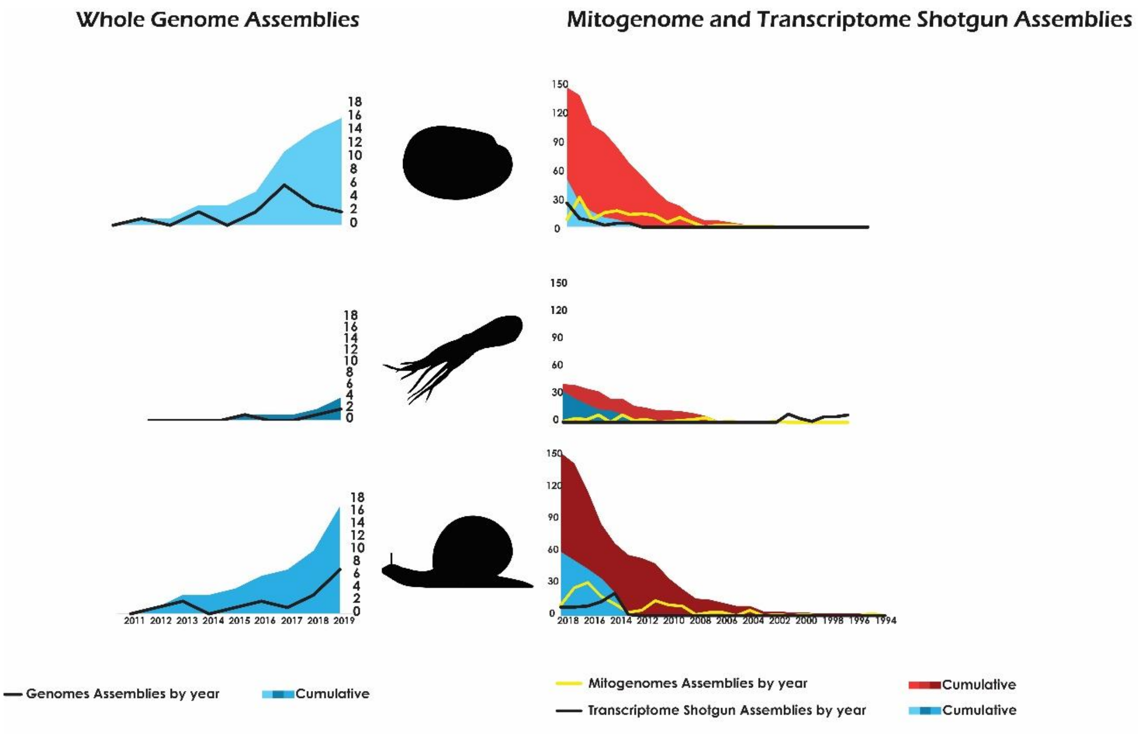


Figure P1. 2 Number of bivalves, cephalopods and gastropods’ whole-genome shotgun (left), transcriptome shotgun assemblies (right) and mitogenome assemblies (right) available on NCBI at 29 June 2019. Black and yellow lines represent the effective number of submits per year.

Further, it is possible to divide the 304 articles analysed (2009 - June 2019) into 13 main categories according to the main aims of the studies. Most of the Restriction Site Associated DNA approaches were applied in population genetic studies or quantitative trait *locus* mapping, using bivalves, followed by gastropods (Figure P1.3; for bibliographic details see Supplementary Table P1.1). On the other hand, Mitogenomes were mostly obtained for gastropods. It is interesting that most of these Mitogenomes were published only as a resource (sequencing and structure), with a lower number being applied in phylogeny, evolution or with other goals (Figure P1.3; for bibliographic details see Supplementary Table P1.1). The opposite pattern is observed for Transcriptomes. These resources have been mostly applied for Gene Identification and characterization, followed by gene expression profiles, mainly using bivalves (Figure P1.3; for bibliographic details see Supplementary Table P1.1). Regarding the “other genetic resources,” the most widely used are still the identification and use of microsatellites. Finally, Whole Genomes have been obtained mostly for gene Identification and characterization (Figure P1.3). Although most represent bivalves, the number of cephalopods and gastropods Whole Genome sequencing on NCBI at the moment has already surpassed those of 2018 (Figure P1.2). As we believe that this resource, i.e., Whole Genome Sequencing is going to become the most important and used one due to their possible applications (see below) the following sections will further detail it.

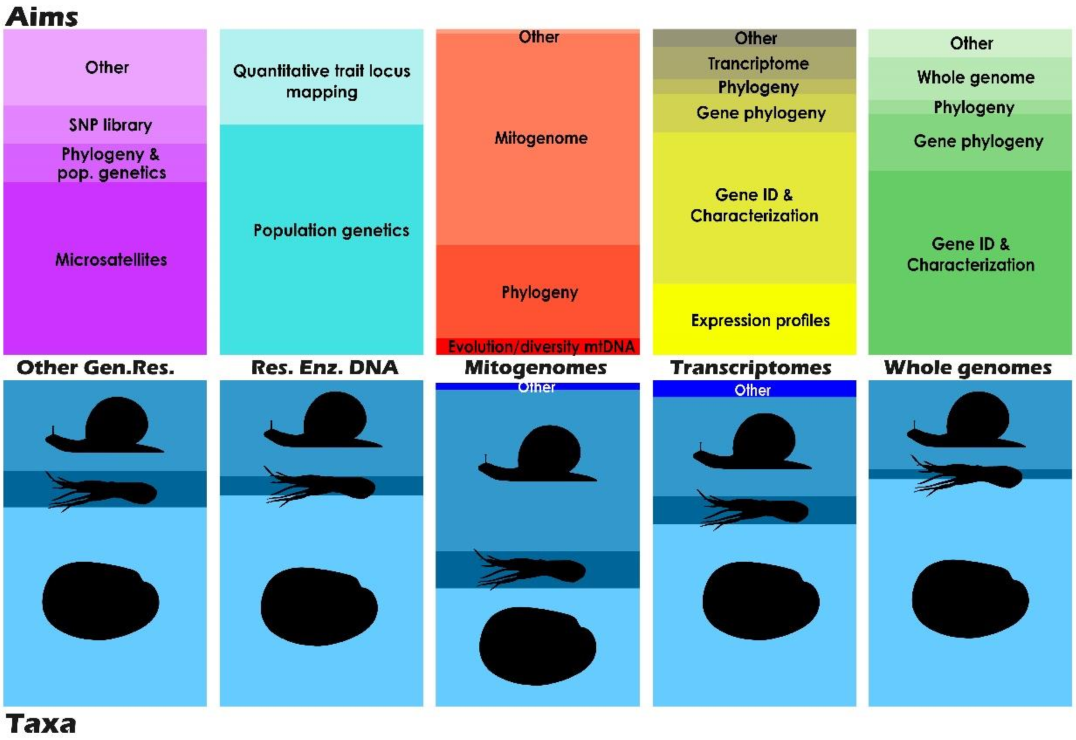


Figure P1. 3 - Compiled data set comprising 304 publications (2009 to June 2019) in peer-reviewed journals by querying the Web of Knowledge using the terms described in Figure P1.1 legend. References for these publications are detailed in Supplementary Table P1.1. The aims were divided into 13 main categories with others referring to the total of less frequently observed aims. Top columns: relative proportion of each of the categories regarding the four main genomic approaches applied to molluscs (restriction site-associated DNA, mitogenome, transcriptomics and whole-genome sequencing, and other for less frequently used genomic resources). Bottom columns: relative proportion of the molluscan taxa (cephalopods, gastropods, bivalves and other) represented by each of the four main genomic approaches applied.

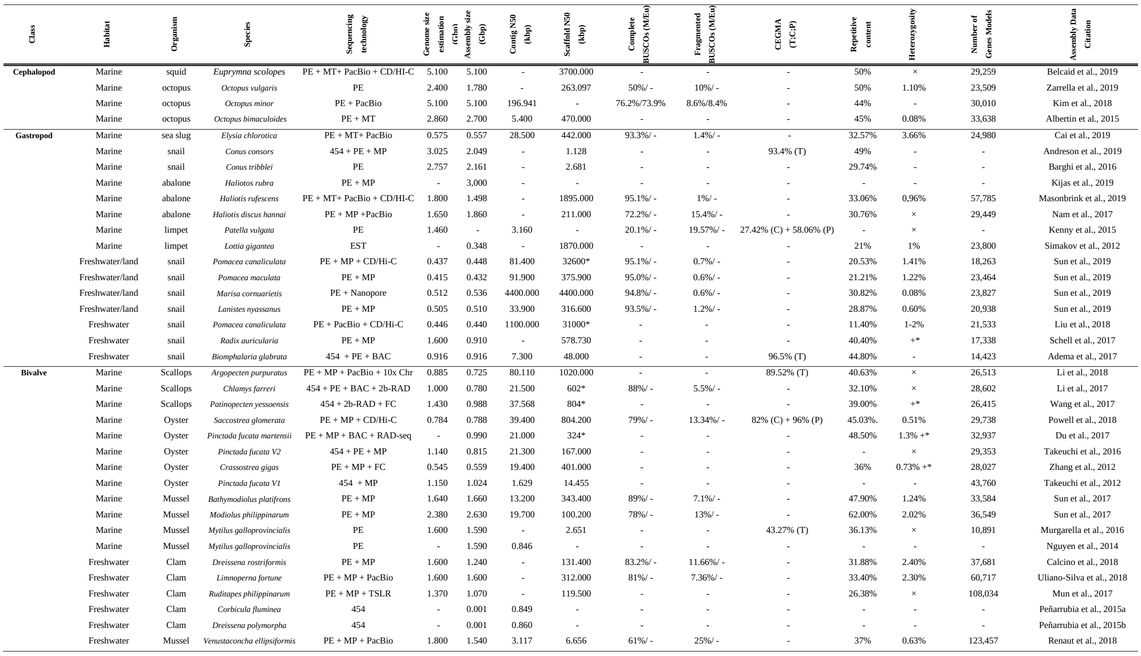
4. Molluscan Whole Genome Assemblies

The number of genome assemblies reported on NCBI for Nematoda and Arthropoda is astonishingly superior compared with the rest of invertebrates (see Table P1.1 and Dunn and Ryan, 2015; Lopez et al., 2019). Considering that Mollusca is the second most species-rich phylum, the number of assemblies available is still exceptionally low, representing only ~0.04% of the species described in the phylum (i.e. around 93,195). Initiatives, such as the Global Invertebrate Genomics Alliance (GIGA) (Lopez et al., 2014; Voolstra et al., 2017) are determined into balancing the genome sequencing within invertebrates and even though they also promote the sequencing of other organisms they give a particular “emphasis on marine taxa”, which neglects a significant fraction of molluscan taxa. Therefore, starting similar initiatives or consortium-based projects slowly focused on molluscs would be an essential step to a more representative genome cataloguing of the phylum. Additionally, by raising awareness to the importance of molluscs, these initiatives would increase the target funding interest, that pale far behind other taxa (Lopez et al., 2014; Voolstra et al., 2017). In fact, obtaining funding for molluscs’ genome projects can be difficult. For instance, the sequencing funding of the golden mussel genome project was obtained through a crowdfunding initiative started by the authors (Uliano-Silva et al., 2018). Nevertheless, since molluscs entered the genome era in 2009, with the first mollusc genome assembly (also the first Lophotrochozoa) of the sea hare *Aplysia californica*, the number of genomes has been increasing on a yearly basis (Figure P1.1 and P1.3; Table P1.2). Both Figure P1.1 and P1.2 and Table P1.2 reveal an increasing trend in the number of papers that included molluscan genomes assemblies and the number of WGS entries available on NCBI.

Table P1. 1 - Number and respective percentages of invertebrate species mitogenomes, transcriptome shotgun assemblies (TSA), whole-genome shotgun (WGS) and total number of sequence read archive (SRA). The percentages were calculated based on the estimated number of species reported in Brusca et al. (2003).

Phylum	Number and Percentage (%) of Mitogenome	Number and Percentage (%) of Unique TSA	Number and Percentage (%) of WGS	Total number of SRA
Tardigrada	1 (0.91)	6 (5.45)	2 (1.82)	277
Onychophora	4 (3.64)	0 (0.00)	1 (0.91)	37
Echiura	2 (1.48)	0 (0.00)	0 (0.00)	15
Rhombozoa	0 (0.00)	0 (0.00)	0 (0.00)	2
Monoblastozoa	0 (0.00)	0 (0.00)	0 (0.00)	0
Gastrotricha	1 (0.22)	0 (0.00)	0 (0.00)	15
Kinorhyncha	2 (1.33)	1 (0.67)	0 (0.00)	3
Acanthocephala	10 (0.91)	0 (0.00)	0 (0.00)	5
Nematomorpha	0 (0.00)	0 (0.00)	0 (0.00)	7
Entoprocta	2 (1.33)	0 (0.00)	0 (0.00)	3
Gnathostomulida	2 (2.50)	0 (0.00)	0 (0.00)	4
Loricifera	0 (0.00)	0 (0.00)	0 (0.00)	1
Cycliophora	0 (0.00)	0 (0.00)	0 (0.00)	2
Sipuncula	3 (0.94)	0 (0.00)	0 (0.00)	13
Echiura	2 (1.54)	0 (0.00)	0 (0.00)	15
Phoronid	0 (0.00)	0 (0.00)	0 (0.00)	53
Ectoprocta	7 (0.16)	0 (0.00)	0 (0.00)	5
Chaetognatha	5 (5.00)	0 (0.00)	0 (0.00)	11
Nemertea	17 (1.89)	1 (0.11)	1 (0.11)	119
Orthonectida	0 (0.00)	0 (0.00)	1 (5.00)	4
Priapulida	1 (6.25)	2 (12.50)	1 (6.25)	23
Placozoa	1 (100.00)	1 (100.00)	1 (100.00)	17
Brachiopoda	4 (1.19)	2 (0.60)	2 (0.60)	58
Porifera	56 (1.02)	5 (0.09)	2 (0.04)	3418
Ctenophora	0 (0.00)	2 (2.00)	3 (3.00)	336
Annelida	74 (0.45)	20 (0.12)	6 (0.04)	3294
Rotifera	1 (0.06)	8 (0.44)	10 (0.56)	224
Echinodermata	49 (0.70)	33 (0.47)	11 (0.16)	3376
Chordata (Cephalochordata/Urochordata/Hemichordata)	28 (0.90)	7 (0.23)	20 (0.64)	3261
Cnidaria	132 (1.32)	76 (0.76)	26 (0.26)	11198
Mollusca	352 (0.38)	107 (0.11)	35 (0.04)	13193
Cephalopoda	42 (4.67)	26 (2.89)	4 (0.44)	994
Bivalvia	144 (0.72)	38 (0.19)	14 (0.07)	7126
Gastropod	152 (0.22)	43 (0.06)	17 (0.02)	5154
Polyplacophora	6 (0.60)	0 (0.00)	0 (0.00)	36
Scaphopoda	2 (0.22)	0 (0.00)	0 (0.00)	7
Aplacophora	4 (1.08)	0 (0.00)	0 (0.00)	38
Monoplacophora	2 (8.00)	0 (0.00)	0 (0.00)	3
Platyhelminthes	91 (0.46)	29 (0.15)	35 (0.18)	7781
Nematoda	109 (0.44)	53 (0.21)	118 (0.47)	35093
Arthropoda	1823 (0.17)	1259 (0.11)	480 (0.04)	192025

Table P1. 2 - Main characteristics of molluscan genomes projects and genomes statistics, as reported in their respective references. The codes used for the sequencing technologies refer to: Illumina Paired-end (PE) + Illumina Mate Paired (MP), Pacific Biosciences (PacBio) + Dovetail Genomics Hi-C or Chicago libraries (CD/Hi-C), Pyrosequencing Roche 454 (454), EST Sanger sequencing (EST), Bacterial Artificial Chromosome (BAC), 10× Chromium Genomics (10x Chr), Fosmid Clones (FC), Linkage maps 2b-RAD (2b-RAD), Linkage maps RAD-seq (RAD-seq) and Synthetic long-read DNA (TSLR). For the Benchmarking Universal Single-Copy Orthologs (BUSCO), the analysis reported with Metazoa (M) and Eukaryotic (Eu) database is shown. For Core Eukaryotic Genes Mapping Approach (CEGMA), the Total (T), Complete (C) and Partial (P) results are shown. Scaffold N50 values marked with * refer to assemblies that reached chromosome level. Heterozygosity values marked with × were reported to be high, in the reference, but the values were not expressed. Heterozygosity values marked with +* refer to individuals artificially inbreeded prior to the assembly (**To see the expanded Figure open the link** https://link.springer.com/article/10.1007/s10750-019-04111-1/tables/2).



Class	Habitat	Organism	Species	Sequencing technology	Genome size estimation (Gbn)	Assembly size (Gbp)	Contig N50 (kbp)	Scaffold N50 (kbp)	Complete BUSCOs (M/Eu)	Fragmented BUSCOs (M/Eu)	CEGMA (T;C;P)	Repetitive content	Heterozygosity	Number of Genes Models	Assembly Data Citation
Cephalopod	Marine	squid	*Euprymna scolopes*	PE + MT+ PacBio + CD/HI-C	5.100	5.100	-	3700.000	-	-	-	50%	×	29,259	Belcaid et al., 2019
	Marine	octopus	*Octopus vulgaris*	PE	2.400	1.780	-	263.097	50%/ -	10%/ -	-	50%	1.10%	23,509	Zarrella et al., 2019
	Marine	octopus	*Octopus minor*	PE + PacBio	5.100	5.100	196.941	-	76.2%/73.9%	8.6%/8.4%	-	44%	-	30,010	Kim et al., 2018
	Marine	octopus	*Octopus bimaculoides*	PE + MT	2.860	2.700	5.400	470.000	-	-	-	45%	0.08%	33,638	Albertin et al., 2015
Gastropod	Marine	sea slug	*Elysia chlorotica*	PE + MT+ PacBio	0.575	0.557	28.500	442.000	93.3%/ -	1.4%/ -	-	32.57%	3.66%	24,980	Cai et al., 2019
	Marine	snail	*Conus consors*	454 + PE + MP	3.025	2.049	-	1.128	-	-	93.4% (T)	49%	-	-	Andreson et al., 2019
	Marine	snail	*Conus tribblei*	PE	2.757	2.161	-	2.681	-	-	-	29.74%	-	-	Barghi et al., 2016
	Marine	abalone	*Haliotos rubra*	PE + MP	-	3,000	-	-	-	-	-	-	-	-	Kijas et al., 2019
	Marine	abalone	*Haliotis rufescens*	PE + MT+ PacBio + CD/HI-C	1.800	1.498	-	1895.000	95.1%/ -	1%/ -	-	33.06%	0,96%	57,785	Masonbrink et al., 2019
	Marine	abalone	*Haliotis discus hannai*	PE + MP +PacBio	1.650	1.860	-	211.000	72.2%/ -	15.4%/ -	-	30.76%	×	29,449	Nam et al., 2017
	Marine	limpet	*Patella vulgata*	PE	1.460	-	3.160	-	20.1%/ -	19.57%/ -	27.42% (C) + 58.06% (P)	-	×	-	Kenny et al., 2015
	Marine	limpet	*Lottia gigantea*	EST	-	0.348	-	1870.000	-	-	-	21%	1%	23,800	Simakov et al., 2012
	Freshwater/land	snail	*Pomacea canaliculata*	PE + MP + CD/Hi-C	0.437	0.448	81.400	32600*	95.1%/ -	0.7%/ -	-	20.53%	1.41%	18,263	Sun et al., 2019
	Freshwater/land	snail	*Pomacea maculata*	PE + MP	0.415	0.432	91.900	375.900	95.0%/ -	0.6%/ -	-	21.21%	1.22%	23,464	Sun et al., 2019
	Freshwater/land	snail	*Marisa cornuarietis*	PE + Nanopore	0.512	0.536	4400.000	4400.000	94.8%/ -	0.6%/ -	-	30.82%	0.08%	23,827	Sun et al., 2019
	Freshwater/land	snail	*Lanistes nyassanus*	PE + MP	0.505	0.510	33.900	316.600	93.5%/ -	1.2%/ -	-	28.87%	0.60%	20,938	Sun et al., 2019
	Freshwater	snail	*Pomacea canaliculata*	PE + PacBio + CD/Hi-C	0.446	0.440	1100.000	31000*	-	-	-	11.40%	1-2%	21,533	Liu et al., 2018
	Freshwater	snail	*Radix auricularia*	PE + MP	1.600	0.910	-	578.730	-	-	-	40.40%	+*	17,338	Schell et al., 2017
	Freshwater	snail	*Biomphalaria glabrata*	454 + PE + BAC	0.916	0.916	7.300	48.000	-	-	96.5% (T)	44.80%	-	14,423	Adema et al., 2017
Bivalve	Marine	Scallops	*Argopecten purpuratus*	PE + MP + PacBio + 10x Chr	0.885	0.725	80.110	1020.000	-	-	89.52% (T)	40.63%	×	26,513	Li et al., 2018
	Marine	Scallops	*Chlamys farreri*	454 + PE + BAC + 2b-RAD	1.000	0.780	21.500	602*	88%/ -	5.5%/ -	-	32.10%	×	28,602	Li et al., 2017
	Marine	Scallops	*Patinopecten yessoensis*	454 + 2b-RAD + FC	1.430	0.988	37.568	804*	-	-	-	39.00%	+*	26,415	Wang et al., 2017
	Marine	Oyster	*Saccostrea glomerata*	PE + MP + CD/Hi-C	0.784	0.788	39.400	804.200	79%/ -	13.34%/ -	82% (C) + 96% (P)	45.03%.	0.51%	29,738	Powell et al., 2018
	Marine	Oyster	*Pinctada fucata martensii*	PE + MP + BAC + RAD-seq	-	0.990	21.000	324*	-	-	-	48.50%	1.3% +*	32,937	Du et al., 2017
	Marine	Oyster	*Pinctada fucata V2*	454 + PE + MP	1.140	0.815	21.300	167.000	-	-	-	-	×	29,353	Takeuchi et al., 2016
	Marine	Oyster	*Crassostrea gigas*	PE + MP + FC	0.545	0.559	19.400	401.000	-	-	-	36%	0.73% +*	28,027	Zhang et al., 2012
	Marine	Oyster	*Pinctada fucata V1*	454 + MP	1.150	1.024	1.629	14.455	-	-	-	-	-	43,760	Takeuchi et al., 2012
	Marine	Mussel	*Bathymodiolus platifrons*	PE + MP	1.640	1.660	13.200	343.400	89%/ -	7.1%/ -	-	47.90%	1.24%	33,584	Sun et al., 2017
	Marine	Mussel	*Modiolus philippinarum*	PE + MP	2.380	2.630	19.700	100.200	78%/ -	13%/ -	-	62.00%	2.02%	36,549	Sun et al., 2017
	Marine	Mussel	*Mytilus galloprovincialis*	PE	1.600	1.590	-	2.651	-	-	43.27% (T)	36.13%	×	10,891	Murgarella et al., 2016
	Marine	Mussel	*Mytilus galloprovincialis*	PE	-	1.590	0.846	-	-	-	-	-	-	-	Nguyen et al., 2014
	Freshwater	Clam	*Dreissena rostriformis*	PE + MP	1.600	1.240	-	131.400	83.2%/ -	11.66%/ -	-	31.88%	2.40%	37,681	Calcino et al., 2018
	Freshwater	Clam	*Limnoperna fortune*	PE + MP + PacBio	1.600	1.600	-	312.000	81%/ -	7.36%/ -	-	33.40%	2.30%	60,717	Uliano-Silva et al., 2018
	Freshwater	Clam	*Ruditapes philippinarum*	PE + MP + TSLR	1.370	1.070	-	119.500	-	-	-	26.38%	×	108,034	Mun et al., 2017
	Freshwater	Clam	*Corbicula fluminea*	454	-	0.001	0.849	-	-	-	-	-	-	-	Peñarrubia et al., 2015a
	Freshwater	Clam	*Dreissena polymorpha*	454	-	0.001	0.860	-	-	-	-	-	-	-	Peñarrubia et al., 2015b
	Freshwater	Mussel	*Venustaconcha ellipsiformis*	PE + MP + PacBio	1.800	1.540	3.117	6.656	61%/ -	25%/ -	-	37%	0.63%	123,457	Renaut et al., 2018

At the time of this review, the number of studies that describe partial or complete genome mollusc assemblies were 33 (Table P1.2): four cephalopods, a squid (Belcaid et al., 2019) and three octopus (Albertin et al., 2015; Kim et al., 2018; Zarrella et al., 2019); fifteen gastropods genomes, most marine, including a sea slug (Cai et al., 2019), two snails (Andreson et al., 2019; Barghi et al., 2016), three abalones (Kijas et al., 2019; Masonbrink et al., 2019; Nam et al., 2017) and two limpets (Kenny et al., 2015; Simakov et al., 2012) but also seven non-marine snails (Adema et al., 2017; Liu et al., 2018; Schell et al., 2017; Sun et al., 2019); eighteen bivalves, most marine, including three scallops (Y. Li et al., 2017; S. Wang et al., 2017), five oysters (Du et al., 2017; Powell et al., 2018; Takeuchi et al., 2012, 2016; Zhang et al., 2012) and fours mussels (Murgarella et al., 2016; Nguyen et al., 2014; Sun et al., 2017) and six freshwater bivalves (Calcino et al., 2019; Mun et al., 2017; Peñarrubia et al., 2015a, 2015b; Renaut et al., 2018; Uliano-

Silva et al., 2018). In more than one occasion more than one mollusc genome has been generated in one paper: in Sun et al., (2019) four snail genomes were assembled; Sun et al., (2017) produced two mussels’ genomes assemblies; and Wang et al., (2017) while producing the first chromosome-level assembly of a mollusc genome (i.e. scallop) by using an already published high-density linkage maps for Pacific oyster (*C. gigas*) and pearl oyster (*P. fucata*), also generated a chromosome-anchored genome for these two species. Additionally, although Kijas et al., (2019) described the production of one abalone genome, the authors used unpublished abalone genome (Botwright et al., unpublished) for an alignment-based assembly and for variation calling in the produced genome. On the other hand, on WGS NCBI Search Engine is available the direct submission of: one additional cephalopod genome *Architeuthis dux* (GenBank assembly accession: GCA_006491835.1); four additional gastropod genomes, *Physella acuta* (GenBank assembly accession: GCA_004329575.1), *Conus consors* (GenBank assembly accession: GCA_004193615.1; later (Andreson et al., 2019) produce a better assembly of this species), *Haliotis rubra* (GenBank assembly accession: GCA_003918875.1), *Aplysia californica* (GenBank assembly accession: GCA_000002075.2); and four bivalves: *Argopecten irradians concentricus* (GenBank assembly accession: GCA_004382765.1), *Argopecten irradians irradians* (GenBank assembly accession: GCA_004382745.1), *Bankia setacea* (GenBank assembly accession: GCA_001922985.1), and *Crassostrea gigas* (GenBank assembly accession: GCA_005518195.1).

Molluscan genome projects have proven to be a challenging task from the beginning, given that obtaining DNA with high quality, quantity and purity is mandatory for library preparation and HTS (Schultzhaus et al., 2019). Particularly, the success of long-read sequencing technologies is entirely dependent on high molecular weight (HMW) DNA (>50kb). Since most recent sequencing technologies do not rely on DNA amplification, a high initial quantity of DNA is required, ranging from a few hundred nanograms, for short-reads, up to tens of micrograms, for long-reads. DNA purity is also critical, as contaminates that bind to the DNA can hinder library preparation and/or sequencing steps (Mayjonade et al., 2017; Schultzhaus et al., 2019). The high concentration of polysaccharides generally found in molluscs´ tissues poses one of the main problems for efficient DNA extraction (Arseneau et al., 2017; Sokolov, 2000). If not effectively removed, polysaccharides may preclude enzymatic reactions, such as ligation and polymerization, essential in library preparation and HTS (Arseneau et al., 2017). Therefore, commonly used DNA extraction kits may not be the best option for molluscan DNA extraction. Consequently, specific mollusc DNA extraction kits or modified protocols, specially intended to remove polysaccharide contaminations while ensuring HMW DNA outputs, are frequently used for molluscs (e.g. Arseneau et al., 2017; D. R. Gonzalez et al., 2019; Richards et al., 2013; Sokolov, 2000; Swart et al., 2019). Finally, for small molluscan species obtaining the necessary high quantities of DNA, from a single individual, may be impossible. Therefore, pooling several individuals in one extraction batch is the most frequent option. However, this strategy may introduce problems in the subsequent assembly steps, given that it may increase polymorphic content within the pooled data (see below for heterozygosity).

Producing quality assemblies for molluscs' genomes could be challenging due to several factors inherent to the genome structure and/or composition, i.e. size, composition of repetitive elements and levels of heterozygosity. Assuming that for larger genomes, higher amounts of sequencing data are needed in order to ensure sufficient sequencing coverage, having an estimated genome size is essential for a cost-effective sequencing project. Databases such as the Animal Genome Size Database (http://www.genomesize.com) have available size estimations for many organisms, including 281 molluscs. However, many groups are poorly represented, such as Solenogastres, Polyplacophora and Scaphopoda, or not represented at all, such as Monoplacophora and Caudofoveata (but see exception in Kocot et al., 2016b). Additionally, within each class, representatives are also unevenly distributed, as is the case of freshwater mussels of the Order Unionida, with only two representatives. Consequently, genome size comparisons with closely related species are the common practice before initiating a genome project (Dominguez Del Angel et al., 2018). In molluscs, this strategy may be problematic, since highly variable genome sizes are observed, even between related groups (Table P1.2; Takeuchi, 2017). Considering the published genomes, Cephalopods have the overall largest estimated genomes sizes, with values ranging between 2.4 Giga base pairs (Gbp) of *Octopus vulgaris* and 5.1 Gbp of the *Euprymna scolopes* and *Octopus minor*. Gastropods, while revealing smaller genome sizes, show an equally large variation with values of genome size ranging between 414.8 Mega base pairs (Mbp) of *Pomacea maculata* and 3.025 Gbp of *Conus consors*. Finally, bivalves' genome sizes ranged between 545 Mbp of *Crassostrea gigas* and 2.38 Gbp of *Modiolus philippinarum*. Another considerably relevant challenge for assembling mollusc's genomes is their generally high heterozygosity (i.e. high number of polymorphic loci), which induces the fragmentation of the genome assembly (Table P1.2 and reviewed in Takeuchi, 2017). Although assemblers generally try to collapse heterozygotic regions into a reference consensus assembly, when the values of heterozygosity are elevated, they may fail to do so, consequently resulting in two separated assemblies and raising the fragmentation of the assembly (Dominguez Del Angel et al., 2018). In molluscs, high heterozygosity, which may be due to several factors (e.g. large population sizes, wide habitat distribution, high fecundity rates), has been a recurrent problem in genome assembly projects (Takeuchi, 2017). As seen in Table P1.2, only on two occasions the observed heterozygosity is very low, i.e. for *Octopus bimaculoides* and for *Marisa cornuarietis* with 0.08%. As for the rest, the values are generally high, with *Elysia chlorotica* showing a 3.66% heterozygosity level (Table P1.2). One of the strategies that has been used to tackle this problem is by rearing inbred lineages before DNA sequencing (Du et al., 2017; Schell et al., 2017; S. Wang et al., 2017; Zhang et al., 2012). Although these strategies can reduce the observed heterozygosity (reviewed in Takeuchi, 2017), sometimes the values are still high (e.g. 1.3% of a third generation of an inbreeding line in Du et al., 2017). Furthermore, these strategies are limited to the species easily cultured, with fast reproductivity and high fecundity rate and success. In addition to the genome sizes and heterozygosity, molluscs also reveal high content of repetitive sequences, which generally represent one of the most challenging obstacles for genome assembly (Table P1.2 and Sedlazeck et al., 2018; Takeuchi, 2017). Unresolved repeats (e.g. when repeats are larger than the sequenced read size) caused the contigs to end during the assembly process, resulting in highly fragmented assemblies (Dominguez Del Angel et al., 2018; Sedlazeck et al., 2018). Furthermore, repeats can result in mis-assemblies by misguiding assembly software into collapsing distinct repetitive regions as unique regions, resulting in lower quality and completeness of the final assembly assemblies (Dominguez Del Angel et al., 2018; Sedlazeck et al., 2018). As seen in Table P1.2, the overall percentage of repetitive sequences within the molluscan genome is very high with a mean overall value of 36.35%. The highest recorded values are observed in the marine mussel *Modiolus philippinarum* (i.e. 62%), the lowest in freshwater snail *Pomacea canaliculata* (i.e. 11.4%) and similarly to the genome size, within closely related groups the estimated repetitive content is highly variable (especially in gastropods). Consequentially, those molluscs genomes whose assembly was performed de novo solely (i.e. no reference guide assembly) based on short-read technologies resulted in highly fragmented genomes (with low values of quality assessment statistics) with large portions of the genome being misrepresented or even absent (low BUSCO/CEGMA scores with high fragmented hits) (Table P1.2 and (Barghi et al., 2016; Kenny et al., 2015; Murgarella et al., 2016; Nguyen et al., 2014). Therefore, to tackle this problem and to ensure high contiguity of the assemblies, molluscan genome projects necessarily need to include long-read sequencing, such as Pacific Biosciences (PacBio) and/or Oxford Nanopore sequencing technologies, which can routinely achieve read lengths over 100 kilo base pairs (kbp)

(Goodwin et al., 2016). Nowadays, there are software programs capable of producing assemblies relying solely on long-read sequencing (reviewed in Sedlazeck et al., 2018). However, due to these technologies higher error rates and higher costs (compared with short-read sequencing) researchers have been using hybrid assembly strategies, i.e. combining low-coverage long-read sequencing and high-coverage short-read (Table P1.2). Hybrid strategies allow the resolution of repetitive regions and heterozygosity related problems by using long-read information while mitigating long-reads' local base-call high error rates by correction with short-read data and keeping the overall sequencing cost considerably low (Besser et al., 2018; Dominguez Del Angel et al., 2018; Goodwin et al., 2016; Miller et al., 2017; Renaut et al., 2018; Sedlazeck et al., 2018). Hybrid assembly strategies dominate most of molluscan genome projects (Table P1.2), generally by a combination of Illumina Paired-end short-reads with one or more long-read sequencing, such as Illumina Mate-paired and PacBio. However, given the high genome diversity observed in molluscs, every paper seems to take a different approach to sequencing and assembly methodologies and, to date, no effective generalized pipeline for molluscan genome assembling has been established. In addition, sequencing strategies have been accompanying the emergence of novel technologies, such as 10x Chromium linked reads and chromatin crosslinking protocols such as Dovetails Genomics Hi-C or Chicago libraries (Table P1.2). Furthermore, the increasing offer of sequencing platforms and bioinformatic tools has also increased the quality of genome assemblies. At the date of this review, seven molluscan genomes have been assembled at the chromosome level, all of them in the last two years. Of these, five used high-density linkage maps 2d-RAD/RAD-seq to anchored draft assemblies to the chromosomes (Du et al., 2017; Y. Li et al., 2017; S. Wang et al., 2017). The other two (from the same species, *Pomacea caniculata*) were produced using long-range chromatin crosslinking Hi-C (Liu et al., 2018; Sun et al., 2019). This last technology has been used in the assembly of four mollusc genomes (Belcaid et al., 2019; Liu et al., 2018; Masonbrink et al., 2019; Sun et al., 2019). Importantly, even when the chromosome level was not achieved, the inclusion of these technologies significantly improved the assemblies. After using Chicago/Hi-C data to scaffold, values of N50 have shown significant increases in size: from 98kbp to 3.7Mbp in the squid *Euprymna scolopes* (Belcaid et al., 2019); from 588kb to 1.895Mbp in the red abalone *Haliotis rufescens* (Masonbrink et al., 2019); from 115Kbp to 806 Kbp in an oyster *Saccostrea glomerata* (Powell et al., 2018); and from 3.40 Mbp to 32.6Mbp and 1.1 Mb to 31Mbp in *Pomacea canaliculata* (Liu et al., 2018; Sun et al., 2019). In fact, Chicago/Hi-C is a very promising technology and, in mammals, when combined with a modest Illumina short-read Paired-end assembly, has been shown to produce chromosome-length scaffolds at incredibly low prices ($1000) (Dudchenko et al., 2018). These outstanding results make Chicago/Hi-C incredibly advantageous, when compared with other long-range chromatin crosslinking technologies and given the efficient results shown in molluscan genome assemblies, it will certainly be a preferable approach for future genome projects.

Following genome assembly, depending on the goals of the genome projects, structural and functional annotation are usually the following steps although not always a requirement (e.g. Kijas et al., 2019; Nguyen et al., 2014; Peñarrubia et al., 2015a, 2015b). Although different tools can be used, molluscan genome's annotations generally are guided by the same conceptual framework, i.e. identification and masking of repeat elements followed by a structural and functional annotation of the masked genome. The overall high content of repetitive elements observed in molluscan genomes makes the first step of masking extremely important. The term repeated elements designate different nucleotide structures, from several low-complexity sequences to highly complex structures such as transposable elements (Dominguez Del Angel et al., 2018; Yandell and Ence, 2012). Furthermore, these structures are generally poorly conserved and therefore highly variable between different taxonomic groups (Yandell and Ence, 2012). In fact, although some molluscan repeat databases are available, they are generally incomplete, and every genome project inevitably needs to produce a de novo repeat library prior to masking (e.g. Murgarella et al., 2016; Renaut et al., 2018). The subsequent step, i.e. gene model predictions, which once again follows a generalized framework in most cases, results from an *ab initio* prediction (i.e. extrapolate genes predictions solely based on nucleotide structure) combined with evidence-based predictions (generally using RNAseq/Transcriptomic) (see Table P1.2 and references within). Gene prediction modules for mollusc's genomes range from 10,891 to 123,457 genes (Table P1.2). However, 70% of the molluscan genomes are between 20,000 to 40,000 genes modules (Table P1.2). Furthermore, the quality of the annotation is intrinsically dependent on the quality of the initial genome assembly (Mudge and Harrow, 2016). Even small improvements in the assembly can have a great impact on the annotation (Mudge and Harrow, 2016). For instance, the annotation of the improved second version of the *Pinctada fucata* genome (Takeuchi et al., 2012, 2016) resulted in a significant reduction in the number of gene predictions (Table P1.2). In the case of *Pomacea canaliculata*, two distinct genome assemblies, from different projects, the total number of predictions varied in around 2,000 genes (Table P1.2).

Finally, the last step is to assign biological information to predictions (i.e. functional annotation), which in molluscan genomes projects is generally accomplished by querying prediction against different public sequence repositories, such as Swissprot (Bairoch and

Apweiler, 1999), KEGG (Kanehisa and Goto, 2000), Uniprot (Bateman et al., 2017), NCBI NR (Pruitt et al., 2007) among others.

In conclusion, the increasing number of genomes produced has been accompanied by increasing the quality of the assemblies and annotations (Table P1.2). With the growing efficiency and lower cost of new sequencing technologies, it is expectable for these trends to continue and probably, in a couple of years, chromosome level assemblies will be the standard for molluscan genome projects. Finally, while the perspective of new high-quality assemblies represents exciting news, they also highlight the importance of updating the first lower qualities assemblies and annotations in order for them to match the progress of new assemblies.

As discussed further down each of the produced genomes are aimed to explore ecological, molecular or economical specific relevant features of the species under study. The increasing number of available genomes offers a unique opportunity to resolve molluscan evolutionary uncertainties, promote the conservation and/or mitigation of molluscs' socioeconomic impacts as well as manage the economic and ecological services they provide while starting a new era of comprehensive genetic research on these widely diverse organisms.

5. Molluscan genomes studies

"***to dive deep, deep, courageously down into some unexploited region of the genome (sic), into some common deep sea of unrecorded knowledge and bring, triumphant, to the surface some treasure buried, lost, forgotten***"

"Tribute to Freud. Writing on the Wall" by Hilda Doolittle, 1933

Sequencing of molluscan genomes has been motivated by the wide range of ecological, economic and scientific reasons that underlie this phylum's immense diversity. The global scale economic impact of molluscan fishery and aquaculture (Takeuchi, 2017) has been leveraging a great share of the produced mollusc genomes, with representatives in the three main groups (e.g. cephalopods Zarrella et al., 2019, gastropods Nam et al., 2017 and bivalves Murgarella et al., 2016). Most of these studies aim to improve aquaculture farming programs and/or promote conservation of threatened stocks. Generating genomic resources will allow a deeper understanding of the molecular mechanisms behind the adaptation to changing or adverse ecological conditions and the immunological response that can improve resistance to diseases that threaten these organisms (Li et al., 2018; Mun et al., 2017; Murgarella et al., 2016; Powell et al., 2018). Additionally, genomic resources may also be applied for more efficient management of breeding programmes by allowing variance detection, selective breeding and pedigree assignment that ultimately can generate new approaches to increase productivity and value of farmed species (Kijas et al., 2019; Masonbrink et al., 2019; Nguyen et al., 2014). Besides fishery and aquaculture, molluscs offer other numerous relevant features that have motivated genome sequencing. Mollusc genomes may have relevant applications in medical and pharmaceutical fields, such as in neuroscience by continuing the studies of the cephalopods and some Heterobranchia neurologic systems and their behavioural adaptations (Albertin et al., 2015; Bailey et al., 1983; Glanzman, 2009; Kim et al., 2018; Zarrella et al., 2019); cancer research by exploring underlying genetic processes behind abalones tumour suppressing features (Nam et al., 2017); search for new drugs by exploring the arrangement and expression of group-specific venom peptide genes present in Conoidea snails (Andreson et al., 2019; Barghi et al., 2016); mitigation the spread of human parasites, by exploring the molecular features and developing novel target strategies for either reducing molluscan suitability as intermediary host or controlling mollusc with synthetically designed genetic control approaches (e.g. molluscicides and gene drivers) (Adema et al., 2017; Liu et al., 2018; Sun et al., 2019). Genomic surveys of problematic invasive species are also essential for understanding the molecular features that confer ecological plasticity as well as producing new tools for invasiveness control (Li et al., 2018; Liu et al., 2018; Sun et al., 2019; Uliano-Silva et al., 2018).

Early molluscan evolution has been extensively explored using fossil records and phylogenetic approaches with many uncertainties persisting and different and conflicting phylogenetic hypotheses being proposed (Kocot, 2013; Wanninger and Wollesen, 2019; Winson F and David R, 2008). Although still largely underrepresented in terms of genomic resources (Figure P1.3, Supplementary Table P1.1), the lesser-known Mollusca classes, i.e. Solenogastres, Caudofoveata, Polyplacophora, Monoplacophora and Scaphopoda, have proved their pivotal role in resolving some long-lasting molluscan phylogenetic relationships (Kocot, 2013; Kocot et al., 2011, 2016b, 2019a; Mikkelsen et al., 2019; Smith et al., 2011; Wanninger and Wollesen, 2019). In fact, comprehensive large-scale phylogenomic studies were essential to generate a consensual molluscan phylogenetic scenario, characterized by a deep dichotomy that splits Mollusca into Aculifera (composed of Polyplacophora and Aplacophora) and Conchifera (composed of Monoplacophora, Cephalopoda, Gastropoda, Bivalvia and Scaphopoda) (Kocot, 2013;

Kocot et al., 2011; Smith et al., 2011; Wanninger and Wollesen, 2019). On the other hand, resolving intraclade phylogenetic relationships is still a challenge, particularly in Conchifera. Although, it is now consensual that Cephalopoda and Gastropoda are not sister taxa (as previously thought) the positioning of Monoplacophora and specially Scaphopoda is still unclear, with different studies showing different results (Kocot, 2013; Kocot et al., 2011; Sigwart and Sumner-Rooney, 2015; Smith et al., 2011; Wanninger and Wollesen, 2019). Consequently, new whole-genome data, especially from unsampled classes (i.e. Monoplacophora, Scaphopoda and “Aquiferans”) will unarguably represent an essential step towards molluscan evolutionary studies and help the clarification of molluscan inter- and intrarelationships. Additionally, molluscs´ genomes through the identification of conserved gene families may further contribute to explore early Lophotrochozoan and even bilaterian evolution (Simakov et al., 2012; Takeuchi et al., 2012, 2016; S. Wang et al., 2017).

In addition, molluscan genomes are also essential for understanding biomineralization of hard structures and pearl formation in bivalves (Adema et al., 2017; Du et al., 2017; Kocot et al., 2016a; Takeuchi et al., 2012, 2016; Zhang et al., 2012), as well as to explore the molecular mechanism of several taxa-specific novelties such as: doubly uniparental inheritance of mitochondrial DNA in many bivalves species, in which biparental inheritance of mitochondrial DNA is frequently observed and differently segregated within male and female’s organism; (Capt et al., 2018; Ghiselli et al., 2018, 2019; Murgarella et al., 2016; Renaut et al., 2018; Zouros, 2013); kleptoplasty of the sacoglossan sea slugs (Cai et al., 2019); symbiotic relationships with microorganisms carried out by specialized symbiotic organs of the bobtailed squid (Belcaid et al., 2019); among many other interesting lineage-specific adaptations.

Numerous relevant questions, some mentioned above, can justify the importance and relevance of molluscan genome sequencing projects. Yet, the generation of answers to those questions is still in the “embryonic stage”. In fact, producing an assembled and annotated genome, alone, is time-consuming, demanding a tremendous amount of specialized effort and structural resources (Dominguez Del Angel et al., 2018). Consequentially, it is frequent to see genome reports that simply characterize the genome assembly and annotation, pointing the direction for future studies without deep exploration of the genome data potential (Cai et al., 2019; C. Li et al., 2018; Nam et al., 2017; Renaut et al., 2018; Schell et al., 2017; Zarrella et al., 2019). Nevertheless, some molluscan genome projects, generally through *multi-omic* approaches (especially transcriptomic), have paved the way for a deeper exploration of these biological questions. Genome studies have relied on comparative analysis of genome structure, organization and genomic composition, as well as, identification of expansion and levels of expression of several gene families in order to identify genomic signatures that may explain some of the molluscan most fascinating features. In fact, these strategies have revealed to be extremely efficient and, in most cases, have resulted in the identification of interesting novelties. In an effort to produce new insights of the neural complexity of cephalopods Albertin et al., (2015) were able to identify two highly expanded gene families, that regulate neural development, and were thought to be exclusive of vertebrates. In a similar way Calcino et al., (2019) while exploring for the genetic signature of adaptation to freshwater lifestyle in the quagga mussel, was able to identify a novel highly expressed aquaporin water channel. This new aquaporin, which was called lophotrochoaquaporin, was associated with the five distinct moments of colonization of freshwater ecosystems by bivalves, highlighting its importance for molluscan freshwater colonization. In several bivalve species, studies have reported the expansion of genes families related to adaptation to environmental settings, innate immune systems response and pathogen recognition (Mun et al., 2017; Murgarella et al., 2016; Powell et al., 2018; Sun et al., 2017; Takeuchi et al., 2012, 2016; G. Zhang et al., 2012). Furthermore, in species that need to cope with tidal fluctuations, gene families related with hypoxia, oxidative stress and anti-apoptosis were also expanded (Mun et al., 2017; Murgarella et al., 2016; Powell et al., 2018; Sun et al., 2017; Takeuchi et al., 2012, 2016; Zhang et al., 2012). On the other hand, deep-sea mussels which are exposed to a very different set of environmental conditions, besides the expansion of immune response genes families, also revealed gene expansion related to protein structure stabilization, elimination of toxic substances, endocytosis and caspase-mediated apoptosis, all of which may be strongly implicated in the deep-sea survival success (Sun et al., 2017). Some genome studies have also explored the expansion and expression of genes related with biomineralization and nacre/shell matrix formation in bivalves (Du et al., 2017; Takeuchi et al., 2012, 2016) and in gastropods (Adema et al., 2017) (see Kocot et al., 2016a for overall review of molluscan biomineralization). In species with particular unique characteristics, such as scallops, expansion of energy-related genes families may explain the large adductor muscle adapted for swimming, light-sensing genes families essential for wide light spectral sensing in the scallop eyes, byssus related genes families for byssal secretion and sodium channel genes that may allow neurotoxins accumulation and transformation (Y. Li et al., 2017; S. Wang et al., 2017). The exploration of two venomous snails’ genomes has allowed the identification and characterization of structure and distribution of several conopepetides (Andreson et al., 2019; Barghi et al., 2016). As for invasive species, genes family expansion related to environmental sensitivity and adaptation, immunological defense, pathogenic resistance, and dietary adaptation were associated with these species' high ecological plasticity (Liu et al., 2018; Sun et al., 2019; Uliano-Silva et al., 2018). The genome survey of *B. glabrata*, a snail that is an intermediary host of a human parasite, allow the identification and description of noble features related to phero-perception, response to stressors, immunological functions and gene expression regulation that can ultimately be used to mitigate the transmission of parasites through the snail host (Adema et al., 2017). Comparative analysis of the molluscan karyotype features, chromosome structure, and rearrangements as well as expansion, expression, and position of conserved genes families (such as Hox and ParaHox, Wnt, G-protein-coupled receptor, nuclear receptors, actins, among other genes families) have also brought an incredible contribution to the study of Bilateria evolution. These studies have started revealing new insights into the origins of bilaterian organs, gene families, genetic pathways, evolutionary rates, and speciation processes. Additionally, the incremental number of molluscan genomes (as well as other unrepresented groups of bilaterians) are fundamental for listing ancestral bilaterian karyotype, chromosome rearrangements and genes families that may eventually allow a reconstruction of the ancestral urbilaterian chromosomes (Adema et al., 2017; Y. Li et al., 2017; Simakov et al., 2012; Sun et al., 2019; Takeuchi et al., 2016; S. Wang et al., 2017). On completely different approaches, the assembled molluscan genomes have also been used to identify microsatellites (Peñarrubia et al., 2015a, 2015b), for variance detection and bloodstock pedigree assignment (Kijas et al., 2019; Masonbrink et al., 2019; Nguyen et al., 2014) and characterization of microRNAs (miRNA) as tools for phylogenetic inference (Kenny et al., 2015).

Mollusc genome assemblies have been critical tools to address long-lasting questions of diversity, evolution and molecular signatures of phenotypic adaptations. Nevertheless, these studies have only explored a small fraction of these features and although new mollusc genomes are necessary, the ones already available represent extremely valuable tools that can be used to explore many of these fascinating questions.

6. Final remarks

This investigation shows that despite the species richness of Molluscs, the available genomic resources are by comparison relatively scarce. Yet, we found a stable increase of published *omic* papers in the past decade. Importantly, we also found a high degree of aim homogenization, probably reflecting that the use of these genomic tools is in its infancy and will become more common once better established and also beneficiating with the growing support from the bioinformatics resources. Furthermore, we note a biased sampling in the characterized molluscan species with the vast majority belonging to the more diverse gastropods and bivalves, with very few being available for cephalopods, and even less or none for other lineages. Finally, in addition to the necessity of producing novel datasets from key molluscan lineages, the already available resources show an immense potential to be explored in comparative and functional genomic analyses.

Acknowledgments

AGS and MLL were funded by the Portuguese Foundation for Science and Technology (FCT) under grants SFRH/BD/137935/2018 and SFRH/BD/115728/2016 respectivelly. This research was developed under ConBiomics: the missing approach for the Conservation of freshwater Bivalves Project Nº NORTE-01-0145-FEDER-030286, co-financed by COMPETE 2020, Portugal 2020 and the European Union through the ERDF, and by FCT through national funds. Additional, Strategic Funding provided by FCT UID/Multi/04423/2019.

Supplementary material

Below is the link to the electronic supplementary material.

https://doi.org/10.1007/s10750-019-04111-1

Table P1.S1 – List of all the peer-reviewed publications resulting from searching the Web of Knowledge by querying the following terms: “RAD”, “RAD-seq”, “mitogenome” and “transcriptome” associated with Mollusca, mollusk, mollusc, mollusks, gastropod, Gastropoda, bivalve, Bivalvia, Cephalopoda, cephalopods, Caudofoveata, Aplacophora, Polyplacophora, Monoplacophora and Scaphopoda.

1.2. Paper 2 – Molluscan mitochondrial genomes break the rules

Ghiselli, F., Gomes-Dos-Santos, A., Adema, C.M., Lopes-Lima, M., Sharbrough, J., Boore, J.L., 2021. Molluscan mitochondrial genomes break the rules. Philosophical Transactions of the Royal Society B: Biological Sciences B376,20200159-20200159. https://doi.org/10.1098/rstb.2020.0159

Molluscan mitochondrial genomes break the rules

Fabrizio Ghiselli [1]*, **André Gomes-dos-Santos** [2], Coen M. Adema [3], Manuel Lopes-Lima [4], Joel Sharbrough [5] and Jeffrey L. Boore [6]

[1] Department of Biological, Geological and Environmental Sciences, University of Bologna, Italy; [2] CIIMAR, Interdisciplinary Centre of Marine and Environmental Research, and Department of Biology, Faculty of Sciences, University of Porto, Portugal; [3] Center for Evolutionary and Theoretical Immunology, Department of Biology, University of New Mexico, Albuquerque, USA; [4] CIBIO/InBIO, Research Center in Biodiversity and Genetic Resources, University of Porto, Vairão, Portugal; [5] Department of Biology, Colorado State University, Fort Collins, USA; [5] Providence St Joseph Health and the Institute for Systems Biology, Seattle, USA

* Corresponding author

DOI: https://doi.org/10.1098/rstb.2020.0159

Abstract

The first animal mitochondrial genomes to be sequenced were of several vertebrates and model organisms, and the consistency of genomic features found has led to a ‘textbook description’. However, a more broad phylogenetic sampling of complete animal mitochondrial genomes has found many cases where these features do not exist, and the phylum Mollusca is especially replete with these exceptions. The characterization of full mollusc mitogenomes required considerable effort involving challenging molecular biology but has created an enormous catalogue of surprising deviations from that textbook description, including wide variation in size, radical genome rearrangements, gene duplications and losses, the introduction of novel genes, and a complex system of inheritance dubbed ‘doubly uniparental inheritance’. Here, we review the extraordinary variation in architecture, molecular functioning and intergenerational transmission of molluscan mitochondrial genomes. Such features represent a great potential for the discovery of biological history, processes and functions that are novel for animal mitochondrial genomes. This provides a model system for studying the evolution and the manifold roles that mitochondria play in organismal physiology, and many ways that the study of mitochondrial genomes are useful for phylogeny and population biology.

This article is part of the Theo Murphy meeting issue ‘Molluscan genomics: broad insights and future directions for a neglected phylum’.

Keywords

Mitochondria; Mollusc; Genome; Evolution; Doubly Uniparental Inheritance.

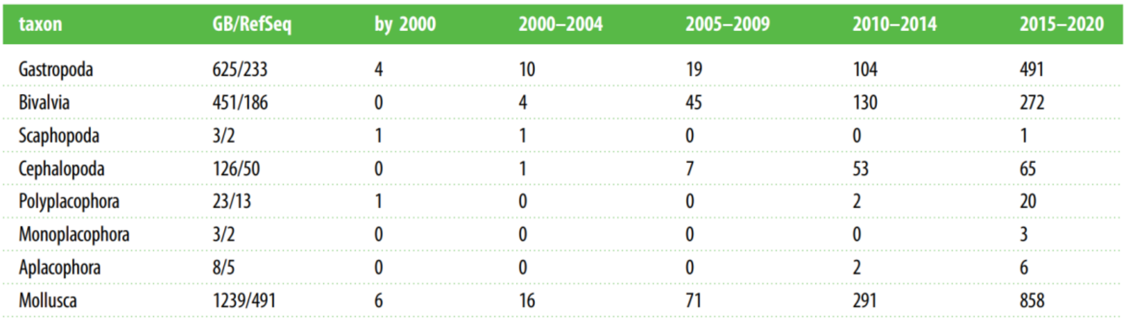
1. Introduction

In the 1980s, as DNA sequencing was becoming common, the fledglings of what we now call ‘genomics’ were diminutive animal mitochondrial genomes. The first reports were of several vertebrates and model organisms, followed quickly by studies of their modes of replication, transcription, RNA processing and other aspects of molecular biology (see Shadel and Clayton, 1997). The consistency of genomic features found and the expectation that these studies were characteristic of all mitochondrial genomes has led to a ‘textbook description’ of mitochondrial genomes that includes a consistent size of about 16 kb, strictly maternal inheritance, a content of 37 genes (encoding 13 proteins, 2 rRNAs and 22 tRNAs) compactly organized in a nearly invariant arrangement, a single large non-coding ‘control region’ with signals for regulating replication and transcription, and transcription of a single polycistron from each strand that is processed by enzymatic removal of tRNAs into gene-specific (or, in the cases of *nad4L-nad4* and/or *atp8-atp6*, bicistronic) mRNAs. Secondary structures were sometimes inferred for regulatory signals or to compensate for lack of tRNA genes where necessary for enzymatically separating the adjacent gene-specific transcripts.

Clearly, understanding these features is important for interpreting the patterns of evolution of these genomes, but this touches also on many other issues, including interactions with the products of nuclear genes, energy generation, wide-ranging aspects of metabolism and physiology, stress tolerance, susceptibility to oxidative stress, aspects of ecology, patterns of inheritance and population genomics. A more broad phylogenetic sampling of complete mitochondrial genomes now belies not only these general genomic features, but also makes clear that there is no potential for some of these functional molecular mechanisms.

Among bilaterian animals, the phylum Mollusca is especially replete with such examples. Due to their modest size and considerable phylogenetic information content both in gene sequences and arrangements, molluscan mitogenomes began to be studied in the early 1990s. Then, characterization of full mitogenomes required considerable effort involving challenging molecular biology including physical isolation of mitochondrial DNA (mtDNA), restriction enzyme mapping, cloning of large inserts, subcloning into a large number of separate plasmid vectors, and Sanger sequencing by directed primer walking, as evident from the first reports of molluscan mitogenomes from *Mytilus edulis* (Bivalvia: Hoffmann et al., 1992), *Katharina tunicata* (Polyplacophora: Boore and Brown, 1994) and several helicid gastropods (Hatzoglou et al., 1995; Terrett et al., 1996; Yamazaki et al., 1997) (Table P2.1). The revolutions in genome sequencing technology since have greatly accelerated these efforts, and we now have available more than 1000 complete mitochondrial genome sequences from more than 700 species. This, plus a modest amount of work to understand the biology of these genomes, has created an enormous catalogue of surprising deviations from that textbook description, including wide variation in size, radical genome rearrangements, gene duplications and losses, the introduction of novel genes and a complex system of inheritance dubbed ‘doubly uniparental inheritance’ (DUI). This creates great potential for the discovery of biological history, processes and functions that are novel for animal mitochondrial genomes. Interestingly, expanded non-coding regions, variable repeat content, frequent gene rearrangements and large numbers of ORFans, while uncommon in other animal lineages, are frequently observed in plants (Wendel et al., 2012).

Table P2. 1 - Number of molluscan mitogenome sequences in GenBank over time. The GenBank (GB) search was structured as follows: ((‘Mollusca’[Organism]) AND (biomol_genomic[PROP] AND mitochondrion[filter] AND (‘8000’[SLEN]: ‘100000’[SLEN]) AND (‘1900/01/01’[PDAT]] : ‘1999/12/31’[PDAT])). The term ‘Mollusca’ was replaced for family-level searches with ‘Gastropoda; Bivalvia; Scaphopoda; Cephalopoda; Polyplacophora; Monoplacophora; Aplacophora’ and the years were adjusted for specific time intervals.



taxon	GB/RefSeq	by 2000	2000–2004	2005–2009	2010–2014	2015–2020
Gastropoda	625/233	4	10	19	104	491
Bivalvia	451/186	0	4	45	130	272
Scaphopoda	3/2	1	1	0	0	1
Cephalopoda	126/50	0	1	7	53	65
Polyplacophora	23/13	1	0	0	2	20
Monoplacophora	3/2	0	0	0	0	3
Aplacophora	8/5	0	0	0	2	6
Mollusca	1239/491	6	16	71	291	858

2. Genome architecture

The first mollusc mitochondrial genome (Hoffmann et al., 1992), sequenced nearly three decades ago with Klenow fragment of *Escherichia coli* DNA polymerase on polyacrylamide gels, documented unprecedented genome architectural variation compared to other metazoans and presaged the amazing variation in mollusc mtDNA genome architecture that was soon to be discovered. Several major patterns of molluscan mitochondrial genome biology were largely present, if not fully understood, in that original *M. edulis* mtDNA. This included, to wit, a dramatic departure in gene synteny from other invertebrate mitochondrial genomes, with all genes encoded on one strand, the presence of DUI, not recognized until 1994 (Skibinski et al., 1994; Zouros et al., 1994), and the seemingly missing ATP synthase gene *atp8* (and the subsequent question of whether bivalves actually have it (Breton et al., 2010) or not (Uliano-Silva et al., 2016).

(a) Extensive natural variation

Mollusc mitochondrial genomes vary widely in size. The smallest reported so far belong to the heterobranch gastropods at approximately 13.6–14.1 kb (e.g. DeJong et al., 2004; Feldmeyer et al., 2010; Grande et al., 2002; Hatzoglou et al., 1995; Kurabayashi and Ueshima, 2000; Terrett et al., 1996; White et al., 2011) and the scaphopods (Boore et al., 2004; Dreyer and Steiner, 2004). These are only slightly larger than the smallest animal mitochondrial genomes (Pett et al., 2011), but still contain all 37 genes typical of metazoan mtDNAs, including 13 protein-coding genes, 22 tRNAs and 2 rRNAs, as well as a putative control region (Kurabayashi and Ueshima, 2000). Not unexpectedly, these compact mitochondrial genomes feature high levels of overlapping gene boundaries. The largest mtDNAs come from the scallop *Placopecten magellanicus* (up to 42.0 kb, Snyder et al., 1987) and the Arcidae clams, with *Scapharca broughtonii* ranging up to approximately 51.0 kb (Y. G. Liu et al., 2013) and a recent report claiming that the *S. kagoshimensis* mitochondrial genome is approximately 56.2 kb in length (Kong et al., 2020). The *S. broughtonii* mtDNA (and that of *S. kagoshimensis*, if verified) represents the largest animal mitochondrial genome yet recorded out of approximately 86 900 mtDNAs from more than 11 600 species present on NCBI. In both scallops and ark shells, the large genome sizes are not primarily a result of duplications or longer intergenic regions, but rather of expansion of the largest non-coding region (la Roche et al., 1990; Y. G. Liu et al., 2013), as is commonly the case for size variation in other mollusc mtDNAs (Figure P2.1). These bivalves are all exceptionally long-lived, especially the Arcidae, raising the question of whether long generation times affect the pace of evolutionary change in mitochondrial genome size, although other long-lived molluscs (e.g. abalone) do not share similar expansions of their mitochondrial genomes (Maynard et al., 2005).

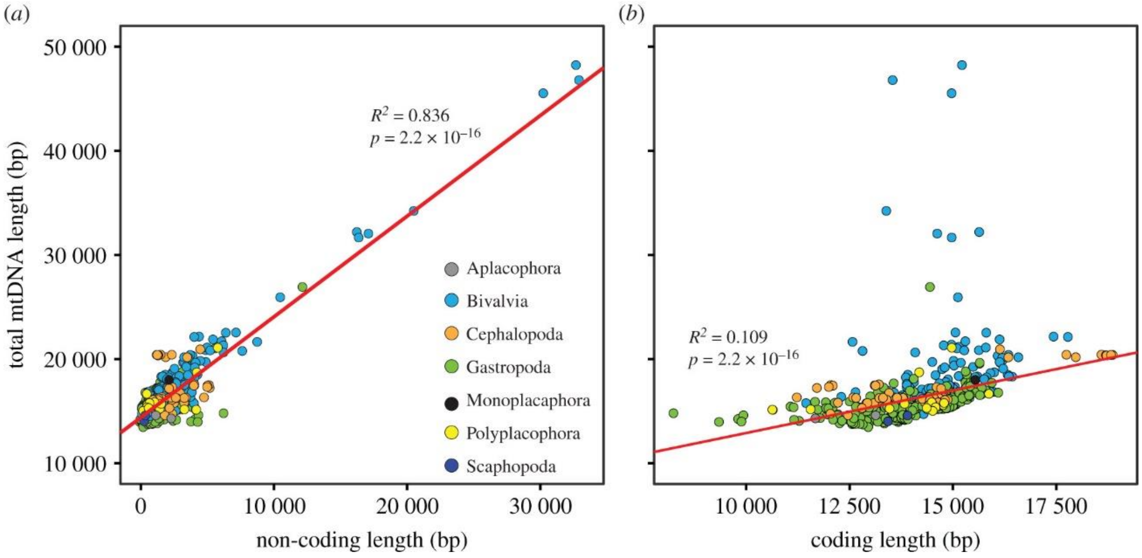


Figure P2. 1 - Relationship between the length of (a) non-coding and (b) coding regions on total mtDNA length in molluscan classes. Variation in non-coding length explains a greater proportion of variation in total mtDNA length compared to variation in coding length. Each circle represents a single species. When multiple mtDNAs were available for a single species, the mean across all individual records was taken as the species value. Colours represent different molluscan classes and are indicated by the key in (a).

Molluscan mitochondrial genomes have substantial variation in nucleotide composition skew asymmetry (i.e. heavy versus light strand, Francino and Ochman, 1997). Strand asymmetry occurs when there are more purines (i.e. adenine and guanine) on one DNA strand than there are pyrimidines. The strand with more purines than pyrimidines is heavier and, therefore, moves farther along in caesium chloride density gradient centrifugation when separated than the complementary strand (Brown et al., 2005) and is therefore termed the heavy or ‘H’ strand, and the other the light or ‘L’ strand. This skew is thought to be caused by the bias in types of spontaneous mutations that occur in single-stranded DNA (i.e. heavy versus light strand, Francino and Ochman, 1997), a condition that occurs for the displaced strand during transcription or replication (see a characterization in Boore, 2006a, a process known to be unusually slow for mtDNA Clayton, 1982). The degree of nucleotide skew is particularly large around the control region, as this region is found in single-stranded conformation more commonly than the rest of the molecule. There have been numerous reversals of strand asymmetry in molluscs (Sun et al., 2018), likely as a result of inversions in the control region, which contains one or both origins of replication (Fonseca et al., 2014; Hassanin et al., 2005).

Molluscs have experienced many changes in the transcriptional orientations (i.e. inversions) of genes, placing them variously on strands of differing nucleotide composition skews. For example, some taxa have all genes on one strand, like all marine bivalves (e.g. scallops, oysters and clams: Danic-Tchaleu et al., 2011; Smith and Snyder, 2007; Y. Yuan et al., 2012) and all protein-coding genes of caenogastropods (Márquez et al., 2014), while others do not, such as unionid mussels (Doucet-Beaupré et al., 2010), heterobranchs (Feldmeyer et al., 2010), vetigastropods (Xin et al., 2011), cephalopods (Akasaki et al., 2006; Boore, 2006b), scaphopods (Boore et al., 2004), aplacophorans (Osca et al., 2014) (but see Mikkelsen et al., 2018, in which all sequenced genes of the *Spathoderm clenchi* mtDNA are on the same strand), monoplacophorans (Stöger et al., 2016) and polyplacophorans (Boore and Brown, 1994). More generally, changes in genome architectures that alter transcriptional patterns across lineages are common and appear to be largely mediated by tRNA transposition and inversion (Feldmeyer et al., 2010), as the secondary structures are hypothesized to form transcriptional barriers (Fernández-Silva et al., 2003) and RNA cleavage signals (Ojala et al., 1981).

Indeed, changes in the gene order are most common for tRNAs. Even families like Haliotidae that exhibit largely conserved synteny of the protein-coding genes exhibit variable tRNA locations (Xin et al., 2011). Duplication of tRNAs appears to be a major contributor to mitochondrial genome rearrangement, as expected for the ‘duplication-random loss model’, with evidence that many molluscs contain extra tRNAs (Y. G. Liu et al., 2013; Smith and Snyder, 2007) beyond the minimal set of the 22 essential for accommodating the ‘super-wobble’ of mitochondrial translation. Interpreting this pattern of tRNA translocations is complicated by cases of remoulding of tRNA anticodons, which occurs sporadically throughout molluscs (Guerra et al., 2018; X. Wu et al., 2012a, 2015) and otherwise (Cantatore et al., 1987). The cases where a single amino acid is specified by two different codon families (serine and leucine) are especially susceptible to this because a switch of anticodons alone would be sufficient since these tRNAs would each have the necessary internal signals for charging with the correct amino acid (Higgs et al., 2003; Rawlings et al., 2010).

Still, there has been a large number of rearrangements of the genes encoding proteins or rRNAs, often via tandem duplication (Nolan et al., 2014; Xie et al., 2019; Xu et al., 2010) or large-scale inversions (e.g. vetigastropods Xin et al., 2011, versus caenogastropods Bandyopadhyay et al., 2006). In contrast with Vertebrata and Arthropoda, in which gene arrangements have remained generally very stable, extensive gene order rearrangements have been documented in every major lineage within Mollusca, including caenogastropods (Rawlings et al., 2001), scaphopods (Dreyer and

Steiner, 2004), cephalopods (Akasaki et al., 2006), heterobranchs (Grande et al., 2008), bivalves (Lopes-Lima et al., 2017a; Zheng et al., 2010), aplacophorans (Mikkelsen et al., 2018; Osca et al., 2014), polyplacophorans (Irisarri et al., 2014) and monoplacophorans (Stöger et al., 2016). The extent of this variation has understandably added complexity to inferring ancestral gene order, as until recently many lineages were too lightly sampled to accurately infer evolutionary paths (e.g., Yokobori et al., 2004, versus Luo et al., 2015; Uribe and Zardoya, 2017).

Across animal life, in nearly all lineages, there has been strong selection to maintain the minimal set of 37 genes (but see Lavrov and Pett, 2016). With the possible exception of *atp8* in bivalves (Breton et al., 2010; Uliano-Silva et al., 2016), the genes encoding proteins or rRNAs are seldom lost and duplicates are rarely maintained for long periods in molluscs (but see Kawashima et al., 2013; S. T. Williams et al., 2017; X. Wu, et al., 2012a), and molluscan mtDNAs rarely contain fewer than the necessary minimal set of 22 tRNAs (but see X. Wu et al., 2009). There has long been speculation about the selection pressures that are responsible for this (Adams and Palmer, 2003), including suggestions that hydrophobic proteins cannot easily move across membranes, that these proteins may be destructive in the cytoplasm, or that there is value in regulating mitochondrial function with this genome that is a remnant of its prokaryotic ancestor (Adams and Palmer, 2003; Allen et al., 2003; Timmis et al., 2004).

Additions to the mitochondrial genetic repertoire are uncommon but, here too, molluscs provide many of the exceptions. For example, lineage-specific open reading frames (ORFs) have been identified in bivalves that exhibit DUI (Breton et al., 2011a), of which the male version in *Ruditapes philippinarum* was proposed to be virally derived (Milani et al., 2014b). Additionally, there is evidence of nuclear-derived genes inserting into the mitochondrial genome. For example, a novel ORF was discovered with no sequence- or domain-based homology to the rest of the mitochondrial genome of the pearl-lip oyster *Pinctada maxima* but has domain-based homology to the nuclear genome (X. Wu et al., 2012b). The mitochondrial genome of the Arcidae clam *Tegillarca granosa* contains 32 novel ORFs, none of which have any homology to the rest of the mitochondrial genome, and eight of which are predicted to have signal peptides, a hallmark of nuclear but not organellar genes (Sun et al., 2015).

Early studies of transcription and translation in mitochondrial systems showed cases where the adjacent gene pairs *atp8-atp6* and *nad4L-nad4* were not enzymatically separated as mRNAs (see more below and Simon and Faye, 1984) and, instead, were separately translated into proteins by initiation on the ribosome, sometimes at the beginning of this bicistron and other times at an internal codon (Barros and Tzagoloff, 2017; Rak and Tzagoloff, 2009; Zeng et al., 2007). Perhaps this is due to difficulties with translating the very small mRNAs from *atp8* and *nad4L*. Early mitogenome sequencing revealed that these pairs were adjacent even in cases of more highly rearranged genes, suggesting this as a universal molecular process. But some molluscs do not have *atp8-atp6* as adjacent (Bettinazzi et al., 2016; Boore, 2006b; Grande et al., 2008; He et al., 2011) and others do not have *nad4L-nad4* as adjacent (polyplacophorans (Boore and Brown, 1994), heterobranchs (Feldmeyer et al., 2010), scaphopods (Boore et al., 2004; Dreyer and Steiner, 2004), unionid mussels (Doucet-Beaupré et al., 2010; Xin et al., 2011), cephalopods (Akasaki et al., 2006; Boore, 2006b), aplacophorans (Osca et al., 2014), monoplacophorans (Stöger et al., 2016) and gastropods (Grande et al., 2008)), indicating that there must be other modes of translation and regulation.

Not only are gene rearrangements rampant in mollusc mitochondrial genomes, but even individual genes exhibit remarkable architectural variation. Perhaps most prominent among these is the splitting of the large ribosomal rRNA gene (*rrnL*) into two distinct genes in *Crassostrea* oysters (Milbury and Gaffney, 2005). The resulting transcripts do not appear to be spliced together into a single RNA, but the ribosome itself appears to be fully functional (Milbury et al., 2010). The partially duplicated *rrnL* and *rrnS* genes of the vermitid snail *Thylacodes squamigerus* mitochondrial genome bear a superficial resemblance to *Crassostrea's* split *rrnL*, but the fragments appear to be pseudogenes (Rawlings et al., 2010).

Evidence for variation in genic architecture also comes from an intriguing case of apparently convergent evolution of the male-specific version of *cox2* in bivalves exhibiting DUI (see more below). In Mytilidae, *cox2* is extended at the 3′ end of the transcript (Curole and Kocher, 2002), but in some Veneridae, *cox2* has a male-specific insertion in the middle of the gene (Bettinazzi et al., 2016). It is unclear whether these *cox2* modifications share similar functions, although the former was hypothesized to have a role in reproduction (Chakrabarti et al., 2007). Finally, tRNAs are commonly found to have truncated D arms, especially in the heterobranchs (Sevigny et al., 2015), and there is even a case in which a tRNA has been inserted into *nad5* (Sun et al., 2014). These evident departures from the typical mode of intense purifying selection acting on mitochondrial genes likely represent lineage-specific mitochondrial adaptations and more work is required to understand their functional importance.

The largest non-coding region, inferred to perform the functions of the ‘control region’, varies widely in location also; see, for example, its varying positions in *Mytilus* [2] versus scallops (Rigaa et al., 1995), squid (Sasuga et al., 1999; Tomita et al., 2002) and caenogastropods (McComish et al., 2010). And the content and structure of control regions are vastly different across the major molluscan lineages, with high rates of evolutionary turnover by novel tandem duplications, often of previously duplicated regions (Akasaki et al., 2006; Simison et al., 2006; Snyder et al., 1987; Sun et al., 2015; Xu et al., 2010; Zbawicka et al., 2014); transpositions, especially of tRNAs, into this region (Breton et al., 2006; Cao et al., 2009; Y. G. Liu et al., 2013; Smith and Snyder, 2007; Sun et al., 2015); and newly evolved simple sequence repeats such as poly(AT) tracts (Brauer et al., 2012; Gao et al., 2018). Together these primary sequence features share the ability to produce secondary structures including stem-loop (X. C. Huang et al., 2013; Y. Yuan et al., 2012), cloverleaf (Bernt et al., 2013a; Grande et al., 2008; Sevigny et al., 2015; Zhu et al., 2012) and cruciform (McComish et al., 2010) structures in the control region, which in other organisms appear to be related to mtDNA replication and transcription (Grande et al., 2008; Shadel and Clayton, 1997).

Some control regions provide especially valuable insight into the biology and evolution of mitochondrial genome architecture. For example, squid control regions harbour relics of tandemly triplicated whole mitochondrial genomes, followed by their subsequent loss (Jiang et al., 2015; Kawashima et al., 2013; Sasuga et al., 1999; Tomita et al., 2002; Uribe and Zardoya, 2017). Heterobranchs have extremely short control regions, reflecting their compact mitochondrial genomes (Kurabayashi and Ueshima, 2000), while caenogastropods have control regions of variable length with an inverted repeat interspersed by a simple sequence repeat (Bandyopadhyay et al., 2006; McComish et al., 2010). Control regions of mussels exhibiting DUI have lineage-specific, tripartite control regions consisting of two variable domains interspersed by a conserved domain (Cao et al., 2009). Recombination between the F-type and M-type control regions in which an F-type mtDNA acquires an M-type control region appears to coincide with the masculinization of F-type mtDNAs (Breton et al., 2006; Burzyński et al., 2003; Stewart et al., 2009; Zouros, 2000); see DUI section below for more details). Thus, although control regions are often omitted in mitochondrial genome assemblies, generally because of technical difficulties in amplifying or sequencing these regions, those that have been sequenced provide rich sources of information for understanding evolution of mitochondrial genome architecture.

(b) Moving forward to understand the processes that contribute to variation in mitochondrial genome architecture

This rich phenomenological record described above makes for an ideal system in which to investigate the underlying molecular, genetic and evolutionary mechanisms contributing to and maintaining variation in genome architecture. Based on this diversity, a few themes have emerged that warrant further investigation. First, tRNA-mediated changes in gene order have been observed across Metazoa (Rawlings et al., 2003). It is hypothesized that at least part of this pattern results from accidental incorporation of tRNAs into the mtDNA when they moonlight as primers for DNA replication (Cantatore et al., 1987). This hypothesis is attractive because it would also help explain why control regions often feature pseudo-tRNAs (e.g. scallops, oysters and clams Cao et al., 2009; Smith and Snyder, 2007; Sun et al., 2015) and other tRNA-like secondary structures (Bernt et al., 2013a; G. G. Brown et al., 1986; Grande et al., 2008; Sevigny et al., 2015; Zhu et al., 2012). Misincorporation of tRNAs might also contribute to the high rates of evolutionary turnover in the control region, as new tRNA incorporation events push older sequences out of the control region. Complicating our understanding of this process are the evolutionary histories of tRNAs, as tRNA remoulding can obscure tRNA evolutionary history (see above). Quantifying the extent of tRNA duplication and remoulding, as well as rates and patterns of control region turnover in molluscan mitochondrial genomes, will provide valuable insight into tRNA-mediated genome architectural change.

Second, tandem duplication, which has been implicated in several molluscan genome rearrangements (e.g. Y. G. Liu et al., 2013; Snyder et al., 1987, cephalopods: Akasaki et al., 2006), can happen through a variety of mechanisms (Boore, 2000; Ludwig et al., 2000) including slipped-strand mispairing (Levinson and Gutman, 1987), imprecise termination of replication (Macey et al., 1997; Stanton et al., 1994), dimerization (Raimond et al., 1999) and illegitimate or non-homologous recombination between repeats (Kajander et al., 2000; Mita et al., 1990). Support for the role of tandem duplication in shaping mitochondrial genomes is undermined by the scarcity of animal mitochondrial genomes that harbour duplicated copies of protein-coding genes (Boore, 1999). It may be that duplicates are lost quickly, perhaps responding to selection favouring the maintenance of cytonuclear stoichiometry (Sharbrough et al., 2017). Evaluating these various possibilities will require better population-level sampling, especially with the help of long-read sequencing technologies like PacBio or Oxford

Nanopore, which can help resolve tandem duplications (Calcino et al., 2020; Ji et al., 2019).

Third, inversions are perhaps the most commonly retained form of structural rearrangements in molluscan mitochondrial genomes (see above paragraph on changes in transcriptional orientation). Inversions can arise via multiple double-stranded breaks or by inverted repeats (see Zampini et al., 2015 for the description of inverted repeat mechanisms) in which one repeat is deleted, likely via recombination (Lobachev et al., 1998). However, inversions would seem to have immediately deleterious consequences for transcriptional control of mitochondrial genomes. There has been speculation of an ‘evolutionary ratchet’, whereby genes rearranging by inversions to be on a single strand would eliminate the selective pressure to maintain transcription of the other strand and, once lost, would make any further inversion of any gene immediately non-functional such that reversion to a state of genes on both strands would be highly unlikely (Boore, 1999). Investigating mitochondrial transcriptional dynamics in closely related species (or M- versus F-type mtDNAs from the same species) that have inversions relative to one another might prove especially useful in understanding how inversions are able to persist longer than other types of mitochondrial genome rearrangements. How these inversions and subsequent changes in expression affect mitochondrial function and fitness will also be of broad interest to the mitochondrial community.

Fourth is the evident selective pressure for genome streamlining, both in terms of gene content and genome size. One of the more surprising observations of animal mitochondrial genomes is the degree to which genes overlap (Boore, 1999; Cheng et al., 2013; X. C. Huang et al., 2013). Overlapping mitochondrial ORFs often exhibit alternative reading frames (Boore, 1999), such that elongation of a gene via nonstop mutations may explain variation in the degree of gene overlap. Once genes do overlap, purifying selection is expected to be intense over the region, as mutations occurring in the overlap could have consequences for two separate genes. The greater degree of overlap between *nad4* and *nad6* in the M-type genome of *Solenaia carinata* compared to the F-type (X. C. Huang et al., 2013) raises the intriguing question of whether the increased intensity of selection engendered by gene overlap might compensate for the reduced efficacy of selection acting on male versus female transmitted mtDNAs (Camus et al., 2012). Comparing whether mitochondrial genomes with high versus low N_e (e.g. F-type versus M-type mtDNAs) have lesser degrees of genic overlap and reduced rates of deleterious mutation accumulation (Neiman and Taylor, 2009) would provide a powerful test of the forces contributing to genome streamlining of animal mitochondrial genomes.

Finally, the extent to which gene order can be used as an effective phylogenetic tool for molluscs (Allcock et al., 2011; Boore and Brown, 1998; Guerra et al., 2018; Uribe and Zardoya, 2017) depends upon low-level taxonomic sampling to infer rates and patterns of structural evolutionary change. The availability of more than 1000 molluscan mitochondrial genomes from over 700 different species as of September 2020 has largely solved that problem, especially for the bivalves (456 mtDNAs from 261 species), gastropods (452 mtDNAs from 358 species) and cephalopods (142 mtDNAs from 60 species). Such gene order analyses should not only take advantage of changes in major gene synteny but also of tRNA movements and inversion events. Together, these five avenues for future research represent central open questions in the evolution of mitochondrial genome architecture and should provide a framework for understanding how genome architecture contributes to mitochondrial function at molecular, cellular and organismal levels.

3. Annotation challenges

Considerable effort is required for annotation of the genes of molluscan mitogenomes. Most protein-encoding genes are easily identified with orthologues by sequencing similarity, with occasional consideration of hydrophobicity plots for *atp8* and *nad4L*, but there are challenges with inferring the correct start codon in cases where there are multiple, closely spaced alternatives. An inference must consider the possibility of overlap with the upstream gene and the extent of evolutionary conservation of the ORF. This is confounded by the fact that molluscs employ the invertebrate mitochondrial genetic code (NCBI Genetic Code 5) that allows for alternative start codons in addition to ATG, including ATA, ATY, TTG and GTG (normally encoding for methionine, isoleucine, leucine and valine, respectively). Each of these would provide a match to at least two nucleotides of the *trnM* anticodon (CAU), which must do double duty in most mitochondrial systems as the tRNA for both methionine and, in the case of protein initiation, formyl-methionine.

Ordinarily, inferring a stop codon for any gene is straightforward but, here too, mitochondrial genomes present a challenge. In many cases, mitochondrial genomes are transcribed as a single polycistronic RNA from each strand (see Garone et al., 2018).

The tRNA genes are then removed enzymatically, which liberates gene-specific mRNAs as proposed in the ‘punctuation model’ (Ojala et al., 1981). In the case of overlapping *atp8-atp6* and of *nad4L-nad4*, these have been shown for yeast (Simon and Faye, 1984), fish (Zardoya et al., 1995) and mammals (Fearnley and Walker, 1986) to remain as bicistrons that are translated on mitochondrial ribosomes, sometimes from the first codon and sometimes from an internal codon that initiates the second gene. In some other cases of adjacent protein-encoding genes without an intervening tRNA, there are potential secondary structures that have been speculated to serve this function (e.g. Boore and Brown, 1994). In many other cases, it remains unknown whether these mRNAs are separated or not. The specific challenge for gene annotation from genome sequence is that, after enzymatic processing to produce gene-specific messages, some will not have a complete TAG or TAA stop codon, but may terminate on just a TA or T that is completed to a TAA stop codon by polyadenylation of the transcript (Clary and Wolstenholme, 1985). Additionally, it is important to consider that some genes are known to overlap even when on the same strand, further complicating an accurate inference of the correct stop codon from genome data alone.

Of course, there are some cases where these features can be directly observed through the sequencing of expressed sequence tags (ESTs) (DeJong et al., 2004), providing the sequences of the full transcripts from which the genomic boundaries can be reliably determined. This has presented some surprises. For example, ORF analysis had predicted that *nad4* of the gastropod *Biomphalaria glabrata* mitogenome (NC_005439) was unusually long, fully overlapping with *trnT*, in contrast with the reported genes in the gastropods *Cepaea nemoralis* (NC_001816) and *Albinaria caerulea* (NC_001761). Independently determined EST data (AA547758) showed the cDNA for the C-terminus of *nad4* to end before the downstream *trnT* gene, more consistent with those of the other gastropods, and to terminate on a single T nucleotide that was extended by polyadenylation to form a TAA stop codon (DeJong et al., 2004).

Based on genome sequence alone, inferring the exact boundaries of rRNA genes is especially difficult. In fact, in most cases, there is simply the presumption that the rRNA gene extends to the boundary of the flanking genes, with this moderated by the extent of similarity matching to homologous genes of other organisms.

The genes for tRNAs diverge in sequence rapidly and are most commonly found by identifying potential secondary structures with a set of typical features (Chan and Lowe, 2019; Hahn et al., 2013; Lowe and Chan, 2016). Some lineages are known to have aberrant structures with some of the arms diminished or even missing, complicating this inference.

The rise of next-generation sequencing (NGS) has been a game-changer for the pace of generating complete mitogenome sequences. These methods generate an enormous number of short sequencing reads, leading to an increased reliance on computational methods for automated genome assembly. Among several alternative software packages that aim to assemble NGS data into large contigs, MITObim was specially designed for the assembly of mitogenomes (Hahn et al., 2013), as well as other tools that were released more recently (Al-Nakeeb et al., 2017; Dierckxsens et al., 2016; Meng et al., 2019). Using a provided mitochondrial genome or even a short (partial) gene sequence as an initial reference to identify sequence data of likely mitogenome origin, this program applies a strategy of BLAST and iterative mapping to select and assemble short reads from a large NGS dataset that provides adequate coverage into a linear representation of a mitogenome. Overlapping, identical sequence termini indicate that the assembly represents the full circular mitogenome. It is worth noting that reliance on computational interpretation of short sequence reads may potentially cause problems in assembling repetitive elements, such as the control region and unsuspected repetitive elements like tandem duplications or repeat regions, that may be resolved only by manual, targeted sequence characterization.

With such relative ease to derive the genome sequences, there is a greater demand for automated annotation. This need was recognized early on by the implementation of semi-automated annotation of genomes of organelles from mitochondria (and plant chloroplasts) through DOGMA (Dual Organellar GenoMe Annotator) that provided predictions of protein- and rRNA-encoding genes through BLAST similarities to previously annotated mitochondrial genomes (Wyman et al., 2004) and provided tools for manually refining the beginning and end of each gene. The identification of tRNA genes employed secondary structure predictions because mitochondrial tRNA sequences share little sequence similarity among animals. Generally, computational predictions were further hindered due to the aberrant structure of several molluscan tRNAs that do not conform to the canonical cloverleaf of animal tRNAs and typically required manual validation (Yamazaki et al., 1997). Current utilities include AGORA (prediction of protein-coding genes in a mitogenome assembly based on BLAST similarities to a reference mitogenome; Jung et al., 2018), MitoZ (Al-Nakeeb et al., 2017; Dierckxsens et al., 2016; Meng et al., 2019) and MITOS (Bernt et al., 2013b). The latter software performs de novo annotation of protein-encoding genes by sequence similarity and secondary structure predictions of both rRNA and tRNA. MITOS reports annotation results in the standardized format that supports the accepted, consistent nomenclature of mitochondrial genes. Updates (MITOS2 is available at http://mitos2.bioinf.uni-leipzig.de/index.py) have improved the prediction accuracy but the results still require manual curation.

Alternative start codons, the potential for incomplete stop codons and molluscan-specific tRNA structures continue to challenge automated annotation. Some possible challenges for annotation are shown in Figure P2.2, using *atp8* from gastropod mitogenomes as an example. *atp8* is the shortest protein-coding gene in mitogenomes and relatively variable among gastropod species, often not detected by BLAST and thus also not recognized by MITOS. Additionally, *atp8* of several gastropod species employs an alternative start codon, like ATT that normally encodes for an I (isoleucine), serving as start codon (specifying formyl-methionine) only at the initiation of protein translation. Automated gene finding and inexperienced annotators may fail to recognize ATT as a true start, choosing an upstream M-encoding nucleotide (ATG or ATA), even if part of a different gene, as an incorrect start codon. As a consequence, annotation of *atp8* often requires manual inspection and comparison to *atp8* from several species (Figure P2.2).

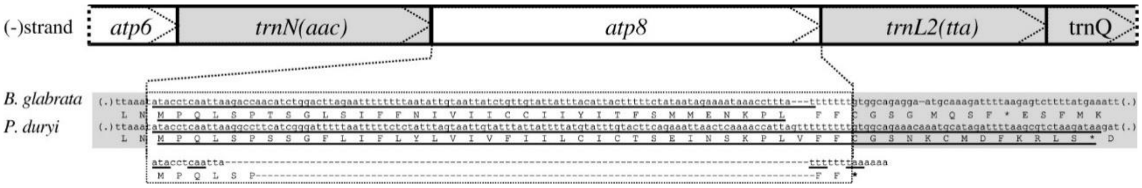

(-)strand
atp6
trnN(aac)
atp8
trnL2(tta)
trnQ
B. glabrata
P. duryi
(.)ttaaatatacctcaattaagaccaacatctggacttagaatttttttaatattgtaattatctgttgtattatttacattactttttctataatagaaaataaaccttta---ttttttgtggcagagga-atgcaaagattttaagagtcttttatgaaatt(.)
L N M P Q L S P T S G L S I F F N I V I I C C I I Y I T F S M M E N K P L F F C G S G M Q S F * E S F M K
(.)ttaaatatacctcaattaaggccttcatcgggatttttaatttttctctatttagtaattgtatttattattttatgtatttgtacttcagaaattaactcaaaaccattagtttttttttgtggcagaaacaaatgcatagattttaagcgtctaagataagat(.)
L N M P Q L S P S S G F L I F L Y L V I V F I I L C I C T S E I N S K P L V F F C G S N K C M D F K R L S * D
atacctcaatta----------ttttttaaaaaa
M P Q L S P----------F F *

Figure P2. 2 - In mitogenomes of planobid gastropods, the atp8 gene is bracketed by trnN(aac) and trnL2(tta). Shaded boxes, tRNA genes; white boxes, protein-coding genes; arrowheads indicate directionality; asterisk, stop codon. ORF analyses of the mitogenome sequences that ignore the concept of tRNA gene excision from polycistronic mitogenomic transcripts frequently yield incorrect prediction of protein-coding sequence intervals. Whereas the start codon is correctly indicated, the ORF for atp8 from *Biomphalaria glabrata* (underlined in both nucleotide and predicted amino acid sequences, NC_005439) falls short, despite an effort to accommodate an incomplete stop codon (T--). Another issue impacts the ORF selected from the *Planorbella duryi* mitogenome (KY514384). It comprises a (correct) start codon and TAA stop codon but overlaps with trnL2 and yields an unusually long protein sequence. For both snail species, considering the boundaries of the (MITOS predicted) tRNA genes, the ATA is the first possible start codon downstream from trnN. At the 3′ end, a single T nucleotide remains after excision of trnL2, completed by polyadenylation to a TAA (underlined) stop codon. Such peculiarities challenge prediction of multiple genes from molluscan mitochondrial sequences, as is evidenced in several GenBank entries, despite the purported curation of submissions by this NCBI database.

Re-evaluation and, if appropriate, updates by contributors of previous GenBank accessions will greatly benefit correct annotation.

A recent paper by Fourdrilis *et al*. (2018) provides a powerful set of criteria to integrate with automated MITOS prediction for correct annotation of gastropod (molluscan) mitogenomes. These criteria include the valid insights into molluscan mitochondrial biology, including the punctuation model, as well as alternative start and stop codons. We summarize these criteria below: (i) Protein-coding genes are assumed to begin at the first eligible in-frame start codon in their 5′ end, that is, the start codon nearest to the preceding gene without overlapping with it, checking that this start codon is suitable regarding the location and gene length by aligning the derived amino acid sequence with that of closely related species. (ii) Due to transcription of mtDNA as polycistronic RNA, it is considered physically impossible to have gene overlap between two protein-coding genes encoded on the same strand and in the same open reading frame, but possible if frames are different. (iii) Protein-coding genes are assumed to end at the first in-frame full stop codon, or an abbreviated stop codon (TA- or T- in invertebrates) ending immediately before the downstream tRNA; such an abbreviated codon results from the cleavage of the transcript at the 5′ and 3′ ends of tRNAs and tRNA-like secondary structures and is subsequently completed to a TAA stop codon with A residues by polyadenylation. (iv) Putatively duplicated genes are evaluated based on quality values provided in the MITOS analysis. (v) The boundaries of tRNA genes are those predicted by MITOS. (vi) The boundaries of rRNA genes were those predicted by MITOS and not extended to flanking genes to avoid overestimating rRNA gene length.

Despite these software packages for assistance and the attention of the scientific community, the entries for mitochondrial genomes at NCBI contain a great number of easily recognized annotation errors even in the ‘Refseq’ portion. Despite having this pointed out over a decade ago with specific, simple recommendations for systematically eliminating these and conducting quality control for new entries (Boore, 2006a), a recent study identified a great number of errors in a systematic search of complete vertebrate mitochondrial genomes at NCBI (Prada and Boore, 2019). To the best of our knowledge, no such systematic study has been made of annotations for complete mollusc mitogenomes, but there is no reason to suspect that they are immune from similar errors during submission or NCBI review (e.g. Fourdrilis et al., 2018). Consistent, accurate, complete annotation of these genomes is critical for comparative and phylogenetic studies. We urge NCBI to implement these simple quality control measures.

4. Inheritance: doubly uniparental inheritance in bivalves

Mitochondrial genomes follow a non-Mendelian inheritance pattern of being transmitted uniparentally in most eukaryotes; in animals, mitochondrial inheritance is usually strictly maternal (from now on: strictly maternal inheritance, SMI) (Barr et al., 2005; Birky, 2003). Perhaps the most striking feature of mollusc mitochondrial biology is the unique doubly-uniparental inheritance pattern so far reported in 100+ species of bivalves (Capt et al., 2020). In species showing DUI, two sex-linked mitochondrial lineages exist: one is inherited through eggs (F-type) the other through sperm (M-type). Differently from the cases of paternal mtDNA leakage reported in several organisms (Breton and Stewart, 2015), in DUI the sperm transmission route is stable across evolutionary time, so the F- and M-type coexist as segregated lineages for millions of years accumulating a remarkable sequence divergence. The F-M nucleotide p-distance ranges from 0.08 to 0.449, and the amino acid p-distance of mitochondrial protein-coding genes can reach 0.534 (Capt et al., 2020).

The dynamics and distribution of F- and M-type in embryos and tissues were first investigated in bivalves of the *Mytilus* species complex, in which DUI was observed for the first time (reviewed in Zouros, 2012). Particularly interesting was the finding that in early embryos (2–8 blastomeres) sperm mitochondria stained with MitoTracker Green showed two different distribution patterns: dispersed versus aggregate. The authors were also able to show a strong link between the pattern and the sex of the progeny: females were associated with the dispersed pattern, males with the aggregated one (Cao et al., 2004; Cogswell et al., 2011). These observations, together with the results of several molecular works, were used to build a first description of the mitochondrial dynamics in DUI, summarized below. Gametes are homoplasmic for the sex-specific type (F-type in eggs, M-type in spermatozoa), so upon fertilization the zygote is heteroplasmic and the fate of sperm mitochondria is tightly linked with sex. If the embryo develops into a female, the M-type mitochondria are dispersed and actively degraded as happens in some species showing SMI (Sato and Sato, 2017), and the animal will be homoplasmic for the F-type. Otherwise, if the embryo develops into a male, sperm mitochondria stay aggregated as they already are in the midpiece of sperm cells and are transported into the blastomere 4d, the precursor of the germline, and survive degradation; males are thus heteroplasmic, containing M-type in the germline and F-type in the somatic tissues.

The main points of this model are: (i) homoplasmy of females due to degradation of M-type; and (ii) heteroplasmy of males with retention of M-type due to the active segregation of sperm mitochondria aggregated in gonad precursors, but not in somatic tissues. A replicative advantage of M-type in males was also hypothesized, to explain its proliferation in spermatogenic tissues (Cogswell et al., 2011). This is still the most commonly used description of the DUI mechanism, but some revisions have become necessary. The existence of the two patterns was confirmed in a distantly related species (divergence time 400+ Ma), the venerid clam *Ruditapes philippinarum* (Milani et al., 2012), but as new data were gathered and new species analysed, evidence of deviations from the mechanism as described above started emerging. The presence of M-type in male somatic tissues is now known to occur in *R. philippinarum* (Ghiselli et al., 2011), *Venustaconcha ellipsiformis* and *Utterbackia peninsularis* (Breton et al., 2017) and in *Mytilus galloprovincialis* (Kyriakou et al., 2010; Obata et al., 2006).

These works showed also that heteroplasmy is more common than previously thought in both males and females of DUI species, and that the presence, abundance and distribution of the F- and M-types is quite variable across species, sexes and tissues. Such differences should be expected when dealing with a quantitative phenomenon like mitochondrial inheritance (Birky, 2003), especially across large evolutionary distances. Recently, immunohistochemistry and microscopy (both confocal and electronic) investigations on *R. philippinarum* showed the presence of heteroplasmy at the organelle level (both types present in the same mitochondrion) in male soma and, quite surprisingly, in undifferentiated germ cells of both sexes, while homoplasmy in both female and male gametes was confirmed (Ghiselli et al., 2019). According to these observations, the strict segregation of F- and M-type in gametes would be achieved during gametogenesis—thus much later in development than hypothesized before—and it was suggested that DUI is based on a mechanism of meiotic drive involving selfish genetic elements associated with mitochondria (Ghiselli et al., 2019; Milani et al., 2016).

(a) Doubly uniparental inheritance molecular mechanism

Hybrid and triploid DUI mussels have been shown to revert to SMI (Kenchington et al., 2009) and the taxonomic distribution of DUI species is scattered across bivalve phylogeny, so DUI must have evolved by the modification of a mechanism of SMI, but which one? There are several different mechanisms by which SMI can be achieved

(Birky, 1995; Sato and Sato, 2017), but that operating in bivalves is still unknown. Similarly to what happens in mammals, it was hypothesized that ubiquitination could be involved (Kenchington et al., 2002) and the results of some investigations seem to be consistent with such supposition (Diz et al., 2013; Ghiselli et al., 2012; Milani et al., 2013b; Punzi et al., 2018). A possible approach to understand which molecular mechanism is involved in DUI is to look at the differences between F- and M-type genomes, and numerous works have investigated this issue in the last 25 years. The main findings can be summarized as follows.

First, bivalve mtDNA shows an abundance of intergenic regions—or at least regions not containing known genes—and the largest are rich in genetic elements such as repeats, motifs and DNA/RNA secondary structures which differ between conspecific F and M genomes in DUI species (e.g. Cao et al., 2009; Ghiselli et al., 2017; Guerra et al., 2014; Passamonti et al., 2011; Robicheau et al., 2017). A strong clue supporting a role of control region elements in DUI comes from observations in the *Mytilus* complex. Several analyses on F- and M-type mtDNAs in *Mytilus edulis*, *M. galloprovincialis* and *M. trossulus* revealed the presence in male gonads of genomes having their coding sequences almost identical to those of the F genome (2–3% divergence). It was hypothesized that these genomes originated from F genomes that invaded the male germline and started to be transmitted through sperm, replacing the M-type and accumulating sequence divergence (which is initially reset to zero when the F-type replaces the M-type). This phenomenon was named ‘role-reversal’ or ‘masculinization’ (reviewed thoroughly in Zouros, 2012), and the aberrant F genomes transmitted through sperm have been defined as ‘masculinized’. Following studies found that the control regions of masculinized genomes contained parts of both the typical F- and M-type mtDNAs, being actually F/M chimaeras. Role-reversal has been observed, so far, only in the *Mytilus* complex. These findings strongly suggest that some elements located in the control region or its proximity have a role in the inheritance mechanism. The identity and the nature of these elements are still unknown and several candidates have been proposed, including DNA and/or RNA secondary structures (Ghiselli et al., 2013; Guerra et al., 2014), specific sequences/motifs (Kyriakou et al., 2015), or peptides encoded by ORFs located near the control region (see second point below).

The second feature that differentiates F and M genomes is the presence of lineage-specific ORFs showing no sequence similarity with known genes, and thus defined ‘ORFans’ (Breton et al., 2009, 2011a; Capt et al., 2020; Ghiselli et al., 2013; Guerra et al., 2019; Milani et al., 2013a, 2014b, 2016; Mitchell et al., 2016). In some cases, a protein product of these ORFans has been detected and localized (Breton et al., 2011b; Ghiselli et al., 2019; Milani et al., 2014b), but their function remains unknown despite extensive *in silico* analyses (Guerra et al., 2019; Milani et al., 2013a, 2014b, 2016; Mitchell et al., 2016). Such bioinformatics work has shown that, despite high evolutionary rates and large sequence divergences, all the analysed ORFans have similar predicted structural features, supporting a similar function. The involvement of the ORFans in the DUI mechanism is still a hypothesis and their mechanism of action is an object of speculation, but it is clear that these elements are maintained in bivalve genomes and some surely produce a novel mitochondrial protein. It would be surprising if these elements turn out to be non-functional.

Third, the cytochrome *c* oxidase subunit 2 gene (*cox2*) shows curious features in bivalves, and in several DUI species, there are important differences between the F-type and M-type *cox2 gene* (see also §2). The *cox2* gene is duplicated in the F-type of *R. philippinarum* (Ghiselli et al., 2013) and the M-type of *Musculista senhousia* (Passamonti et al., 2011), with paralogous copies showing different lengths. In some other cases, *cox2* has a different length in the two mtDNAs, due either to 3′ coding extensions (550 bp) or large in-frame insertions (up to 3.5 Kb) (Capt et al., 2020). It is still not clear if such modifications of *cox2* are linked to DUI for some functional reason, or are a more general feature of bivalve mtDNAs, maybe due to modifications in Complex IV of oxidative phosphorylation.

The fourth and last feature characterizing the differences between the two mitochondrial lineages concerns small non-coding RNAs (sncRNAs). Pozzi *et al*. (Pozzi et al., 2017) sequenced sncRNA libraries from gonads of *R. philippinarum*, and found miRNA-like sequences transcribed by intergenic regions for which a stable hairpin structure was predicted. *In silico* analyses showed that F and M genomes produce different mitochondrial sncRNAs with different nuclear targets. The authors hypothesized that such sncRNAs might affect nuclear gene expression through RNA interference and might influence gonad formation. More recently, Passamonti *et al*. (2020) reported *in vivo* clues of the activity of two sncRNAs in *R. philippinarum*. Small mitochondrial RNAs have also been predicted *in silico* in several species of amniotes (Pozzi et al., 2019), and in *Drosophila melanogaster*, *Danio rerio* and *Mus musculus* (Passamonti et al., 2020).

(b) MtDNA evolutionary patterns in doubly uniparental inheritance

It is still unclear how DUI emerged and why it has been maintained for hundreds of millions of years. Traits that last so long in evolution are usually maintained by natural selection because they have a function that affects organismal fitness. For this reason, and given the tight link between mitochondrial inheritance pattern and sex in DUI species, it was hypothesized that DUI has a role in sex determination and/or gonad differentiation (Breton et al., 2007, 2011b; Capt et al., 2018; Ghiselli et al., 2012; Milani et al., 2016; Passamonti and Ghiselli, 2009; Yusa et al., 2013; Zouros, 2012).

Studies on the patterns of molecular evolution of mitochondrial proteins in DUI bivalves clearly show that M-type evolves faster than F-type and both mtDNAs evolve faster than the mitochondrial genomes of other metazoans (Breton et al., 2007; Zouros, 2012). The reasons behind this pattern are the subject of debate. Relaxed selection is one possible explanation; Stewart *et al*. (Stewart et al., 1996) suggested that F- and M-type mtDNAs evolve under different degrees of selective constraints as a consequence of different ‘selective arenas’. Supposing that F-type mtDNA is functional in all somatic tissues and the female germline, while M-type functions only in the male germline, F-type would be subject to more stringent constraints, hence the faster sequence evolution of M-type. However, the more recent findings of F- and M-type distribution across tissues (discussed above), and the findings of M-type transcriptional activity in the soma (Breton et al., 2017; Milani et al., 2014a), may suggest that the above-mentioned arenas of function are not that distinct. Moreover, even if M-type mitochondria are functional only in the male germline, they have a crucial function of providing energy for sperm swimming. This is a fundamental function, especially in a broadcast spawning animal, and the relaxation of natural selection on such a trait could have long-term consequences on DUI species. Many DUI species are quite successful; for example, *R. philippinarum* is highly invasive, and *Arctica islandica* (in which DUI has been reported Dégletagne et al., 2016) is the longest-living non-colonial animal known (maximum reported lifespan approximately 507 years), so it seems that DUI is not manifestly disadvantageous.

A high-throughput analysis of mtDNA single nucleotide polymorphisms (SNPs) in F- and M-type of *R. philippinarum* (Ghiselli et al., 2013) revealed a similar amount of polymorphism in the two genomes, but a different distribution of allele frequencies (probably due to different bottleneck sizes), the M-type having a lower proportion of SNPs with a predicted deleterious effect. According to these data, the faster evolution of M-type is likely due to the roles of mitochondria in spermatogenesis and sperm motility, the latter being especially important in the intense sperm competition of an animal using broadcast fertilization. Indeed, one interesting feature of DUI is that mtDNA is under selection also for male functions, differently from what happens in all the SMI organisms, in which mitochondria are an evolutionary dead-end in males. This opens a series of interesting consequences and deserves thorough investigations. Recently, two comparative analyses of OXPHOS activity in gametes and somatic tissues of SMI and DUI bivalves reported a metabolic remodelling in M-type mitochondria that suggests an adaptive value of mtDNA variation, and a link between male-energetic adaptation, fertilization success and the preservation of paternally inherited mitochondria (Bettinazzi et al., 2020, 2019).

DUI is generally unknown or considered just a ‘freak of nature’, but it represents a unique and precious model to study mitochondrial biology and evolution. Thanks to its unusual features, it can be used as a tool to better understand mitochondrial heteroplasmy, inheritance, recombination, and the role of mitochondria in germline formation, meiosis, gametogenesis and fertilization, in some cases providing the exceptions that address general phenomena in other animal groups. Up to now, DUI has not been found outside bivalves, but, to the best of our knowledge, it has been specifically investigated in just five gastropod species (Gusman et al., 2017).

5. The utility and limitations of mitochondrial genomes for phylogeny

During the last three decades, mitochondrial markers, either individually, combined or as a whole, have been commonly used for phylogenetic reconstruction within Metazoa (Bernt et al., 2013a; Gissi et al., 2008; Kern et al., 2020; Stöger and Schrödl, 2013). This preference is due to several features that make mitochondrial sequences a well-suited and reliable molecular marker for phylogenetic assessment. First, all Metazoa (except some Loricifera Danovaro et al., 2010) possess a mitochondrial genome that can be obtained with relative ease compared with any particular genome region of similar size due to its high abundance and copy numbers within animal cells (Bernt et al., 2013a; Gissi et al., 2008). Second, gene orthology, essential for a successful phylogenetic assessment, is expected in the mitogenome, since genes from eventual duplication events shown to occur in molluscan mtDNA are rarely retained, and quickly lost or pseudogenized (Bernt et al., 2013a; Gissi et al., 2008; Kern et al., 2020). Furthermore, uniparental inheritance (see exception in bivalves in the DUI section above) and a general lack of recombination (Elson and Lightowlers, 2006) greatly favour the reliable inference of population structure. The variable substitution rates within the different genes/regions of the mitogenome grant a range of phylogenetic signals that might potentially be useful for accessing shallow and deep relationships (Bernt et al., 2013a; Gissi et al., 2008; Kern et al., 2020). Mitogenomes also possess several structural features that, when thoroughly studied, can be phylogenetically informative, such as genome size, gene arrangement and content (Boore and Brown, 1998), as well as the presence and composition of non-coding regions and repetitive sequences, and even RNA secondary structures (Gissi et al., 2008; Kern et al., 2020).

Despite the overall unarguable utility of mitogenomes for phylogenetic assessments, several limitations may affect their reliability for the same purposes. By being an ‘independent genetic unit’, that is usually uniparentally inherited with very little recombination, the mitogenome as a whole is itself a single locus that reflects the evolutionary history of the mitochondria, which for several reasons may not be the same as the species evolutionary history (e.g. due to introgression and sex-biased reproductive dispersal Kern et al., 2020). Furthermore, the presence of non-functional nuclear copies of mitochondrial sequences (numts) may lead to a false interpretation of phylogenetic relationships (Kern et al., 2020), particularly when single genes are amplified by PCR, and the highly variable substitution rates and base composition between taxa can make direct comparisons difficult (Bernt et al., 2013a; Kern et al., 2020). Inversions can also complicate phylogenetic analysis using mtDNA gene sequences, as it is likely that genes equilibrate in nucleotide composition to their strand skew, even to the point of having convergent amino acid substitutions within physico-chemically similar groups that have arisen independently in different lineages (Masta et al., 2009).

Despite these drawbacks, overall mitogenomes represent a complete and ‘isolated’ genomic feature, easily available from a wide range of taxa, whose genetic information is comparable and compact enough to be both phylogenetically informative and investigated with low computational effort and therefore a logical choice for a comprehensive phylogenetic study. Consequently, mitochondrial DNA has been used, with a variable range of success, to assess phylogenetic relationships at several taxonomic levels ranging from shallow population-level relationships (e.g. Froufe et al., 2014), up to phyla (Mollusca: Stöger and Schrödl, 2013; e.g. Annelida: Bleidorn et al., 2006, Platyhelminthes: Park et al., 2007, Rotifera: Min and Park, 2009) and even Metazoa as a whole (Bernt et al., 2013a).

Although mitophylogenetics have been successfully used to infer deeper evolutionary relationships within other metazoan taxa, the same success has not been achieved for the Mollusca. The reconstruction of the molluscan deep-level relationships has been extremely challenging, and consistently recovering the monophyly of the Mollusca or even of the eight molluscan classes, both presumed to be correct based on other data, has not been possible using mitochondrial markers alone (Bernt et al., 2013a; Osca et al., 2014; Schrödl and Stöger, 2014; Stöger and Schrödl, 2013; S. Yokobori et al., 2008). Moreover, only recently and through the application of phylogenomic approaches relying on several nuclear loci, have consistent monophyletic Mollusca and monophyletic molluscan classes started to be recovered (Kocot et al., 2011; Smith et al., 2011; Vinther et al., 2012; Wanninger and Wollesen, 2019). These studies, by contradicting the generally accepted morphocladistic Testaria hypothesis, have resulted in a fundamental reinterpretation of the phylogenetic history of Mollusca. The Testaria hypothesis placed worm-like Aplacophora (Solenogastres and Caudofoveata) as a paraphyletic basal group of the Mollusca and thus postulated a progressive evolution of body complexity, with a true shell occurring only once (Wanninger and Wollesen, 2019). Conversely, all the recent phylogenomic studies unambiguously support a basal dichotomy that splits the Mollusca into two major groups, the Aculifera (including the Polyplacophora and the reciprocally monophyletic Aplacophora) and the Conchifera (including the Monoplacophora, Cephalopoda, Scaphopoda, Gastropoda and Bivalvia), thus postulating that the worm-like body plan of Aplacophora was acquired secondarily and has derived from a more complex-bodied ancestor (Kocot et al., 2020; Smith et al., 2011). However, the relationships within Conchifera are more controversial, with conflicting results regarding the positioning of Monoplacophora as either basal to all other Conchifera (Kocot et al., 2020) or sister taxa to Cephalopoda (Kocot et al., 2020; Smith et al., 2011), as well as the positioning of Scaphopoda as sister to Gastropoda (Kocot et al., 2020; Smith et al., 2011; Vinther et al., 2012) or sister to a clade composed of Gastropoda and Bivalvia (Kocot et al., 2020, 2011; Smith et al., 2011). Nevertheless, phylogenomic studies have been fundamental to understanding early molluscan evolution and although whole genome-scale resources are now easier to obtain, the taxon sampling is still considerably reduced when compared with the mitogenomic data already available (reviewed in Gomes-dos-Santos et al., 2020).

The effectiveness of mtDNA markers to infer deep Molluscan phylogeny has been a thoroughly discussed subject in recent studies (Osca et al., 2014; Schrödl and Stöger, 2014; Stöger and Schrödl, 2013), describing several factors that may lead to the lack of phylogenetic signal and conflicting tree topologies. Phylogenies often show long-branch attraction artefacts (LBA), with molluscan mitogenomes revealing high differentiation in nucleotide abundance and strand bias. All of these features are a probable consequence of highly frequent gene order rearrangements observed in Molluscan mitogenomes, resulting in heterogeneous substitution rates and generating systematic analytical errors (see Bernt et al., 2013a; Schrödl and Stöger, 2014; Stöger and Schrödl, 2013; Uribe et al., 2019 and references within). Furthermore, ancient (Cambrian) incomplete lineage sorting and uneven taxon sampling may also play a role in the inconsistency of the inferred phylogenetic relationships (Schrödl and Stöger, 2014). These authors also explored the phylogenetic utility of other molluscan-specific mitogenome features, such as mitogenome size variation, the highly variable (sometimes absent) protein-coding gene *atp8*, and even the coupling behaviour of particular genes (such as *atp8-atp6* and *nad4L-nad4*) (Schrödl and Stöger, 2014). However, a clear phylogenetic signal is once again hindered, probably by homoplasy of these features.

Within the molluscan classes, deeper relationships based only on mitochondrial markers have also been showing a variable range of success. Recent studies on the Aculifera have expressed promising results using phylomitogenomics, supporting the usefulness of both whole mitogenome sequences and structural features (Irisarri et al., 2020, 2014; Mikkelsen et al., 2018). For instance, new phylogenetic informative mitogenome rearrangements were detected within Polyplacophora and Caudofoveata, which along with the only Solenogastres published mitogenome, revealed a conserved protein-coding gene order likely consistent with the ancestral molluscan gene order (Irisarri et al., 2020, 2014; Mikkelsen et al., 2018). However, mitogenome availability is still scarce for groups within the Aculifera clade. For example, mitogenome sequences for all the main lineages of the best sampled Aplacophora group, Polyplacophora (*n* = 18), only recently became available (Irisarri et al., 2020) (Figure P2.3). Similarly, Scaphopoda, for which several phylogenetic and systematics doubts persist within its major groups, is very poorly represented regarding mitogenome availability (Kocot et al., 2019b). Furthermore, although phylogenetic analysis using complete mitogenomes revealed promising results for the phylogenetic assessment within the Scaphopoda, using *cox1* alone did not and, therefore, a more comprehensive and intensive whole mitogenome sequencing within the group is urgently needed (Kocot et al., 2019b).

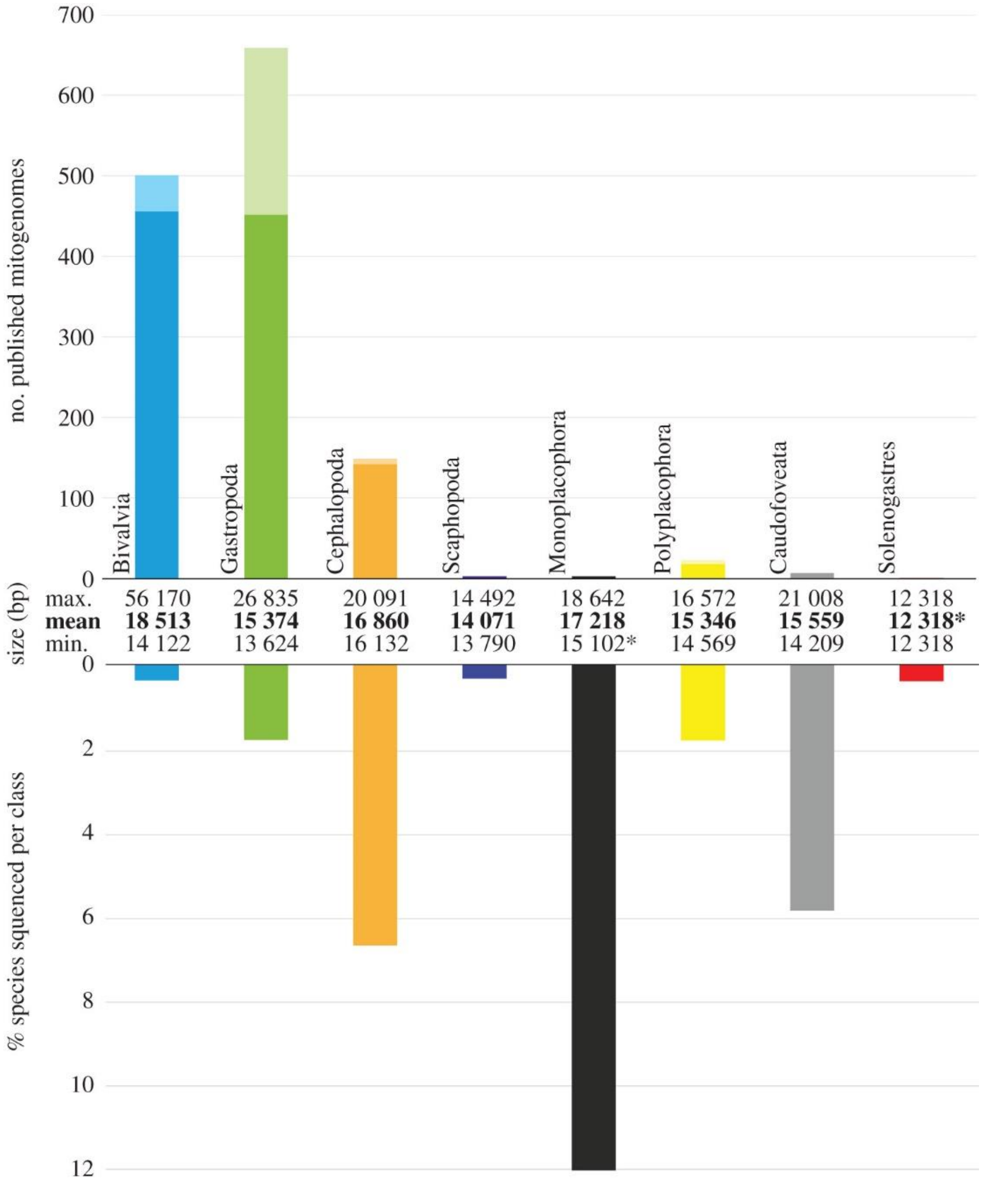


Figure P2. 3 - Top: Graphic showing the number of complete (dark colours) and partial (light colours: min. size 10 000 bp) mitogenomes available in GenBank; middle: mean, minimum and maximum size (bp) of complete mitogenomes per Mollusca class; bottom: graphic showing the percentage of total species with complete mitogenomes published in GenBank. Asterisk superscripts refer to unverified size values, due to assembly challenges; critical evaluation of these publicly available mitogenome sizes and sequence content is highly recommended.

Monoplacophoran mitogenomes have been recently sequenced to test their positioning within the Mollusca. However, consistent with the low resolution of mitochondrial markers for deep molluscan classes (see above) the results were inconclusive (Stöger et al., 2016). Nevertheless, once again unique structural features (e.g. gene arrangement and presence of large intergenic regions) that may be phylogenetically informative were detected and further sampling of the group is needed (Stöger et al., 2016).

Of the three most economically important molluscan classes, Cephalopoda is the best represented in terms of mitogenome availability, which nonetheless represents only 5.5% of the total species of the group. Unlike in other molluscan classes, mitochondrial markers have shown to be informative regarding the deeper Cephalopoda phylogenetic relationships, revealing their potential to resolve long-lasting phylogenetic questions within the group (Stöger and Schrödl, 2013; Uribe and Zardoya, 2017).

As for the two most speciose classes of Mollusca (i.e. Bivalvia and the megadiverse Gastropoda), deep-level phylomitogenomics have been constantly inefficient. Both bivalves and gastropods have very unusual mitochondrial evolutionary patterns at both nucleotide and structural level, which render them prone to analytical inconsistencies (e.g. LBA) and hamper a consistent phylogenetic inference (Combosch et al., 2017; Stöger and Schrödl, 2013; Uribe et al., 2019). Inevitably, only through the application of large-scale genomic approaches are the interrelationships within both classes starting to be clarified (Combosch et al., 2017; Cunha and Giribet, 2019; V. L. Gonzalez et al., 2015; Zapata et al., 2014).

Contrary to these difficulties in the resolution of deeper, older evolutionary relationships, mitochondrial genes and genomes have been much more useful in resolving more recent, intrafamilial phylogenies (Cong et al., 2020; Froufe et al., 2019). Most shallow phylogeny, phylogeographic and populations genetics studies on molluscs have relied so far on one or two mitochondrial gene fragments sometimes coupled with the same number of nuclear counterparts (Fernández-Pérez et al., 2017; Froufe et al., 2016c; Ye et al., 2015). However, use of these gene fragments alone may lead to biased results and fail to reveal the mitochondrial evolutionary history of species. Furthermore, obtaining a complete mitogenome is not always a possibility, either due to the higher cost of sequencing (when compared with Sanger sequencing of a single gene) or due to logistic limitations (e.g. lack of computational resources). It is therefore important to identify the genes or regions of the mitogenome that better correspond and may be used as surrogates of the whole mitogenome evolutionary history. A study on 41 unionid bivalves statistically evaluated the coherence of the individual mitochondrial gene trees and the whole mitogenome tree, indicating that the trees using *nad5* sequences were the most similar to whole mtDNA trees (Fonseca et al., 2016). The results of the gene fragments more widely used in molecular studies within this bivalve taxon (i.e. *cox1*, *rrnL* and *nad1*) were less robust. This study also tested pairs of these widely used gene markers with much higher success, indicating that the trees constructed with the large ribosomal subunit *rrnL* concatenated with *cox1* or *nad1* are highly coherent with the whole mitogenome trees (Fonseca et al., 2016). Another study within the cephalopod Octopodidae family comparing the whole mitogenome with the individual gene tree topologies also showed that the *nad5* trees best represented the whole mitogenome topologies (Abalde et al., 2017; Tang et al., 2018). However, these results were obtained in specific groups of molluscs and should be tested across the Mollusca to evaluate the usefulness of individual and pairs of gene fragments in representing the whole mitochondrial genome phylogenies.

Comparisons of mitochondrial genes have great potential for revealing hidden cryptic diversity aiding in species delimitation and identification (X. Shen et al., 2014; Tang et al., 2018) and in understanding molluscan species phylogeographical patterns and population genetic structure, since they have already been used successfully for these purposes in other taxa (Morin et al., 2010; Teacher et al., 2012). However, to our knowledge, no comprehensive phylogeographic or population genetics study on mollusc species has used this type of marker.

In summary, studies with phylogenetic analyses of whole mitochondrial sequences and structural features of molluscs have been increasing steadily over the last decade. These studies have shown limited success in representing deeper evolutionary patterns within the Mollusca and molluscan classes. However, below the family level, robust phylogenies consistent with results of other genomic and morphological studies have been obtained. Given the high potential of whole mitogenomes for barcoding, revealing cryptic diversity and obtaining robust shallow phylogenetic relationships, it is expected that an increasing number of phylogeographic and population genetics studies using whole mitogenomes will be published shortly.

6. Summary and conclusion

Despite widespread misunderstanding based on early studies that animal mitochondrial genomes are consistent in structure, function and inheritance patterns, there is actually enormous diversity among these diminutive genomes across animal life. The phylum

Mollusca, in particular, is replete with examples of extraordinary variation in genome architecture, molecular functioning and intergenerational transmission. This provides a model system for studying the evolution of these features in concert with the diverse and manifold roles of mitochondria in organismal physiology and the many ways that the study of mitochondrial genomes are useful for phylogeny and population biology.

Acknowledgements

F.G. was funded by the ‘Ricerca Fondamentale Orientata’ (RFO) funding from the University of Bologna, and the Canziani bequest. A.G.S. was funded by FCT—Fundação para a Ciência e a Tecnologia—under grant SFRH/BD/137935/2018. M.L.L. was funded under project ConBiomics: the missing approach for the Conservation of Freshwater Bivalves Project no. NORTE-01-0145-FEDER-030286, co-financed by COMPETE 2020, Portugal 2020 and the European Union through the ERDF, and by FCT through national funds. J.S. was funded by the National Science Foundation (DEB – 1753695, DEB – 1753851).

The authors would like to thank Maurine Neiman and Angus Davison for organizing the Royal Society meeting ‘Pearls of Wisdom: Synergising Leadership and Expertise in Molluscan Genomics' and for guest-editing this issue of *PhilosophicalTransactions B*.

Chapter 3 – Freshwater mussels (Bivalvia: Unionida)

2.1. Paper 3 – The male and female complete mitochondrial genomes of the threatened freshwater pearl mussel *Margaritifera margaritifera* (Linnaeus, 1758) (Bivalvia: Margaritiferidae)

Gomes-dos-Santos, A., Froufe, E., Amaro, R., Ondina, P., Breton, S., Guerra, D., Aldridge, D.C., Bolotov, I.N., Vikhrev, I. v., Gan, H.M., Gonçalves, D. V., Bogan, A.E., Sousa, R., Stewart, D., Teixeira, A., Varandas, S., Zanatta, D., Lopes-Lima, M., 2019. The male and female complete mitochondrial genomes of the threatened freshwater pearl mussel *Margaritifera margaritifera* (Linnaeus, 1758) (Bivalvia: Margaritiferidae). Mitochondrial DNA B Resour 4, 1417–1420. https://doi.org/10.1080/23802359.2019.1598794

The male and female complete mitochondrial genomes of the threatened freshwater pearl mussel *Margaritifera margaritifera* (Linnaeus, 1758) (Bivalvia: Margaritiferidae)

André Gomes-dos-Santos [1,14]*, Elsa Froufe [1], Rafaela Amaro [2], Ondina Paz [2], Sophie Breton [3], Davide Guerra [3], David Aldridge [4], Ivan Bolotov [5], Ilya Vikhrev [5], Han Ming Gan [6], Duarte Gonçalves [1], Arthur Bogan [7], Ronaldo Sousa [8], Donald Stewar [9], Amílcar Teixeira [10], Simone Varandas [11], David Zanatta [12], Manuel Lopes-Lima [1,13]

[1] CIIMAR/CIMAR – Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Terminal de Cruzeiros de Leixões. Av. General Norton De Matos s/n 4450-208 Matosinhos, Portugal Matosinhos, Portugal; [2] Department of Zoology, Genetics, and Physical Anthropology, Faculty of Veterinary Science, University of Santiago de Compostela, Lugo, Spain; [3] Departement de Sciences Biologiques, Universite de Montreal, Canada; [4] Department of Zoology, University of Cambridge, Cambridge, United Kingdom; [5] Federal Center for Integrated Arctic Research Russian Academy of Sciences, Northern Arctic Federal University, Arkhangelsk, Russia; [6] Centre for Integrative Ecology, School of Life and Environmental Sciences, Deakin University, Victoria, Australia; [7] Research Laboratory North Carolina State Museum of Natural Sciences, Raleigh, United States; [8] CBMA – Centre of Molecular and Environmental Biology, Department of Biology, University of Minho, Braga, Portugal; [9] Department of Biology, Acadia University, Wolfville, Canada; [10] CIMO-ESA-IPB – Mountain Research Centre, School of Agriculture, Polytechnic Institute of Braganca, Bragança, Portugal; [11] CITAB-UTAD – Centre for Research and Technology of Agro-Environment and Biological Sciences, University of Trás-os-Montes and Alto Douro, Vila Real, Portugal; [k] Institute of Great Lakes Research, Central Michigan University, United States; [12] Biology Department, Institute for Great Lakes Research, Central Michigan University, Mount Pleasant, MI, USA; [13] CIBIO– Centro de Investigação em Biodiversidade e Recursos Genéticos, InBio Laboratório Associado, Universidade do Porto, Campus Agrário de Vairão, Vairão, Portugal; [14] Department of Biology, Faculty of Sciences, University of Porto, Rua do Campo Alegre1021/1055, Porto, Portugal.

* Corresponding authors

DOI: https://doi.org/10.1080/23802359.2019.1598794

Abstract

The complete mitogenomes of one (M-)ale (North America), one Hermaphroditic (Europe), and two (F-)emale (North America and Europe) individuals of the freshwater pearl mussel *Margaritifera margaritifera* were sequenced. The M-type and F-type (Female and Hermaphroditic) mitogenomes have 17,421 and 16,122 nucleotides, respectively. All with the same content: 13 protein-coding genes, 22 transfer RNA, two ribosomal RNA genes, and one sex-related ORF. The M-type is highly divergent (37.6% uncorrected p-distance) from the F-type mitogenomes. North American and European F-type mitogenomes exhibit low genetic divergence (68 nt substitutions), and the Female and Hermaphroditic European mitogenomes are almost identical, and matching sex-related ORFs.

Keywords

Doubly uniparental inheritance; mitogenome; Unionida

1. Mito Communication

The Margaritiferidae (Bivalvia: Unionida), comprising 16 extant species, represents the most threatened freshwater mussel family (Lopes-Lima et al., 2018a). Within this family, the freshwater pearl mussel *Margaritifera margaritifera* (Linnaeus, 1758) is one of the most threatened species; it is subject to numerous conservation projects and is listed as Endangered globally (Geist, 2010; Moorkens et al., 2018). *Margaritifera margaritifera* is a long lived species (reaching over 100 years) that generally inhabits cool oligotrophic running waters throughout freshwater systems of northwest Europe and northeast North America (Geist, 2010; Lopes-Lima et al., 2017c). As noted for other margaritiferid bivalves, *M. margaritifera* shows an unusual mitochondrial inheritance process called doubly uniparental inheritance (DUI). Under DUI, both males and females inherit F-type mitochondrial DNA from their mothers, while males also inherit M-type mitochondrial DNA from their fathers, which predominates in gonad tissue (Amaro et al., 2016; Hoeh et al., 1996). Furthermore, many hermaphroditic species of freshwater mussels seem to have lost DUI and do not possess the M-type mitochondrial genome in their gonad tissues (Breton et al., 2011b).

The M-type and F-type mitochondrial lineages show high levels of divergence within species of Unionida freshwater mussels and even distinct gene order arrangements

(Fonseca et al., 2016; Froufe et al., 2016a; Guerra et al., 2017). The Margaritiferidae also exhibit a unique M-type and F-type gene order (Lopes-Lima et al., 2017a). Available phylogenetic studies within the family are based on only a few markers still lacking a more robust multi-marker approach (Lopes-Lima et al., 2018a).

Given this background, the aims of this study are to (1) obtain the whole mitogenomes of male, female, and hermaphroditic specimens of *M. margaritifera* from North America and Europe; (2) determine and compare the gene order and content of those mitogenomes; and (3) produce phylogenetic analyses using all available F-types and M-type mitogenomes of the Margaritiferidae family.

Four complete mitogenomes of *M. margaritifera* were sequenced: one M-type and one F-type from a North American male specimen (River Annapolis near Auburn, Canada: 45.014999, -64.856344) and two F-type from European specimens, one from a female (River Ulla near Barazon, Galicia, Spain: approximate coordinates 42.846676, -8.025244) and another from a hermaphrodite (River Tuela near Vinhais, northeast Portugal: approximate coordinates 41.862414, -6.931596). DNA samples are stored at the CIIMAR Institute Unionoid DNA and Tissue Databank (Voucher numbers P2, MM63, 155G, and 165G). Sex was determined for all specimens under a microscope following Hinzmann et al., (2013).

DNA was sheared to ~500 bp using an M220 Covaris Ultrasonicator (Covaris, Woburn, MA, USA) and processed with the NEBultra Illumina library preparation kit (NEB, Ipswich, MA, USA). Sequencing was performed on the MiSeq (Illumina, San Diego, CA, USA) located at Monash University Malaysia using a run configuration of 2 × 250 bp. Mitogenomes were assembled from the paired-end reads and annotated using an established pipeline (Gan et al., 2014). The four mitogenomes have been deposited in the GenBank database under the accession numbers (MK421959 and MK421956; M-type and F-type, respectively for the North American specimens), and (MK421957 and MK421958; for the Spanish and Portuguese F-type European specimens). Sequence divergence (uncorrected p-distance) was assessed using MEGA7 software (Kumar et al., 2016). The length of both mitogenome types (M-type: 17,421 nt; and F-type: 16,122 nt) of *M. margaritifera* sequenced in this study is within the expected range for each gender-specific haplotypes within Margaritiferidae (Guerra et al., 2017; Lopes-Lima et al., 2017a). The larger size of the M-type genomes is expected given the larger *cox2* gene and the presence of M-specific coding regions (Breton et al., 2009). Both haplotypes have the same gene content: 13 protein-coding genes (PCGs), 22 transfer RNA (trn) and two ribosomal RNA (rrn) genes. ORFs specific to each type of mtDNA, F-orf in the F mitogenome and M-orfs in the M, are also present. Regarding the gene orientation, again, both have the same genes (four PCGs, 20 tRNAs, and two rRNAs) encoded on the heavy strand and the remaining (nine PCGs and two tRNAs) encoded on the complementary strand. The exception is the sex related ORFs, with the M-orf on the complementary strand and the F-orf on the heavy strand, located at different positions. A nucleotide alignment of the mitochondrial genomes shows that the M-type mitogenome is highly divergent (37.6% uncorrected p-distance) from the F-type mitogenomes. The F-type mitogenomes from North America and Europe exhibit a low genetic divergence (68 nt substitutions = 0.04% uncorrected p-distance), with the European mitogenomes of the female and hermaphroditic individuals being almost identical with only 5 nt substitutions. This pattern may reflect a recent (Pleistocene) dispersal event of freshwater pearl mussels from Europe to North America or slow mtDNA substitution rates in this species (Lopes-Lima et al., 2018a; Zanatta et al., 2018). The F-orfs of the European hermaphroditic and female individuals are identical. Secondarily hermaphroditic species generally contain a distinct and longer F-like ORF (Breton et al., 2011b). Therefore, these results seem to indicate that hermaphroditic individuals of typically dioecious species may maintain their F-type ORFs unchanged.

For the phylogenies, additional mitogenome sequences (M-type and F-type) available from all Margaritiferidae and three Unionidae species were downloaded from GenBank. Each gene sequence was aligned using GUIDANCE v. 1.5 (Penn et al., 2010) with the MAFFT v. 7.304 multiple sequence alignment algorithm (Katoh and Standley, 2013). To build single gene alignments the following GUIDANCE parameters were used: score algorithm: GUIDANCE; bootstraps replicates: 100; Sequence cut-off score: 0.0 (no sequences removed); Column cut-off score: below 0.8. The final concatenated data set included the 13 mitochondrial PCG and the 2 rrn genes for each mitogenome reaching a total length of 13,505 nt. Phylogenetic relationships were estimated by Bayesian inference using MrBayes v. 3.2.6 (Ronquist et al., 2012) and Maximum Likelihood using RAxML v. 8.2.10 (Stamatakis, 2014) HPC Black Box with 100 rapid bootstrap replicates and 20 ML searches at the San Diego Supercomputer Center through the CIPRES Science Gateway (https://www.phylo.org). The final alignment was partitioned in 11 subsets according to the best scheme determined using PartitionFinder v.2.1.1 (Lanfear et al., 2017). For the ML a unique GTR model was applied for each partition with corrections for gamma distribution. For the BI, the GTR + G, GTR + I+G, HKY + G, HKY + I+G models were used. Each chain started with a randomly generated tree and ran for 1 × 10^6 generations with a sampling frequency of 1 tree for every 100 generations. The resultant trees, after discarding the first 25% as burn-in, were combined in a 50% majority rule consensus tree. The final trees were rooted at the split between Male and Female haplotypes (based on previous studies, e.g. X. C. Huang et al., 2013).

The best obtained phylogenetic BI and ML trees revealed an identical topology (Figure P3.1). Both the F and M clades are divided into the two Unionida families, Margaritiferidae, and Unionidae. Maximum support values were obtained for all nodes with two exceptions for the relationships of *Pseudunio marocanus* both in the female and in the male clades (Figure P3.1). The phylogenies are consistent with the systematic divisions of the Margaritiferidae in four genera (*Margaritifera*, *Cumberlandia*, *Pseudunio*, and *Gibbosula*) and two subfamilies (Margaritiferinae (*Margaritifera* + *Cumberlandia* + *Pseudunio*) and Gibbosulinae (*Gibbosula*)) (Lopes-Lima et al., 2018a). The newly sequenced *M. margaritifera* genomes cluster inside the *Margaritifera* genus in the F-type clade, being the M-type mitogenome sequence the first available for this genus, following the most recent systematics for the family (Lopes-Lima et al., 2018a).

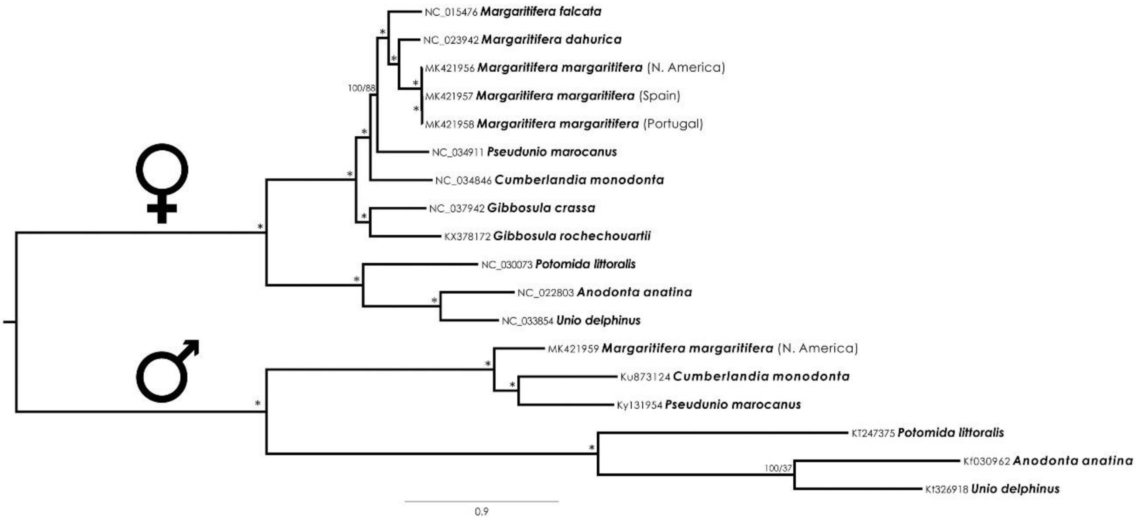

NC_015476 Margaritifera falcata
NC_023942 Margaritifera dahurica
MK421956 Margaritifera margaritifera (N. America)
MK421957 Margaritifera margaritifera (Spain)
MK421958 Margaritifera margaritifera (Portugal)
NC_034911 Pseudunio marocanus
NC_034846 Cumberlandia monodonta
NC_037942 Gibbosula crassa
KX378172 Gibbosula rochechouartii
NC_030073 Potomida littoralis
NC_022803 Anodonta anatina
NC_033854 Unio delphinus
MK421959 Margaritifera margaritifera (N. America)
Ku873124 Cumberlandia monodonta
Ky131954 Pseudunio marocanus
KT247375 Potomida littoralis
Kf030962 Anodonta anatina
Kt326918 Unio delphinus
100/88
100/37
*
0.9

Figure P3. 1 - Margaritiferidae and Unionidae Bayesian phylogenetic tree of Male and Female mitogenomes sequences based on concatenated nucleotide sequences of 13 mitochondrial protein-coding genes and the two rRNA genes. GenBank accession numbers are behind species names, numbers at the nodes indicate the percentage posterior probabilities and bootstrap support values. The * above the branches indicate posterior probabilities and bootstrap support values > 95%.

The present results confirm the usefulness of the mitogenomes gene arrangements as diagnostic character for the Margaritiferidae and provide additional confirmation for the systematics of the family as recently proposed by Lopes-Lima et al., (2018a). These results also highlight the low intraspecific genetic divergence of *M. margaritifera* even between specimens from the edges of distribution. Furthermore, the current study provides novel information about mtDNA structure and sequence of hermaphroditic individuals of typical dioecious species providing opportunities for further studies on the sex determination mechanism and mtDNA evolution of freshwater bivalves.

Acknowledgements

This work was supported by Portuguese Foundation for Science and Technology (FCT) [grant number SFRH/BD/115728/2016 (MLL), grant number SFRH/BD/137935/2018 (AGS)]; Russian Foundation for Basic Research [grant number 18-34-20033 (IVV)]; Dawson Fellowship at St. Catharine’s College, Cambridge (DCA); Life Margal Ulla [number LIFE09 NAT/ES/000514 (RA and PO)]; COMPETE 2020, Portugal 2020 and the European Union through the ERDF, and by Portuguese Foundation for Science and Technology (FCT) through national funds [UID/Multi/04423/2019] under project ConBiomics: the missing approach for the Conservation of freshwater Bivalves Project, and [project number NORTE-01-0145-FEDER-030286]; Federal Agency for Scientific Organizations under Grants [grant number 0409-2015-0143 (INB and IVV)]

2.2. Paper 4 – A novel assembly pipeline and functional annotations for targeted sequencing: A case study on the globally threatened Margaritiferidae (Bivalvia: Unionida)

Gomes-dos-Santos, A., Froufe, E., Pfeiffer, J., Smith, C., Machado, A., Castro, L.F.C., Do, V., Hattori, A., Garrison, N., Whelan, N., others, 2022. A novel assembly pipeline and functional annotations for targeted sequencing: A case study on the globally threatened Margaritiferidae (Bivalvia: Unionida). Authorea Preprints. DOI: 10.1111/1755-0998.13802

A novel assembly pipeline and functional annotations for targeted sequencing: A case study on the globally threatened Margaritiferidae (Bivalvia: Unionida)

André Gomes-dos-Santos[1,2]*, Elsa Froufe[1]*, John Pfeiffer[3], Nathan Johnson[4], Chase Smith[5], André M. Machado[1,2], L. Filipe C. Castro[1,2], Van Tu Do[6,7], Akimasa Hattori[8], Nicole Garrison[9], Nathan Whelan[9,10], Ivan N. Bolotov[11], Ilya V. Vikhrev[11], Alexander V. Kondakov[11], Mohamed Ghamizi[12], Vincent Prié[13,14], Arthur E. Bogan[15], Manuel Lopes Lima[14,16]*

[1]CIIMAR/CIMAR - Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Matosinhos, Portugal; [2]Department of Biology, Faculty of Sciences, University of Porto, Porto, Portugal; [3]National Museum of Natural History, Smithsonian Institution, Washington, DC, USA; [4]U.S. Geological Survey, Wetland and Aquatic Research Center, Gainesville, FL, USA; [5]Department of Integrative Biology, University of Texas, Austin, Texas, USA; [6]Institute of Ecology and Biological Resources, Vietnam Academy of Science and Technology, 18 Hoang Quoc Viet, Cau Giay, Ha Noi, Viet Nam; [7]Graduate University of Science and Technology, Vietnam Academy of Science and Technology, 18 Hoang Quoc Viet, Cau Giay, Ha Noi, Viet Nam; [8]Matsuyama High School, 1-6-10 Matsuyama-cho, Higashimatsuyama, Saitama 355-0018, Japan; [9]School of Fisheries, Aquaculture, and Aquatic Sciences, College of Agriculture, Auburn University, Alabama, USA; [10]Southeast Conservation Genetics Lab, Warm Springs Fish Technology Center, US Fish and Wildlife Service, Auburn, Alabama, USA; [11]N. Laverov Federal Center for Integrated Arctic Research of the Ural Branch of the Russian Academy of Sciences, Northern Dvina Emb. 23, 163069, Arkhangelsk, Russia; [12]Department of Biology, Natural History Museum of Marrakech, University of Cadi Ayyad, Marrakech, Morocco; [13]Research Associate - Institute of Systematics, Evolution, Biodiversity (ISYEB), National Museum of Natural History (MNHN), CNRS, SU, EPHE, UA CP 51, 57 rue Cuvier, 75005 Paris, France; [14]CIBIO/InBIO - Research Center in Biodiversity and Genetic Resources, Universidade do Porto, Vairão, Portugal; [15]North Carolina Museum of Natural Sciences, 11 West Jones St., Raleigh, NC. USA; [16]IUCN SSC Mollusc Specialist Group, c/o IUCN, David Attenborough Building, Pembroke St., Cambridge, England

* Corresponding authors

DOI: 10.1111/1755-0998.13802

Abstract

The proliferation of genomic sequencing approaches has significantly impacted the field of phylogenetics. Target capture approaches provide a cost-effective, fast, and easily applied strategy for phylogenetic inference of non-model organisms. However, many existing pipelines used to create phylogenomic datasets from target capture data are incapable of incorporating whole genome sequencing data into their workflows. Here, we develop a highly efficient pipeline for capturing and *de novo* assembly of the targeted regions using whole genome re-sequencing reads. This new pipeline allows capturing targeted loci accurately and efficiently, and given its unbiased nature, can easily be expanded to be used with any other target capture probe set. We demonstrate the utility of our approach by incorporating whole genome sequencing data into a recently developed target capture probe set to reconstruct the evolutionary history of the freshwater mussel family Margaritiferidae, reconstructing supraspecific relationships outside the Unionidae family, providing the first comprehensive multi-loci phylogeny of the Margaritiferidae. We also provide a catalogue of well-curated functional annotations of the targeted regions for the target capture probe set, representing a complementary tool for scrutinizing phylogenetic inferences while expanding future applications of the probe set.

Keywords

Target capturing assembly; Anchored Hybrid Enrichment; Phylogenomics; Freshwater Mussels.

1. Introduction

Two decades separate the announcements of the first human genome sequence (Craig Venter et al., 2001; Lander et al., 2001) and the recently assembled gapless human genome (Nurk et al., 2022). In between, genome biology has been revolutionized by a fundamental shift in the process and the scale on which DNA and RNA are studied (Goodwin et al., 2016; L. Koch et al., 2021; Sedlazeck et al., 2018; Stephan et al., 2022). The initial prohibitive cost of whole genome-scale sequencing dropped numerous orders of magnitude at an astonishing pace, which has revolutionized biodiversity research (Goodwin et al., 2016; Sedlazeck et al., 2018; Stephan et al., 2022). In particular, new genomic approaches have transformed the field of phylogenetics and helped to usher in the new era of “phylogenomics”. The emerging phylogenomic approaches offer several cost-effective strategies to simultaneously sequence hundreds or thousands of loci, which was one of the major constraints of Sanger sequencing (Kapli et al., 2020; Lemmon and Lemmon, 2013; Smith and Hahn, 2021).

Although whole-genome sequencing (WGS) presents an ideal option for phylogenomic studies of non-model organisms, it is still a rarely used resource (Stephan et al., 2022). However, the costs of WGS have constantly decreased in the last two decades, which will expand its application when coupled with increasing reference genome assemblies of non-model organisms (Goodwin et al., 2016; Sedlazeck et al., 2018; Stephan et al., 2022). Nevertheless, when compared with genome subsampling strategies (e.g., genotype-by-sequencing, target capture), WGS data processing is more computationally demanding, especially for large and complex genomes (Goodwin et al., 2016; Sedlazeck et al., 2018; Stephan et al., 2022). Consequently, phylogenomic studies have tendentially favoured genotype-by-sequencing, whole mitogenomes, target capture, or transcriptomes (e.g., Alda et al., 2021; Breinholt et al., 2018; Fernández et al., 2018; V. L. Gonzalez et al., 2015; D. D. Houston et al., 2022; Hughes et al., 2018; Ilves et al., 2018; Kocot et al., 2019a; Leebens-Mack et al., 2019; Lemmon and Lemmon, 2013; Pfeiffer et al., 2019; Schwentner et al., 2017; Vieira et al., 2022). Target capture methods have become increasingly popular due to the ability to simultaneously sequence hundreds or thousands of evolutionarily conserved loci at multiple phylogenetic scales (Faircloth et al., 2012; Lemmon et al., 2012; Lemmon and Lemmon, 2013). One target capture method that is suitable at both the shallow and deep phylogenetic scales is anchored hybrid enrichment (AHE) (Lemmon et al., 2012; Lemmon and Lemmon, 2013), which targets highly conserved regions and their rapidly evolving flanks (Lemmon et al., 2012; Lemmon and Lemmon, 2013). Anchored hybrid enrichment probe sets have been designed for a variety of taxa, including vertebrates (Lemmon et al., 2012), butterflies and moths (Breinholt et al., 2018), flies (Young et al., 2016), spiders (Maddison et al., 2017), flowering plants (Buddenhagen et al., 2016), freshwater gastropods (Whelan et al., 2022), and freshwater mussels (Pfeiffer et al., 2019). With the increasing availability of WGS, new methods that incorporate them into target capture datasets are still limited (but see J. M. Allen et al., 2017; Faircloth, 2016; Knyshov et al., 2021) and there is a need to develop efficient pipelines that can incorporate WGS data into their workflow.

Freshwater mussels (Bivalvia: Unionida) represent an ecologically and taxonomically diverse group of bivalves composed of six families and nearly 1,000 species (Graf and Cummings, 2021). These organisms play fundamental roles in freshwater ecosystems, such as water filtration, nutrient cycling and sediment bioturbation and oxygenation (Graf and Cummings, 2021; Lopes-Lima et al., 2017c; Vaughn et al, 2015). Unfortunately, the group is also among the most threatened worldwide, showing the second-highest percentage of threatened species (43%) and among the highest percentage of assessed wild extinctions (5.9%) (Díaz et al., 2019; Lopes-Lima et al., 2021). Freshwater mussels have a fascinating evolutionary history that includes obligatory parasitism of freshwater vertebrates (primarily fishes), doubly uniparental inheritance (DUI) of mitochondrial DNA (mtDNA), parental care, and a wide range of habitats preferences, resulting in complex evolutionary histories (Graf and Cummings, 2021; Lopes-Lima et al., 2017c). Despite their ecological importance and conservation concern, the evolutionary history of many lineages remains uncertain. Phylogenetic studies of this group have primarily relied on either reduced character sampling, generally using a few mitochondrial and nuclear markers or, more recently, the whole mitogenome (e.g., Araujo et al., 2018; Combosch et al., 2017; Froufe et al., 2019; X. C. Huang et al., 2019; Lopes-Lima et al., 2017a, 2018a; Whelan et al., 2011; R. W. Wu et al., 2019). However, many of these studies have highlighted that resolving evolutionary relationships within Unionida based on these datasets is challenging, either based on limited signal (Lopes-Lima et al., 2017b) or mitonuclear discordance (Chong and Roe, 2018; Sano et al., 2022), highlighting the need for phylogenomic strategies to reconstruct the evolutionary history of the group. Although their genomic and transcriptomic resources are limited (e.g., Capt et al., 2018a; Chen et al., 2019; Gomes-dos-Santos et al., 2021; D. Huang et al., 2019; Renaut et al., 2018; Rogers et al., 2021; Smith, 2021; Q. Yang et al., 2021), the recent development of an AHE probe set (Unioverse) has provided a useful large-scale phylogenomic tool for the group (Pfeiffer et al., 2019). This probe set has accelerated our ability to understand various aspects of freshwater mussel evolution (Pfeiffer et al., 2019, 2021; Smith et al., 2020) but has not yet been meaningfully applied to the families outside the Unionidae, which remain dramatically undersampled (i.e., Hyriidae – 1/62 species; Mulleriidae – 1/53 species; Irididinidae – 1/39; Margaritiferidae – 2/15 species; Etheriidae – 1/4 species).

The Unioverse probes were designed from a set of exonic regions from eight unionid transcriptomes and the reference genome of one marine bivalve, *Bathymodiolus* *platifrons* Hashimoto and Okutani 1994 (Bivalvia, Mytillidae) (Pfeiffer et al. 2019). Since these AHE probe sets are regarded as tools for phylogenetic inferences, their functional relevancy is often ignored (Lemmon et al., 2012; Lemmon and Lemmon, 2013; Pfeiffer et al., 2019). Accurate and unbiased phylogenetic reconstruction is a multi-dependent task which requires a set of decisions regarding evolutionary modelling, orthology assessment, and matrix reconstruction to properly balance information from loci with dissimilar underlying evolutionary histories (Bernt et al., 2013a; Buddenhagen et al., 2016; Chen et al., 2007; Edwards et al., 2016; Hosner et al., 2016; Lemmon and Lemmon, 2013; Zhang et al., 2018). Those decisions, as well as the assessment of the results, will therefore be facilitated by proper structural and functional characterization of loci, as annotations will provide a framework to guide data processing, test robustness, and identify biases. Moreover, annotations will widen the applications for the data, such as targeted functional analyses, using the probes as a reference.

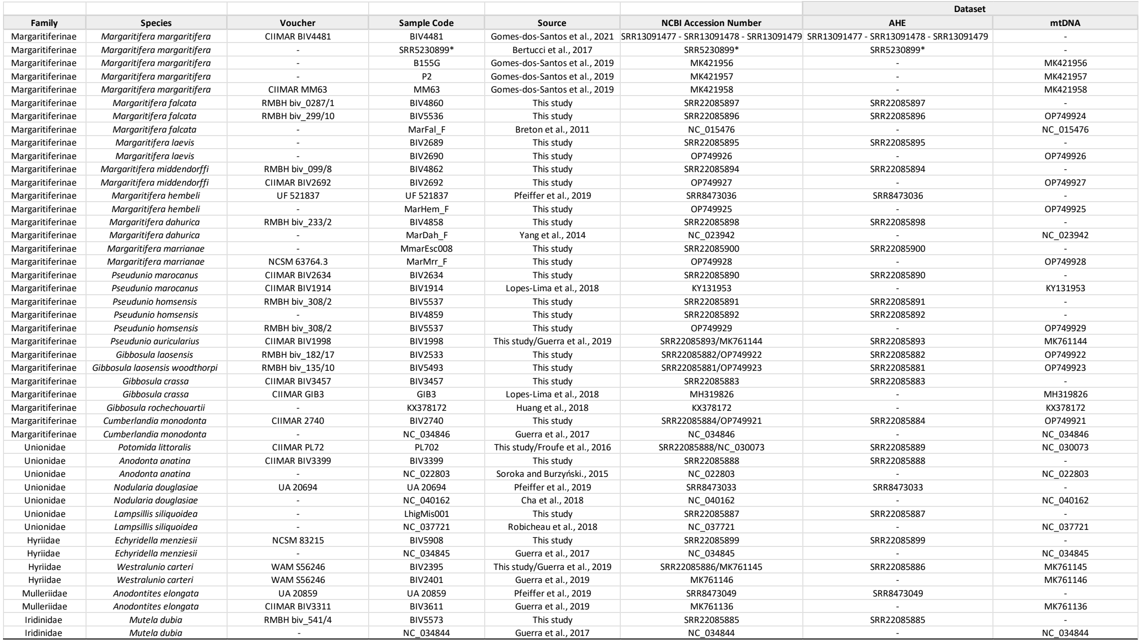
Here, we develop an assembly pipeline for target capture data that can incorporates WGS data. The pipeline is designed to be applied to all types of target capture sequencing data and can easily be used with any probe set. We demonstrated the utility of this approach by using the pipeline to reconstruct the evolutionary history of the family Margaritiferidae. We incorporated the recently published WGS dataset of the freshwater pearl mussel *Margaritifera margaritifera* (Linnaeus, 1758) (Gomes-dos-Santos et al., 2021) and increased the available AHE datasets for Margaritiferidae, from 2 to 11 species, including representatives of all four margartiferid genera. Within the tested dataset, the new pipeline allowed the best balance between loci capturing, sequence length and duplication rate. To compare and complement this primarily nuclear data set, nine additional whole mitogenomes were produced for direct comparison. We also produce a catalogue of functional annotations for the Unioverse targets based on a highly comprehensive multi-evidence database search (i.e., INTPRO, pfam, GO, KEGG and BUSCO), representing a complementary tool for scrutinizing phylogenetic inferences while expanding future application of the probe set.

2. Material and Methods

2.1. Sampling and DNA extraction

Tissue samples from 22 margaritiferid specimens and eight outgroup taxa (from families Unionidae, Hyriidae, Iridinidae and Mycetopodidae) were collected (Table P4.1). A small piece of foot tissue was subsampled from the animals following Naimo, Damschen, Rada, and Monroe, (1998) and placed in 96% ethanol and the specimen returned to its habitat. Genomic DNA was extracted for all samples using a conventional high-salt protocol (Sambrook et al., 1989) or the Qiagen DNeasy Blood and Tissue Kit.

Table P4. 1 - List of samples used in each dataset and respective NCBI accession codes. * Only used in Figure P4.S1.



						Dataset	
Family	**Species**	**Voucher**	**Sample Code**	**Source**	**NCBI Accession Number**	**AHE**	**mtDNA**
Margaritiferinae	*Margaritifera margaritifera*	CIIMAR BIV4481	BIV4481	Gomes-dos-Santos et al., 2021	SRR13091477 - SRR13091478 - SRR13091479	SRR13091477 - SRR13091478 - SRR13091479	-
Margaritiferinae	*Margaritifera margaritifera*	-	SRR5230899*	Bertucci et al., 2017	SRR5230899*	SRR5230899*	-
Margaritiferinae	*Margaritifera margaritifera*	-	B155G	Gomes-dos-Santos et al., 2019	MK421956	-	MK421956
Margaritiferinae	*Margaritifera margaritifera*	-	P2	Gomes-dos-Santos et al., 2019	MK421957	-	MK421957
Margaritiferinae	*Margaritifera margaritifera*	CIIMAR MM63	MM63	Gomes-dos-Santos et al., 2019	MK421958	-	MK421958
Margaritiferinae	*Margaritifera falcata*	RMBH biv_0287/1	BIV4860	This study	SRR22085897	SRR22085897	-
Margaritiferinae	*Margaritifera falcata*	RMBH biv_299/10	BIV5536	This study	SRR22085896	SRR22085896	OP749924
Margaritiferinae	*Margaritifera falcata*	-	MarFal_F	Breton et al., 2011	NC_015476	-	NC_015476
Margaritiferinae	*Margaritifera laevis*	-	BIV2689	This study	SRR22085895	SRR22085895	-
Margaritiferinae	*Margaritifera laevis*	-	BIV2690	This study	OP749926	-	OP749926
Margaritiferinae	*Margaritifera middendorffi*	RMBH biv_099/8	BIV4862	This study	SRR22085894	SRR22085894	-
Margaritiferinae	*Margaritifera middendorffi*	CIIMAR BIV2692	BIV2692	This study	OP749927	-	OP749927
Margaritiferinae	*Margaritifera hembeli*	UF 521837	UF 521837	Pfeiffer et al., 2019	SRR8473036	SRR8473036	-
Margaritiferinae	*Margaritifera hembeli*	-	MarHem_F	This study	OP749925	-	OP749925
Margaritiferinae	*Margaritifera dahurica*	RMBH biv_233/2	BIV4858	This study	SRR22085898	SRR22085898	-
Margaritiferinae	*Margaritifera dahurica*	-	MarDah_F	Yang et al., 2014	NC_023942	-	NC_023942
Margaritiferinae	*Margaritifera marrianae*	-	MmarEsc008	This study	SRR22085900	SRR22085900	-
Margaritiferinae	*Margaritifera marrianae*	NCSM 63764.3	MarMrr_F	This study	OP749928	-	OP749928
Margaritiferinae	*Pseudunio marocanus*	CIIMAR BIV2634	BIV2634	This study	SRR22085890	SRR22085890	-
Margaritiferinae	*Pseudunio marocanus*	CIIMAR BIV1914	BIV1914	Lopes-Lima et al., 2018	KY131953	-	KY131953
Margaritiferinae	*Pseudunio homsensis*	RMBH biv_308/2	BIV5537	This study	SRR22085891	SRR22085891	-
Margaritiferinae	*Pseudunio homsensis*	-	BIV4859	This study	SRR22085892	SRR22085892	-
Margaritiferinae	*Pseudunio homsensis*	RMBH biv_308/2	BIV5537	This study	OP749929	-	OP749929
Margaritiferinae	*Pseudunio auricularius*	CIIMAR BIV1998	BIV1998	This study/Guerra et al., 2019	SRR22085893/MK761144	SRR22085893	MK761144
Margaritiferinae	*Gibbosula laosensis*	RMBH biv_182/17	BIV2533	This study	SRR22085882/OP749922	SRR22085882	OP749922
Margaritiferinae	*Gibbosula laosensis woodthorpi*	RMBH biv_135/10	BIV5493	This study	SRR22085881/OP749923	SRR22085881	OP749923
Margaritiferinae	*Gibbosula crassa*	CIIMAR BIV3457	BIV3457	This study	SRR22085883	SRR22085883	-
Margaritiferinae	*Gibbosula crassa*	CIIMAR GIB3	GIB3	Lopes-Lima et al., 2018	MH319826	-	MH319826
Margaritiferinae	*Gibbosula rochechouartii*	-	KX378172	Huang et al., 2018	KX378172	-	KX378172
Margaritiferinae	*Cumberlandia monodonta*	CIIMAR 2740	BIV2740	This study	SRR22085884/OP749921	SRR22085884	OP749921
Margaritiferinae	*Cumberlandia monodonta*	-	NC_034846	Guerra et al., 2017	NC_034846	-	NC_034846
Unionidae	*Potomida littoralis*	CIIMAR PL72	PL702	This study/Froufe et al., 2016	SRR22085888/NC_030073	SRR22085889	NC_030073
Unionidae	*Anodonta anatina*	CIIMAR BIV3399	BIV3399	This study	SRR22085888	SRR22085888	-
Unionidae	*Anodonta anatina*	-	NC_022803	Soroka and Burzyński., 2015	NC_022803	-	NC_022803
Unionidae	*Nodularia douglasiae*	UA 20694	UA 20694	Pfeiffer et al., 2019	SRR8473033	SRR8473033	-
Unionidae	*Nodularia douglasiae*	-	NC_040162	Cha et al., 2018	NC_040162	-	NC_040162
Unionidae	*Lampsillis siliquoidea*	-	LhigMis001	This study	SRR22085887	SRR22085887	-
Unionidae	*Lampsillis siliquoidea*	-	NC_037721	Robicheau et al., 2018	NC_037721	-	NC_037721
Hyriidae	*Echyridella menziesii*	NCSM 83215	BIV5908	This study	SRR22085899	SRR22085899	-
Hyriidae	*Echyridella menziesii*	-	NC_034845	Guerra et al., 2017	NC_034845	-	NC_034845
Hyriidae	*Westralunio carteri*	WAM S56246	BIV2395	This study/Guerra et al., 2019	SRR22085886/MK761145	SRR22085886	MK761145
Hyriidae	*Westralunio carteri*	WAM S56246	BIV2401	Guerra et al., 2019	MK761146	-	MK761146
Mulleriidae	*Anodontites elongata*	UA 20859	UA 20859	Pfeiffer et al., 2019	SRR8473049	SRR8473049	-
Mulleriidae	*Anodontites elongata*	CIIMAR BIV3311	BIV3611	Guerra et al., 2019	MK761136	-	MK761136
Iridinidae	*Mutela dubia*	RMBH biv_541/4	BIV5573	This study	SRR22085885	SRR22085885	-
Iridinidae	*Mutela dubia*	-	NC_034844	Guerra et al., 2017	NC_034844	-	NC_034844

2.2. Sample selection and sequencing

A total of 14 DNA extractions, from all known species of the genera *Margaritifera, Pseudunio*, *Cumberlandia,* and two species of the genus *Gibbosula* were selected for the AHE sequencing, as well as six outgroup taxa (Table P4.1). Genomic DNA was used for capturing phylogenetically informative nuclear protein-coding loci using the Unioverse probe set developed by Pfeiffer et al., (2019). Capture, library preparation, and Illumina sequencing were performed at RAPiD Genomics (Gainesville, FL, USA) (Table P4.1). For all species aside from *M. hembeli* and *M. marrianae*, genomic DNA was sheared to

~400 bp fragments, end-repaired and adenine residues were ligated to the 3-end of each blunt-end. Subsequently, barcoded adapters were ligated to the library and amplified by PCR. SureSelectxt Target Enrichment System for Illumina paired-end Multiplexed Sequencing Library protocol was used for solution-based target enrichment of pooled libraries. Probes were synthesized as Custom SureSelect probes from AgilentTechnologies (Santa Clara, CA, USA). Finally, 150-bp long paired-end reads were generated in an Illumina HiSeq 3000 (San Diego, CA, USA).

A total of 11 DNA extractions, encompassing all genera, including 5 species from the genus *Margaritifera*, 1 species from *Pseudunio*, one *Cumberlandia* and two from *Gibbosula* were selected for mitogenome sequencing (Table P4.1). Briefly, genomic DNA was sheared to ~500 bp using an M220 Covaris Ultrasonicator (Covaris, Woburn, MA, USA). Illumina library preparation was constructed using NEBultra Illumina library preparation kit (NEB, Ipswich, MA, USA). Sequencing was performed at Monash University Malaysia using a MiSeq (Illumina, San Diego, CA, USA) Illumina machine, producing 2 × 250 bp paired-end reads.

For *M. hembeli* and *M. marrianae*, Illumina library preparation was done with an Illumina Nextera XT library preparation kit. Libraries were dual-barcoded and sequencing was performed at the U.S. Fish and Wildlife Service Northeast Fisheries Center using an Illumina MiSeq and 2 x 150 bp paired-end sequencing; because of an issue with sequencing chemistry, the second read resulted in only 57 bp reads, but this did not affect genome assembly (see below).

2.3. Targeted sequence assembly

2.3.1. AHE and RNA-seq sequencing outputs

The novel and previously generated AHE sequenced samples (Pfeiffer et al 2019) and the RNA-seq samples (Bertucci et al., 2017) (Table P4.1) were processed according to the pipeline developed by (Breinholt et al., 2018). Briefly, reads with less than 30 nt were filtered, and reads with Phred score < 20 were trimmed using Trim Galore! v0.4.4 (www.bioinformatics.babraham.ac.uk/ projects/trim_galore/). Locus assemblies were produced with the iterative bait assembly script IBA.py (Breinholt et al., 2018) using Unioverse reference probes sequences as baits (Pfeiffer et al., 2019). Briefly, IBA filters reads with high similarity to the reference taxa probe using USEARCH v 10.0.240 (Edgar, 2010), producing a *de novo* assembly of isoforms from the selected reads using the transcriptome assembler Bridger v2014-12-01 (Chang et al., 2015). Afterward, MAFFT v 7.453 (Katoh and Standley, 2013) was used to add the *de novo* assemblies to the Unioverse reference alignment using the parameters "-addlong" and "adjustdirectionaccurately". Exonic and flanking regions were split using the script extract_probe_region.py (Breinholt et al., 2018). Gene orthology was checked by single hit to the same regions of the *B. platifrons* genome using the ortholog_filter.py (Breinholt et al., 2018). Individual alignments for each locus were retrieved with script split.py (Breinholt et al., 2018) and subsequently aligned with MAFFT. At each locus, a single consensus sequence (of each isoform) was produced using FASconCAT-G v 1.04 (Kück and Longo, 2014). Finally, the script remove_duplicates.py (Breinholt et al., 2018) was used to discard loci with more than one sequence per taxon.

2.3.2. Whole genome sequencing outputs

Since the IBA assembly has only been developed for assembling loci derived either from AHE sequencing outputs (IBA.py) or RNA-seq sequencing outputs (IBA_trans.py) (Breinholt et al., 2018), here we develop a new pipeline for assembling loci using whole-genome sequencing outputs (Illumina short read). This pipeline is based on an alternative iterative bait assembly, i.e., SRAssembler (McCarthy et al., 2019), that relies on two distinct whole-genome assemblers for the assembly stage, i.e., ABySS (Jackman et al., 2017) and SOAPdenovo2 (Luo et al., 2012). The freshwater pearl mussel *M. margaritifera* whole genome sequencing reads were retrieved from NCBI (SRR13091477, SRR13091478, SRR13091479). Although SRAssembler allows to directly use raw sequencing outputs, it rapidly escalates the storage requirements, especially if a large amount of initial sequencing reads is provided. Therefore, a prefiltering was performed before running SRAssembler. Using the AHE assembly datasets (Margaritiferidae and outgroup) generated as described above, a set of reference sequences composed of both probe and flanking regions was produced. This set of sequences was used as a reference for read filtering using BBmap tool from BBtools (https://jgi.doe.gov/data-and-tools/software-tools/bbtools/bb-tools-user-guide/bbmap-guide/) specifying the parameters "outm" and "pairedonly". The output was converted in paired fastq files using the tool reformat.sh, from BBtools and after quality processed with Trim Galore!. Since the output reads were still considerably large, they were subsampled using the script extractfq.py from MitoZ (G. Meng, Li, Yang, and Liu, 2019), specifying the parameter “-size_required 5”.

The trimmed and subsampled reads were subsequently used for the individual loci assembly in SRAssembler, using ABySS as the assembler and specifying each Unioverse probe reference loci individually at each run with the parameter “-t dna -Z 450 -R 5000 -A 1 -k 15:10:65 -S 1 -s drosophila -G 1 -i 200 -m 200 -M 5000 -e 0.5 -c 0.8 -n 20 -a 2 -x 2 -b 4 -d 500 -z 1 -y -f” (for a detailed description see https://github.com/BrendelGroup/SRAssembler).

All the individual assemblies were concatenated and the names of the sequences transformed to resemble the names generated by Bridger assembler (e.g., L{$}_{$}__comp0_seq0) to be used in the post-assembly processing AHE pipeline developed from (Breinholt et al., 2018). Subsequently, this new assembly was added to the IBA assemblies and, the combined dataset was processed using the (Breinholt et al., 2018) pipeline, as described in the section above.

2.3.3. Mitochondrial genomes data processing

The mitogenomes were assembled using NovoPlasty v4 (Dierckxsens et al., 2016) using a COI gene sequence retrieved from NCBI as reference bait for each sample. Annotations were obtained using MITOS2 web server (Bernt et al., 2013b) and visually validated by comparison with the other Margaritiferidae mtDNA available on NCBI.

2.4. Alignment construction

2.4.1. AHE datasets

To access the correct positioning of the assemblies produced by the new assembly strategy, an initial dataset that included the *M. margaritifera* AHE probe regions assemblies produced from RNA-seq, was constructed (Figure P4.S1). However, RNA-seq based assemblies only encompass exonic regions, whereas many non-exonic regions are represented in the AHE dataset by sequenced regions that flank the genomic regions targeted by the AHE probes. Consequently, the RNA-seq sample was excluded from the remaining alignments, which were all composed of both the exonic and hypervariable flanking regions.

Individual loci were aligned to reference sequences using MAFFT and only the loci with over 70% of gene occupancy were kept. The python scripts alignment_DE_trim.py and flank_dropper.py (Breinholt et al., 2018) were used to trim and filter sequences and then split using extract_probe_region.py. This resulted in three alignment files, one corresponding to the probe region (hereafter referred to as probe) and two corresponding to flanking regions (hereafter referred to as head for the 5' flanking regions and tail for the 3' flanking region), that were visually inspected using AliView v 1.26 (Larsson, 2014). The final datasets were realigned using MAFFT and were concatenated using both the probe and flanking regions ("head+probe+tail"), except for the dataset including RNA-seq data, which was solely composed of the probe region ("probe"). Finally, trimAL v. 1.2rev59 (Capella-Gutiérrez et al., 2009) was used to remove positions with gaps in 50% or more of the sequences.

2.4.2. Mitogenomes alignments

The newly sequenced Female Type (F-type) mitogenomes and the 17 F-type previously sequenced mitogenomes available on NCBI were used (Table P4.1). The nucleotide sequences of the protein-coding genes (PCG) (except atp8) and the two ribosomal RNA genes (rRNA) were aligned using GUIDANCE2 (Penn et al., 2010) with the MAFFT v.7 multiple sequence alignment algorithms, specifying the following parameters: score algorithm: GUIDANCE2; bootstrap replicates 100; sequence cut-off score: 0.0 (no sequence removal); column cut-off score: below 0.8; site masking score: below 0.6 (for codon and amino acids alignments) and 0.8 (for the rRNA alignments). The resulting alignments were concatenated to include both 12 PCG and the two rRNA ("PCG+rRNA").

2.5. Phylogenomic analyses

All phylogenomic inferences for each AHE distinct dataset were performed using a Maximum Likelihood (ML) implemented with IQ-TREE v 1.6.12 (Nguyen et al., 2015). Best-fit substitution models were inferred with ModelFinder (Kalyaanamoorthy et al., 2017), implemented in IQ-TREE using a 10% relaxed clustering (option -rcluster 10) (Lanfear et al., 2014). IQ-TREE analyses were conducted with 10 independent runs of an initial tree search and 10000 ultrafast bootstrap replicates (ufBS). For AHE and AHE+mtDNA combined datasets, substitution models were estimated in ModelFinder with no partition. For the mtDNA dataset, substitution models were obtained specifying each PCG codon position as a single partition and the two rRNA genes as two independent partitions.

2.6. AHE probe region functional annotation

All the individual reference probe sequences from *B. platifrons* (i.e., L1 to L813 ref sequences) were aligned to the genome assembly inferred transcripts (Sun et al., 2017), using blastn v.2.11.0 (Camacho et al., 2009). After, transcripts with 100% identity were filtered in the annotation table of the *B. platifrons* genome (Sun et al., 2017) and linked to their corresponding probe references code. Using the final annotation table, a frequency count of probe per *B. platifrons* gene was produced. To obtain a list of pfam descriptions, hmmfetch h3.1b2 (Punta et al., 2012) was used to retrieve the HMM(s) profile of each AHE gene using the list of pfam IDs from the annotations file. Gene Ontology IDs were used to produce GO Terms Classification Count, applying module "find_enrichment.py" of the python library GOATOOLS (Klopfenstein et al., 2018), by using a list of Unioverse loci as the "study" and "population" and specifying the parameter "--pval=1.00 --prt_study_gos_only --no_propagate_counts". To access putative functional biases in the final AHE loci used for the final Margaritiferidae matrix alignment, "find_enrichment.py" was also applied to these loci, using the same parameters.

The complete *B. platifrons* proteins corresponding to the Unioverse loci were used for Benchmarking Universal Single-Copy Orthologs (BUSCO) search. For that, the recently released version, i.e., BUSCO v5.2.2 which includes a curated list of Mollusca Universal Single-Copy Orthologs (Manni et al., 2021) was used. In total, three curated lists were searched for near-universal single-copy orthologous, eukaryotic (n = 255), metazoa (n = 978), and mollusca (n = 5295). The annotation of the matching results was acquired from OrthoDB v10 (Kriventseva et al., 2019).

3. Results

3.1. Sequencing outputs and data description

All newly sequenced AHE sequencing reads are available on the GenBank SRA database (BioProject PRJNA895396). Furthermore, mitogenomes are available on GenBank. Further details on the data usage for different datasets, along with respective sources and accession numbers, are provided in Table P4.1. The provisory link for reviewers is https://dataview.ncbi.nlm.nih.gov/object/PRJNA895396?reviewer=fu7ulv4vd37us50mo1qecvmc10.

3.1.1. AHE assembly and probe capturing

The number of initial assembled sequences for each sample ranged between 493 to 1241, with the two samples from Mulleriidae and Iridinidae showing the lowest number of sequences (< 550) and two Unionidae samples showing the largest numbers (> 1000), with duplication rates ranging from 1.07 to 2.30 (Table P4.2). Within Margaritiferidae samples, an average of 764 sequences was observed, ranging between 631 and 903, with the smallest number observed for the RNA-seq-based assembly of *M. margaritifera* (hereafter referred to as “RNAassembly”) and the largest number observed in one of the *M. falcata* samples (Table P4.2). The assembly resulting from the pipeline here designed (hereafter referred to as “SRAassembly”) has a total of 790 sequences and shows the highest number of individual captured loci (81% Table P4.2). Furthermore, the RNA assembly also shows the smallest mean sequence length (177.8 bp) of all samples, while the SRAassembly shows the third largest mean sequence length (1334.0 bp).

Table P4. 2 - General statistics of AHE assemblies and probe capturing count. * The pipeline to generate matrices alignments followed (Breinholt et al., 2018).

Family	Species	Sample Code	Total assmbled sequences	Mean assmbled sequence length	Assembly Duplication rate	Captured loci count	Captured loci count (% of Unioverse probes)	Loci maintained after data processing pipeline *
Mulleriidae	*Anodontitinae elongata*	UA 20859	510	770.0	1.36	374	46.12	298
Iridinidae	*Mutela dubia*	BIV5573	493	1018.9	1.23	400	49.32	330
Hyriidae	*Echyridella menziesii*	BIV5908	858	1164.1	1.47	584	72.01	512
Margaritiferidae	*Gibbosula laosensis*	BIV2533	737	1222.9	1.26	584	72.01	522
Margaritiferidae	*Pseudunio marocanus*	BIV2634	776	1313.3	1.33	585	72.13	513
Hyriidae	*Westralunio carteri*	BIV2395	799	967.6	1.37	585	72.13	511
Margaritiferidae	*Margaritifera falcata*	BIV4860	741	1233.4	1.27	585	72.13	521
Margaritiferidae	*Margaritifera falcata*	BIV5536	903	1314.5	1.54	586	72.26	468
Margaritiferidae	*Margaritifera middendorffi*	BIV4862	760	1271.0	1.30	586	72.26	517
Margaritiferidae	*Margaritifera dahurica*	BIV4858	735	1230.3	1.25	587	72.38	524
Margaritiferidae	*Margaritifera margaritifera*	SRR5230899	631	177.8	1.07	588	72.50	519
Margaritiferidae	*Pseudunio homsensis*	BIV4859	738	1056.6	1.25	589	72.63	519
Margaritiferidae	*Pseudunio auricularius*	BIV1998	792	1344.7	1.34	589	72.63	519
Margaritiferidae	*Gibbosula laosensis woodthorpi*	BIV5493	750	1247.0	1.27	589	72.63	522
Margaritiferidae	*Margaritifera laevis*	BIV2689	814	874.7	1.38	590	72.75	513
Margaritiferidae	*Pseudunio homsensis*	BIV5537	748	1173.7	1.27	590	72.75	518
Margaritiferidae	*Margaritifera marrianae*	MmarEsc008	700	709.6	1.18	591	72.87	526
Margaritiferidae	*Margaritifera hembeli*	UF 521837	731	1127.0	1.24	591	72.87	525
Margaritiferidae	*Cumberlandia monodonta*	BIV2740	798	1041.9	1.35	592	73.00	525
Margaritiferidae	*Gibbosula crassa*	BIV3457	816	1114.9	1.36	600	73.98	536
Unionidae	*Potomida littoralis*	PL702	840	1267.5	1.39	604	74.48	546
Unionidae	*Anodonta anatina*	BIV3399	886	1419.9	1.46	606	74.72	554
Unionidae	*Nodularia douglasiae*	UA 20694	1396	1059.0	2.30	608	74.97	553
Unionidae	*Lampsillis siliquoidea*	LhigMis001	1241	1144.2	2.02	613	75.59	565
Margaritiferidae	*Margaritifera margaritifera*	BIV4481	790	1334.0	1.20	660	81.38	558

The largest number of loci captured after the assemblies was observed in the SRAassembly with 660 loci (Table P4.2). In total, the number of loci captured after the assemblies ranged between 484 and 660 for Margaritiferidae samples, with an average of 593 loci (Table P4.2). Outside the Margaritiferidae family, loci captured varied between 374 and 613, with an average of 546 but with a clear distinctiveness between families, with Unionidae showing the highest number of captured loci (Table P4.2).

3.1.2. Mitochondrial genomes characteristics

All the newly sequenced whole mitogenomes, of *Margaritifera marrianae* Johnson, 1983, *Margaritifera hembeli* (Conrad, 1838), *Margaritifera middendorffi* (Rosen, 1926), *Margaritifera laevis* (Haas,1910), *Margaritifera falcata* (Gould, 1850), *Pseudunio auricularius* (Spengler, 1793), *Pseudunio homsensis* (Lea, 1864), *Gibbosula laosensis* (Lea, 1863), *G. laosensis* ssp. *Woodthorpi* Godwin-Austen, 1919 and *Cumberlandia monodonta* (Say, 1829) (Table P4.1), include the typical 13 protein-coding genes (PCGs), 22 transfer RNA (tRNA) and two ribosomal RNA (rRNA) genes. The gene order of all newly sequenced mitogenomes is the expected for Margaritiferidae F-type mtDNA, i.e., MF1 (Lopes-Lima et al., 2017a).

3.2. Phylogenetic analyses

The general characteristics of the alignments used for the various phylogenetic inferences are reported in Table P4.3. All the phylogenetic tree files can be found in the supplementary information (Data S1–S3).

Table P4. 3 - General characteristics of the alignment matrices used for phylogenetic inferences using AHE and mtDNA. + Includes a *Margaritifera margaritifera* sample originating from RNA-seq probe assembly (RNAassembly). * TPM2u+F+I+G4: 1atp6 1nad3, GTR+F+I+G4: 2atp6 2nad4 2nad4l 2nad5, TPM3+F+R3: 3atp6 3nad3, TIM2+F+I+G4: 1cob 1nad1 rrnL rrnS, GTR+F+I+G4: 2cob 2nad1, TIM3+F+R3: 3cob, TN+F+I+G4: 1cox1 1cox2 1cox3 2nad3, TVM+F+R2: 2cox1 2cox2 2cox3, TPM2u+F+R3: 3cox1, K3Pu+F+I+G4: 3cox2 3cox3, TN+F+G4: 3nad1 3nad2 3nad6, TIM2+F+I+G4: 1nad2 1nad6, TPM3u+F+G4: 2nad2 2nad6, TIM2+F+I+G4: 1nad4 1nad4l 1nad5, K3Pu+F+G4: 3nad4, K3Pu+F+I+G4: 3nad4l, TPM2u+F+R2: 3nad5

Data Source	Dataset	Phylogentic inference tool	Number of Samples	Number of loci	Length (nt)	Missing data (%)	ModelFinder infered substitution model (BIC)	Figure
AHE	head+probe+tail	IQ-TREE	24	1541	455738	19.565	GTR+F+R3	Fig. 1a
AHE + RNA-seq⁺	probe	IQ-TREE	25	514	98681	8.295	TVM+F+R3	Fig. S1
mtDNA	F-Type	IQ-TREE	27	14	12923	1.408	*	Fig 1b

3.2.1. AHE datasets

The two AHE-based phylogenetic inferences reveal the same topology, recovering the glochidia-bearing mussels (i.e., (Hyriidae (Margaritiferidae+Unionidae)) as sister to the lasidia-bearing mussels (Mulleriidae+Iridinidae) (Figure P4.1a, Figure P4.S1). Margaritiferidae is two subfamily-level groups, one containing taxa belonging to the genus *Gibbosula* and the other including the remaining genera. The monotypic genus *Cumberlandia* is sister to the group including all *Pseudunio* species, which is sister to *Margaritifera*. The *Margaritifera* group is further divided into a clade including *M. margaritifera* and *M. dahurica* which is sister to a clade encompassing the remaining species, with *M. falcata* sister to the reciprocally monophyletic groups represented by *M.* marrianae + *M.* hembeli and *M. middendorffi* + *M. laevis* (Figure P4.1a, Figure P4.S1). In the phylogeny inferred using the RNAassembly and SRAassembly samples, both *M. margaritifera* samples were grouped with maximum support (Figure P4.S1). All the above-described splits show maximum support for both phylogenies (Figure P4.1a, Figure P4.S1).

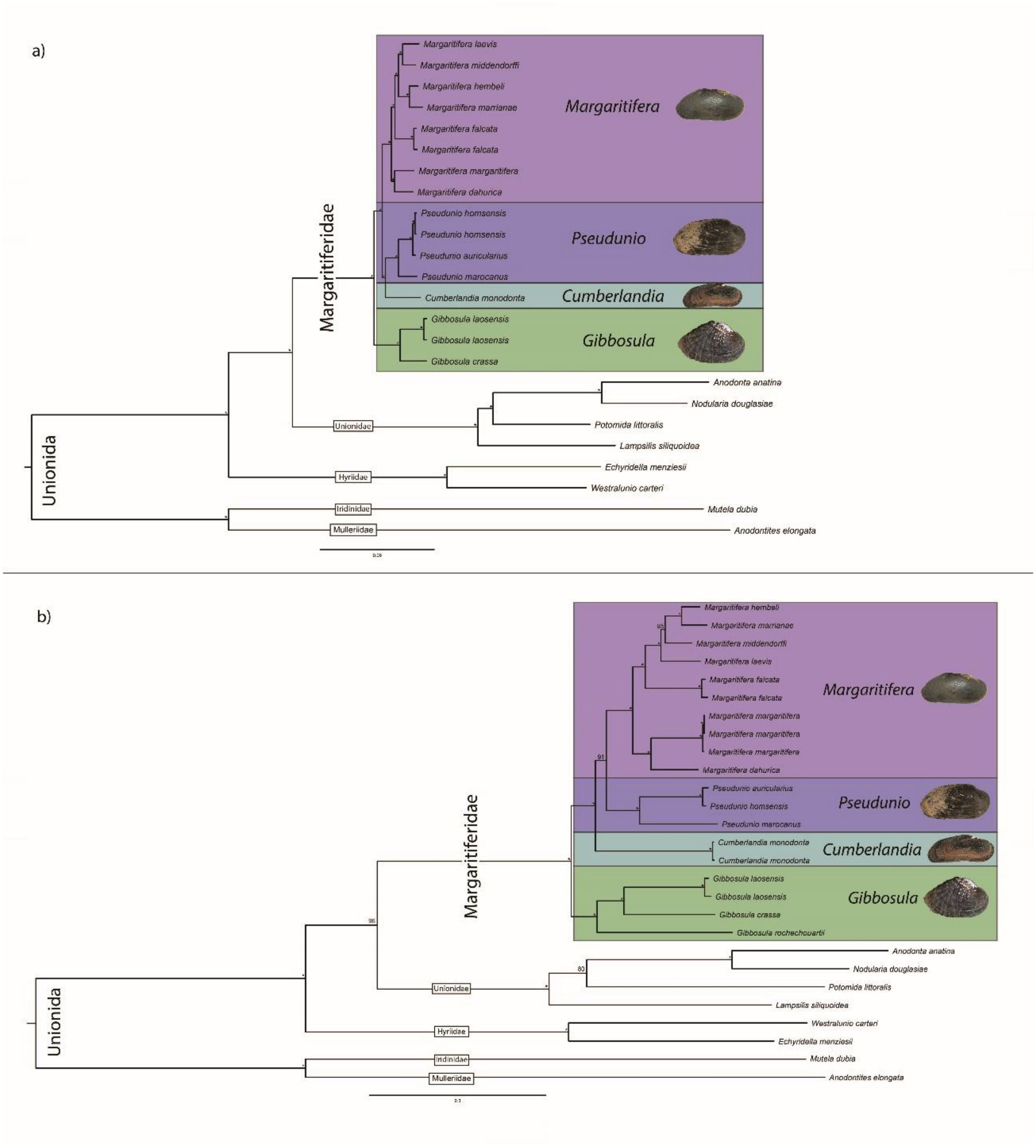


Figure P4. 1 - Maximum Likelihood phylogenetic trees of family Margaritiferidae based on: a) concatenated alignments of 514 Unioverse loci (head+probe+tail; n=1541); b) concatenated alignments of the mitochondrial DNA from 12 PCG and 2 rRNA; *Above the nodes refer to bootstrap with maximum support.

3.3. Mitogenome datasets

The mtDNA F-type derived phylogeny (with the outgroup taxa *A. elongata* and *M. dubia*) is identical to the AHE phylogenies, except for genus-level relationships (Figure P4.1b). All the margaritiferid genera are still recovered as monophyletic with *Gibbosula* as sister to the remaining genera, but the sister group of *Cumberlandia* is *Pseudunio* + *Margaritifera* (vs only *Pseudunio* in the AHE topology) (Figure P4.1b).

3.4. Unioverse probe regions functional annotation

The blast search of the Unioverse loci to the *B. platifrons* transcripts resulted in a total of 811 probe regions being assigned with a single hit at a unique position with 100% identity (Tables 5.S1-S2). A total of 460 *B. platifrons* transcripts were linked to loci, with 507 AHE probe regions occurring on different parts of the same gene, and the remaining 304 AHE probe regions occurring on 304 independent genes (Table P4.S1, Figure P4.2). The average number of loci per gene is around 1.76, with the largest observed number of loci assigned to the same gene being 19 (for loci L21-L35 and transcript Bpl_scaf_47521-1.5), covering nearly 40% of the whole transcript (Table P4.S1, Figure P4.2).

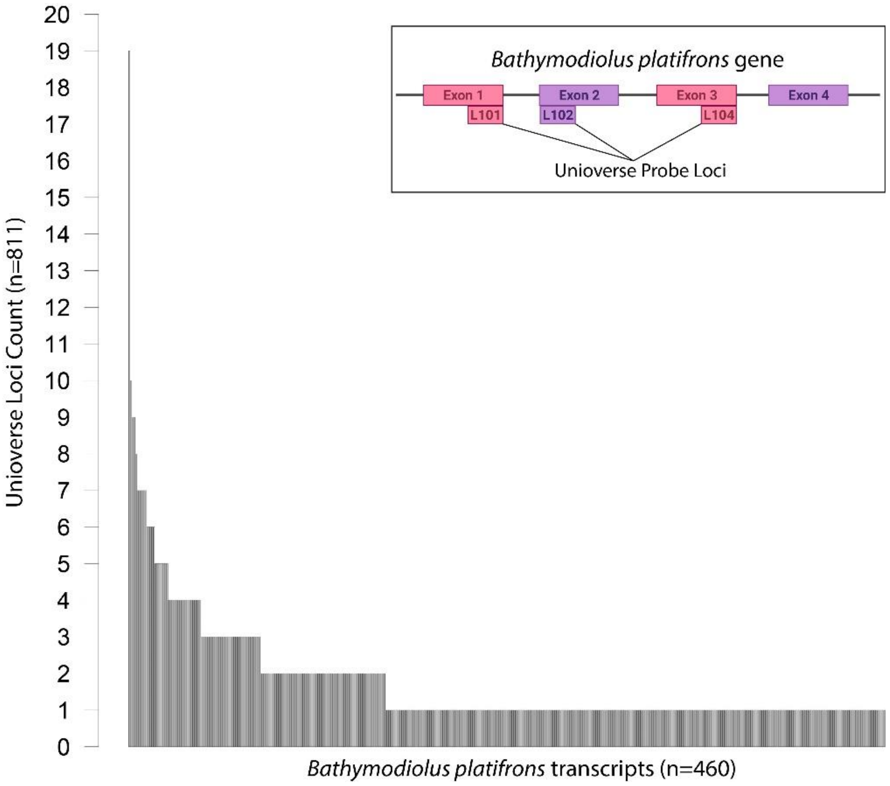


Figure P4. 2 - Frequency distribution of Unioverse probe regions per *Bathymodiolus platifrons* whole transcripts. Top right corner is a schematic representation of several probe regions originating from the same gene.

The complete functional annotation of the probe regions (i.e., from the corresponding to the *B. platifrons* transcripts) is presented in Table P4.S3, which includes the putative genes descriptions, Gene Ontology IDs, InterPro IDs, pfam IDs and Description and Kyoto Encyclopedia of Genes and Genomes (KEGG) Orthology (KO).

GOATOOLS GO Terms Classification Count provided a classification for a total of 649 GO IDs corresponding to a total of 607 loci (Table P4.S4). The GO Terms assignment to the three main GO categories showed 47.61% (n=309) of IDs belonging to Biological Process (BP), followed by Molecular Functions (MF) with 34.05% (n=221) and Cellular Component (CC) with 18.34% (n=119) (Figure P4.3a). Within each of the main categories, the highest loci count was classified to ATP binding, within MF with 105 counts, followed by metabolic process, within BP with 77 counts followed by cytoplasm, within CC with 59 counts (Table P4.S4). The same overall percentages were maintained in the final 514 loci AHE Margaritiferidae alignment dataset, i.e., BP with 46.91% (n=228), MF with 33.95% (n=165), and CC with 19.13% (n=93) (Table P4.S5, Figure P4.3b) and within each category, the highest loci count were for ATP binding with 64 counts within MF, metabolic process with 42 counts within BP and cytoplasm with 39 counts within CC (Table P4.S5).

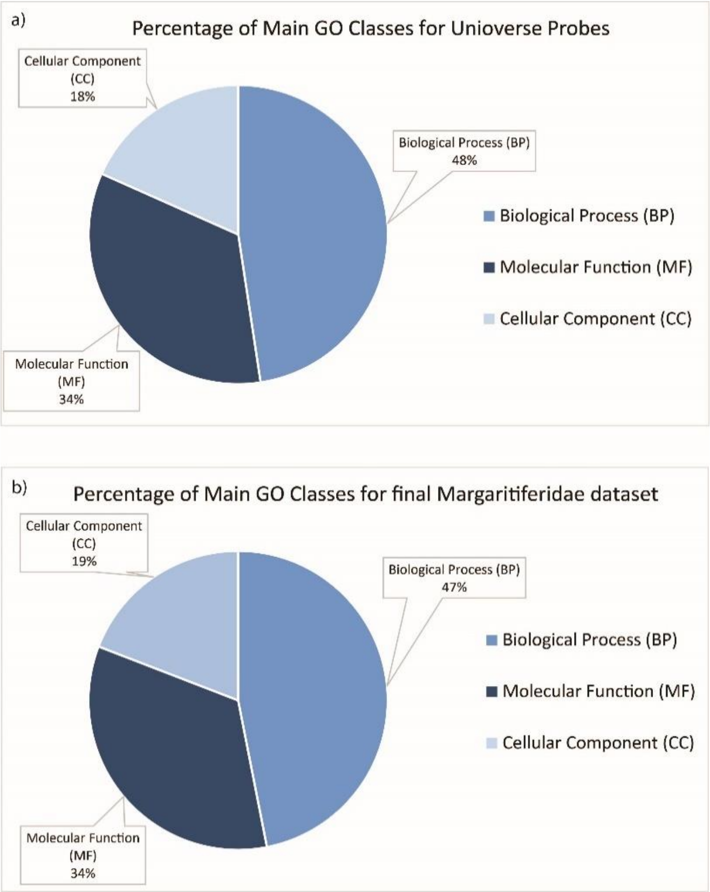


Figure P4. 3 - Percentages distributions of the Gene Ontology (GO) Terms main categories count for: a) All Unioverse probes; b) The Unioverse loci used for the final AHE Margaritiferidae samples alignments.

The Benchmarking Universal Single-Copy Orthologs (BUSCO) search revealed that around 73% of the probe regions match at least one of the three searched OrthoDB databases (Table P4.S6, Figure P4.4), with more than half of the loci being assigned to Mollusca, 15% assigned to metazoan and 5% assigned to eukaryote (Figure P4.4).

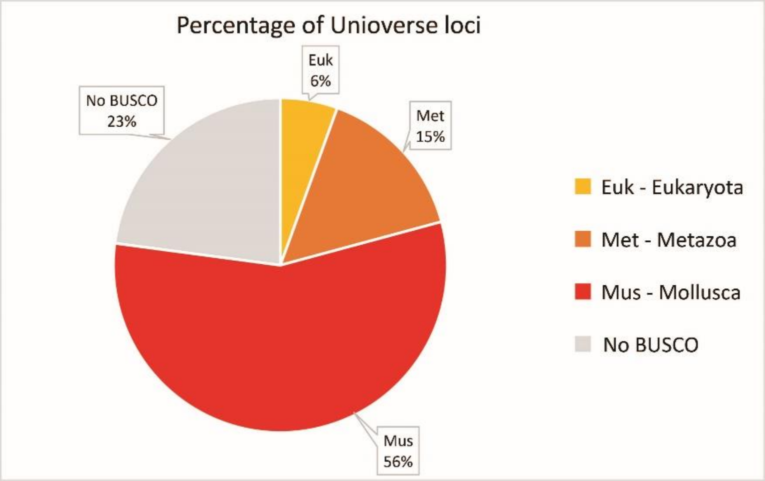


Figure P4. 4 - Percentage of total Unioverse loci assigned to three BUSCO databases. Euk - Dataset with 255 genes of Eukaryota library profile; Met - Dataset with 954 genes of Metazoa library profile; Mus - Dataset with 5295 genes of Mollusca library profile; No BUSCO – No match found.

4. Discussion

4.1. Assembly pipelines and probe capturing results

Here, we develop a new assembly strategy (referred to here as SRAassembly) for capturing and *de novo* assembly of the targeted regions using whole genome re-sequencing reads. Given no taxon-specific parameter is required, our method can be replicated to any other reduced representation sequencing dataset based on the reference sequences for region capturing. We demonstrated the utility of SRAassembly using the Unioverse probe set, and it consistently outperformed previous methods. SRAassembly had the best balance between the number of assembled sequences and the number of captured loci while generating large sequences (Table P4.2). Although many AHE samples show a higher number of initially assembled sequences, most of these sequences represented duplicates (Table P4.2), and SRAassembly assembled the highest proportion of individual captured loci (81%; Table P4.2). While keeping duplication levels low from the beginning and maintaining high loci capturing, SRAassembly reduces errors introduced by post-assembly duplication removal. Further, SRAassembly maintains the higher overall mean sequence length (1333.97bp), which has obvious advantages over RNA-seq methods (177.8 bp) that only capture the probe region (Breinholt et al., 2018; Lemmon et al., 2012; Pfeiffer et al., 2021, 2019) (Table

P4.2). Given these statistics, whole genome sequencing outputs are preferable over RNA-seq to augment target capture datasets.

At the date of this manuscript, high-quality reference genomes are still a scant resource for many non-model species However, the increasing accessibility to fast and affordable sequencing approaches will likely generate new whole genome assemblies followed by re-sequencing studies (Gomes-dos-Santos et al., 2020; Hotaling et al., 2021). Therefore, the availability of whole genome sequencing data will increase sharply, making this pipeline a timely tool for future studies. Further, the availability of reference genomes will provide the opportunity to make more robust reference gene sets for capture, which can easily be incorporated into our pipeline given that targeted capture is not a requirement for sequencing. Additionally, target sequencing may no longer represent an affordable option compared to WGS due to the constant and accentuated decline of sequencing costs. However, the availability of already carefully selected target regions represents a valuable set of markers for phylogenomics. Consequently, new pipelines that integrate WGS for target capturing, such as the one here presented, represent fundamental tools for future studies.

4.2. Phylogenetic analyses

Early margaritiferid systematics was based on highly variable or homoplastic morphological characters that often produced conflicting classifications (Lopes-Lima et al., 2018a). Molecular phylogenetic studies on the group, which have primarily been based on Sanger sequencing data, have dramatically improved the classification of the Margaritiferidae (Araujo et al., 2017; Bolotov et al., 2016; X. C. Huang et al., 2018). The family now includes two subfamilies, i.e., Margaritiferinae and Gibbosulinae, and four genera, i.e., *Margaritifera*, *Pseudunio*, *Cumberlandia*, and *Gibbosula* (Lopes-Lima et al., 2018a). Here we demonstrated the utility of SRAassembly by using AHE and novel whole genome sequencing data to reconstruct the evolutionary history of the group and complement it with a new set of whole mitogenome dataset. We produced the first Margaritiferidae family-wide AHE sequencing study while including representatives of all unioniod families. Although samples from the family Unionidae show an overall higher number of captured probe regions, the numbers observed within Margaritiferidae are similar, which was maintained after whole data processing pipeline to generate matrix alignments (Table P4.2). This reinforces the efficiency of the Unioverse AHE probe dataset in isolating the target regions across Margaritiferidae (Pfeiffer et al., 2019). Both AHE and mtDNA phylogenies retrieve the genus-level groups recently described (Lopes-Lima et al., 2018a), corresponding to the four genera and placing the *Gibbosula* clade sister to all the remaining genera (Figure P4.1a, b). However, the position of *Cumberlandia* concerning *Margaritifera* and *Pseudunio* differs among the phylogenies (Figure P4.1a, b). The results of the AHE phylogeny agree with the published works based on combined mitochondrial and nuclear markers (Araujo et al., 2017; Huff et al., 2004; Lopes-Lima et al., 2018a). On the other hand, the mitogenome phylogeny is congruent with other mitochondrial makers-based studies (Araujo et al., 2009; Gomes-dos-Santos et al., 2019; Inoue et al., 2014). These expected differences do not contradict the inferred relationships within genera, which have high support among both phylogenies (Figure P4.1a, b). Mitochondrial genomes (or markers) have many intrinsic features that make them attractive for phylogenetic inferences in Metazoa (Bernt et al., 2013a; Ghiselli et al., 2021), however, not always reflecting the evolutionary history of the species (Ghiselli et al., 2021; Hurst and Jiggins, 2005; Kern et al., 2020). Here, the consistent, well-supported but disagreeing results between nuclear and mtDNA markers, suggest a divergent evolutionary history of both markers. Moreover, given the notably low mitochondrial evolutionary rates observed within Margaritiferidae (Bolotov et al., 2016), it is unlikely that the results are due to nucleotide substitution saturation. On the other hand, the M-type phylogeny provided by (Gomes-dos-Santos et al., 2019) groups with maximum support for *Cumberlandia* and *Pseudunio* as sister clades, a relationship also observed here when using the AHE dataset. The split between the F and M type Unionida mitochondrial lineages can be traced back to the origin of the order, thus reflecting two independently evolving and phylogenetic informative units (Guerra et al., 2019). However, future studies should aim to increase M-type mitogenomes sequencing.

4.3. Unioverse probe region annotation

Phylogenetic inferences are highly sensitive to data selection and studies relying on reduced sequencing approaches commonly treating loci as independent regions (e.g., Hipp et al., 2014; Pfeiffer et al., 2019, 2021; Rosenfeld et al., 2016; Smith et al., 2020; Zhang et al., 2020). However, this may be unrealistic. Here we show that less than half (n=304; ~37%) of the regions targeted by the target capture set Unioverse originate from a single gene, with a considerable number of targets belonging to different regions of the same gene (n=507; ~63%) (Tables 4.S1-S2, Figure P4.2). Specifically, a total of 507 probe regions belong to different regions of the same genes (from 2 to 19 loci in the same gene), which may introduce bias when using individual locus methods. Although there is a high level of redundancy in gene targets, the entire coding sequence (CDS) for these genes may be assembled with *a priori* locus information provided in this study. This approach would allow for more conservative and less biased phylogenetic reconstruction by building alignments based on entire CDS rather than restricting the analyses to small, linked exons. We recommend future studies utilize this approach and capturing the CDS of targeted genes could be increased by using larger inserts in the sequencing library preparation (Lemmon et al., 2012).

We were able to provide functional annotations of the 811 exons targeted by Unioverse (Tables 4.S3-S6), which will allow future studies to employ functional validation of their datasets. Moreover, we complemented these annotations with pfam Descriptions, Gene Ontology categorization, and the Benchmarking Universal Single-Copy Orthologs (BUSCO) annotation. Due to the superimposed underlying assumptions, it is expected that the target capture methods will capture a significant number of genes within BUSCO databases (Lemmon et al., 2012; Lemmon and Lemmon, 2013; Manni et al., 2021). In total, 76% of the Unioverse AHE probe regions could be matched to three curated lists of BUSCO genes for Eukaryota, Metazoa, and Mollusca, with more than half of the hits being assigned to Mollusca (Figure P4.4). This further demonstrates the utility of the Unioverse probe set for phylogenetic inferences (Pfeiffer et al. 2019) and highlights the possibility of the probe set being integrated with extensive BUSCO databases to test macroevolutionary hypotheses. This demonstrates the importance of annotating target capture probe sets, which can expand their utility outside targeted focal groups (Fleming and Arakawa, 2021; Johnston et al., 2019; Y. Li et al., 2021; Pfeiffer et al., 2019; X. X. Shen et al., 2016; Taylor et al., 2020; Zhao et al., 2020).

5. Conclusion

In this study, we provide a new assembly strategy highly efficient for target capturing that allows us to easily combine AHE datasets with the increasingly emerging whole genome sequencing outputs. To explore the results of the new pipeline we provide a phylogenetic study with a comprehensive sampling of the most threatened Unionida family, Margaritiferidae. Furthermore, we provide a complementary phylogenetic analysis using whole mitogenome assemblies. Moreover, we provide a thorough structural and functional annotation of the Unioverse AHE probes dataset, that will allow future studies to adjust loci sampling to their desire.

Data Accessibility

The raw reads for each AHE sample were deposited at the NCBI Sequence Read Archive (see Table P4.1 for accessions), under BioProject PRJNA895396. Whole mitogenome assemblies were deposited in GenBank (see Table P4.1 for accessions). The provisory link for reviewers is https://dataview.ncbi.nlm.nih.gov/object/PRJNA895396?reviewer=fu7ulv4vd37us50mo1qecvmc10. The structural and functional annotation tables of the AHE probe regions are provided in the Supplementary Data (Tables 4.S1-S6). All software with respective versions and parameters used to assemble the AHE loci here presented are listed in the methods section. Software programs with no parameters associated were used with the default settings.

Acknowledgments

Portuguese Foundation for Science and Technology (FCT) funded AGS (SFRH/BD/137935/2018), MLL (2020.03608.CEECIND) and EF (CEECIND/00627/2017). This research was developed under the project EdgeOmics - Freshwater Bivalves at the edge: Adaptation genomics under climate-change scenarios (PTDC/CTA-AMB/3065/2020) funded by FCT through national funds. Additional strategic funding was provided by FCT UIDB/04423/2020 and UIDP/04423/2020. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the United States Government. The findings and conclusions in this study are those of the authors and do not necessarily represent the views of the United States Fish and Wildlife Service. The RSF supported I.N.B. (No. 19-04-00066). The Ministry of Science and Higher Education funded A.V.K. and I.V.V. (No. FUUW-2022-0056).

Supplementary material

Below is the link to the electronic supplementary material.

https://doi.org/10.22541/au.166799900.04572038/v1

https://www.dropbox.com/sh/mfye70025ccq3lg/AAAWdrkB7kJwmLfHECcYrS16a?dl=0

Table P4.S1 – List of *Bathymodiolus platifrons* transcripts names and their corresponding Unioverse loci.

Table P4.S2 – Tabular blastn output resulting from blast search of *Bathymodiolus platifrons* Unioverse loci against all *Bathymodiolus platifrons* transcripts. qseqid - query or source sequence id; sseqid - subject or target sequence id; pident - the percentage of identical matches; length - alignment length (sequence overlap); mismatch - number of mismatches; gapopen - number of gap openings; qstart - start of alignment in query; qend - end of alignment in query; sstart - start of alignment in the subject; send - end of alignment in the subject; evalue - expect value; bitscore - bit score.

Table P4.S3 – Unioverse loci functional annotations extracted from the *Bathymodiolus platifrons* whole genome annotation: GO_ID - Gene Ontology (GO) IDs; KEGG_ko - Kyoto Encyclopedia of Genes and Genomes Orthology.

Table P4.S4 – GO Terms Classification Count for all Unioverse loci produced by the GOATOOLS module “find_enrichment.py”.

Table P4.S5 – GO Terms Classification Count for the Unioverse AHE probes used in the final Margaritiferidae matrices alignments produced by the GOATOOLS module “find_enrichment.py”.

Table P4.S6 –Unioverse AHE probes functional annotation from three BUSCO databases. Euk - Dataset with 255 genes of Eukaryota library profile; Met - Dataset with 954 genes of Metazoa library profile; Mus - Dataset with 5295 genes of Mollusca library profile.

2.3. Paper 5 – The gill transcriptome of threatened European freshwater mussels

Gomes-dos-Santos, A., Machado, A.M., Castro, L.F.C., Prié, V., Teixeira, A., Lopes-Lima, M., Froufe, E., 2022. The gill transcriptome of threatened European freshwater mussels. Scientific Data 9, 494, 1–10. https://doi.org/10.1038/s41597-022-01613-x

The gill transcriptome of threatened European freshwater mussels

André Gomes-dos-Santos [1,2], André Machado M. [1,2], L. Filipe C. Castro [1,2], Vincent Prié [3], Amílcar Teixeira [4], Manuel Lopes-Lima [1,5,6], Elsa Froufe [1]

[1] CIIMAR/CIMAR — Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Terminal de Cruzeiros do Porto de Leixões, Avenida General Norton de Matos, S/N, P 4450-208 Matosinhos, Portugal; [2] Department of Biology, Faculty of Sciences, University of Porto, Rua do Campo Alegre 1021/1055, 4169-007 Porto, Portugal; [3] National Museum of Natural History (MNHN), CNRS, SU, EPHE, UA CP 51, 57 rue Cuvier, 75005 Paris, France; [4] Centro de Investigação de Montanha (CIMO), Instituto Politécnico de Bragança, Bragança, Portugal; [5] CIBIO/InBIO - Research Center in Biodiversity and Genetic Resources, Universidade do Porto, Campus Agrário de Vairão, Rua Padre Armando Quintas, 4485-661 Vairão, Portugal; [6] IUCN SSC Mollusc Specialist Group, c/o IUCN, David Attenborough Building, Pembroke St., Cambridge, England

* Corresponding authors

DOI: https://doi.org/10.1038/s41597-022-01613-x

Abstract

Genomic tools applied to non-model organisms are critical to design successful conservation strategies of particularly threatened groups. Freshwater mussels of the Unionida order are among the most vulnerable taxa and yet almost no genetic resources are available. Here, we present the gill transcriptomes of five European freshwater mussels with high conservation concern: *Margaritifera margaritifera*, *Unio crassus*, *Unio pictorum*, *Unio mancus* and *Unio delphinus*. The final assemblies, with N50 values ranging from 1069–1895 bp and total BUSCO scores above 90% (Eukaryote and Metazoan databases), were structurally and functionally annotated, and made available. The transcriptomes here produced represent a valuable resource for future studies on these species’ biology and ultimately guide their conservation.

1. Background & Summary

Ever since genomics approaches have been applied to non-model organisms, they have been recognized as fundamental tools to study biodiversity and guide conservation actions, coining the term conservation genomics (Allendorf et al., 2010; Formenti et al., 2022; Hohenlohe et al., 2021; Meek and Larson, 2019). Genomic data provides a comprehensive and accurate framework enhancing the characterization of genetic variation, population structure and dynamics, selective pressures and adaptative traits that ultimately guide and prioritize applied conservation efforts (Allendorf et al., 2010; Formenti et al., 2022; Hohenlohe et al., 2021; Meek and Larson, 2019). Furthermore, genomic data are fundamental to construct predictive models to access the impact of human-mediated threats, such as biological invasions, resource depletion, and climate change (Allendorf et al., 2010; Hohenlohe et al., 2021; McCartney et al., 2022).

Freshwater mussels (Order Unionida) are molluscs extremely important to freshwater ecosystems where they play key ecological roles, such as nutrient and energy cycling and retention (Lopes-Lima et al., 2014a; Vaughn, 2017; Vaughn et al., 2015). They also provide important direct (e.g., as food, pearls, and other raw materials) and indirect (e.g., water clearance, sediment mixing) services to humans (Lopes-Lima et al., 2014a; Vaughn, 2017; Vaughn et al., 2015). These organisms are among the most threatened worldwide, with many species near extinction (Cuttelod et al., 2011; Lopes-Lima et al., 2017c, 2018b). Of the thousand known species, only four whole genomes (Gomes-dos-Santos et al., 2021; Renaut et al., 2018; Rogers et al., 2021; Smith, 2021) and less than 20 transcriptomes are available (Bertucci et al., 2017; Capt et al., 2018, 2019; Chen et al., 2019; Cornman et al., 2014; D. Huang et al., 2019; Luo et al., 2014; Patnaik et al., 2016; Robertson et al., 2017; Roznere et al., 2018; R. Wang et al., 2015; X. Wang et al., 2017; Q. Yang et al., 2021). Of these, only one is from the European continent (Bertucci et al., 2017). Here, we produce reference transcriptomes of five European species as baseline tools to support future studies. Genomic tools, such as transcriptomes, are key resources to study evolutionary and adaptive traits. Examples include, in the case of freshwater mussels, the unique obligatory parasitic interaction with a freshwater fish host (and occasionally other vertebrates), essential to disperse their larvae and complete the life cycle or the response to human-mediated threats, including climate change and habitat degradation (Lopes-Lima et al., 2014a, 2017c). Moreover, these species are ecological indicators, and the transcriptomes provide a catalogue of key genes and pathways, related to important stressors (e.g., temperature, oxygen availability), as well as basic mechanisms underlying freshwater mussel’s stress adaptation (Bertucci et al., 2017; Ganser et al., 2015; Haney et al., 2020; Luo et al., 2014; Robertson et al., 2017; Roznere et al., 2018).

We present the gill transcriptome of the most emblematic freshwater pearl mussel, *Margaritifera margaritifera* (Linnaeus, 1758). This species was famous as a source of pearls throughout the last two millennia (Gomes-dos-Santos et al., 2021). Currently, is among the most threatened freshwater mussel species in Europe, with many populations suffering massive declines, with up to 90% of European populations depleted by the 90 s, which is reflected in the current scattered distribution (Geist, 2010) (Figure P5.1). Recently, a whole-genome assembly was published (Gomes-dos-Santos et al., 2021), adding to unique transcriptomic dataset of a very specialized tissue (i.e., kidney, Bertucci et al., 2017). The current species conservation status is Endangered by the IUCN and is also listed in the EC Habitats Directive (Moorkens et al., 2017). The other four transcriptomes are from the *Unio* genus, the type genus of the order Unionida, i.e., *Unio delphinus* Spengler, 1793, *Unio crassus* Philipsson in Retzius, 1788, *Unio pictorum* (Linnaeus, 1758) and *Unio mancus* Lamarck, 1819, for which no genomic resources have been produced at all. Two of these species, i.e., *U. crassus* and *U. pictorum*, although widely distributed (Figure P5.1), have also suffered recent declines, with *U. crassus*, once considered the most abundant unionid in Europe, now listed as Endangered by the IUCN and also listed in the EC Habitats Directive (Lopes-Lima et al., 2014b). The other two species have much more restricted distributions (Figure P5.1), both suffering strong population losses, with *U. delphinus* listed as Near Threatened and *U. mancus* as Endangered by the IUCN (Araujo, 2011; Lopes-Lima and Seddon, 2014). The depleted conservative state of Unionida mussels is a global concern, being the second group with the highest percentage of threatened species (43%) and the group with the highest number of wild extinct species (6.3%) (Díaz et al., 2019).

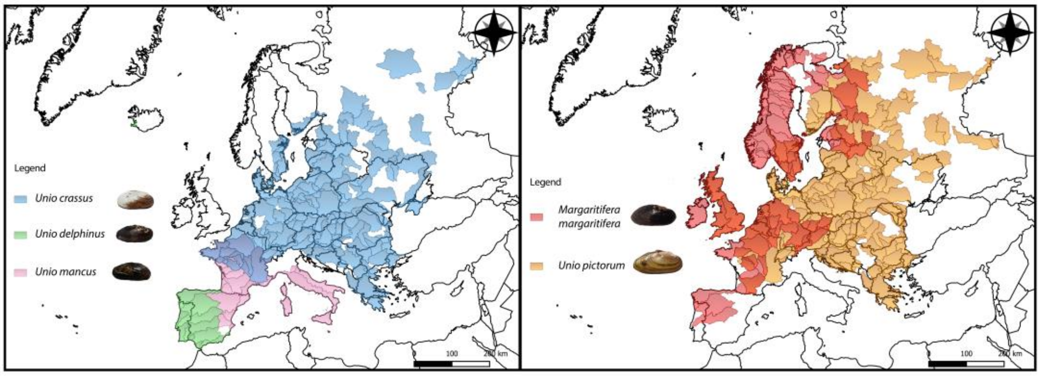


Figure P5. 1 - Maps of the five species’ potential distributions produced by overlapping points of recent presence records (obtained from Lopes-Lima et al., 2017c) with the Hydrobasin level 5 polygons (Lehner and Grill, 2013). Overlapping distribution polygons between *Unio mancus* and *Unio crassus* are represented by a light purple shade, in the left panel. Overlapping distribution polygons between *Unio pictorum* and *Margaritifera margaritifera* are represented by an orange shade, in the right panel.

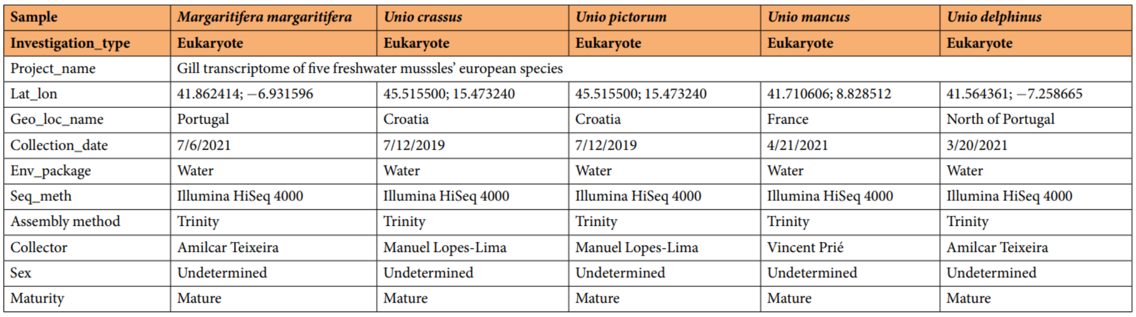
In this context, increasing the genomic resources available for freshwater mussels, particularly of European species, is vital. The transcriptomes produced here offer a unique opportunity to explore and decipher the capability of these species to cope with current and future threats and ultimately guide conservation genomic studies to protect this highly threatened group of organisms.

2. Methods

2.1. Animal sampling

One individual of *M. margaritifera* was collected from the Tuela River in Portugal, one *U. crassus*, and one *U. pictorum* from the Dobra River in Croatia, one *U. mancus* from the Taravu River in France and one *U. delphinus* from the Rabaçal River in Portugal (Table P5.1), all adult individuals. Differentiated tissues were promptly flash frozen and stored at −80 °C, at CIIMAR tissue and mussels’ collection, as well as their respective shells.

Table P5. 1 - MixS descriptors for the five freshwater mussel species.



Sample	***Margaritifera margaritifera***	***Unio crassus***	***Unio pictorum***	***Unio mancus***	***Unio delphinus***
Investigation_type	**Eukaryote**	**Eukaryote**	**Eukaryote**	**Eukaryote**	**Eukaryote**
Project_name	Gill transcriptome of five freshwater musssles’ european species				
Lat_lon	41.862414; −6.931596	45.515500; 15.473240	45.515500; 15.473240	41.710606; 8.828512	41.564361; −7.258665
Geo_loc_name	Portugal	Croatia	Croatia	France	North of Portugal
Collection_date	7/6/2021	7/12/2019	7/12/2019	4/21/2021	3/20/2021
Env_package	Water	Water	Water	Water	Water
Seq_meth	Illumina HiSeq 4000	Illumina HiSeq 4000	Illumina HiSeq 4000	Illumina HiSeq 4000	Illumina HiSeq 4000
Assembly method	Trinity	Trinity	Trinity	Trinity	Trinity
Collector	Amilcar Teixeira	Manuel Lopes-Lima	Manuel Lopes-Lima	Vincent Prié	Amilcar Teixeira
Sex	Undetermined	Undetermined	Undetermined	Undetermined	Undetermined
Maturity	Mature	Mature	Mature	Mature	Mature

2.2. RNA extraction, library construction, and sequencing

Total RNA of gills was extracted using the NZY Total RNA Isolation kit (NZYTech, Lda. - Genes and Enzymes), following the manufacturer’s instructions. RNA concentration (ng/µl) and quality measurement (OD260/280 ratio values) were obtained using a DS-11 Series Spectrophotometer/Fluorometer (*M. margaritifera* - 380.75 ng/µl, *U. crassus* – 478.290 ng/µl, *U. pictorum* - 375.461 ng/µl, *U. mancus* - 225.815 ng/µl, *U. delphinus* – 230.234 ng/µl). The extracted total RNA from the five samples was sent to Macrogen, Inc to build strand-specific libraries, with an insert size of 250–300 bp and sequenced using 150 bp paired-end reads on the Illumina HiSeq 4000 platform.

2.3. Pre-assembly processing

Raw reads datasets for each sample were first inspected with FastQC (version 0.11.8) software (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Afterwards, reads were quality-filter and Illumina adaptors were removed using Trimmomatic (version 0.38) (Bolger et al., 2014), using the parameters LEADING:5 TRAILING:5 SLIDINGWINDOW:5:20 MINLEN:36 (Figure P5.2-3). Trimmed reads were corrected for random sequencing errors using a kmer-based error correction approach in Rcorrector (version 1.0.3) (Song and Florea, 2015) with default parameters and after imported to Centrifuge (version 1.0.3-beta) (Kim et al., 2016) to taxonomically classify them using a pre-compiled nucleotide database from NCBI (ftp://ftp.ccb.jhu.edu/pub/infphilo/centrifuge/data/) (version nt_2018_3_3). All reads whose classification did not belong to the Mollusca superclass (Taxon Id: 6447) were removed (Figure P5.3).

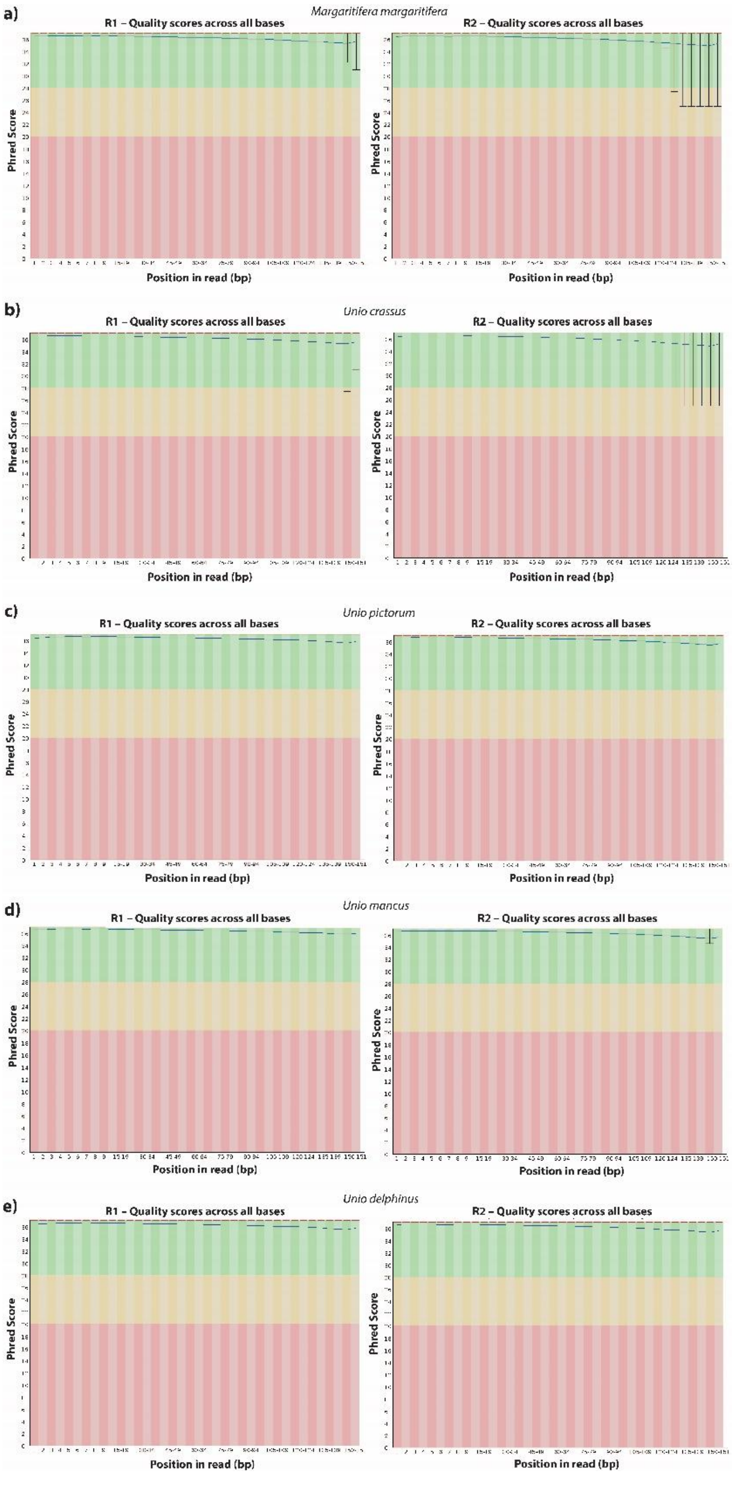


Figure P5. 2 - FastQC quality report of the trimmed and decontaminated RNA-seq reads (after Centrifuge for each species. (a) *Margaritifera margaritifera*; (b) *Unio crassus*; (c) *Unio pictorum*; (d) *Unio mancus*; and (e) *Unio delphinus*.

2.4. *De novo* transcriptome assembly

The fully processed reads were used for the whole transcriptome *de novo* assembly for each sample, with Trinity (version 2.13.2) (Grabherr et al., 2011; Haas et al., 2013) using the default parameters. To ensure the removal of contamination, the assembled transcripts were queried against the nucleotide database of NCBI (NCBI-nt; (Download; 24/08/2021) (Agarwala et al., 2016) and Univec (Download; 02/04/2019) databases using BLAST-n (version 2.11.0) (Camacho et al., 2009) (Figure P5.3). Afterwards, transcripts that held a minimum alignment length of 100 bp, an e-value cut-off of 1e-5, identity score of 90%, and a match to Mollusca phylum (NCBI: taxid 6447) or without matches at all, were retained. On the other hand, transcripts matching other taxa in the NCBI-nt database or any match to the Univec database were considered contaminants and removed from the datasets.

2.5. Redundancy removal

Before proceeding to open reading frame (ORF) prediction, transcript redundancy was removed using a hierarchical contig clustering approach, implemented with Corset (version 1.0.9) (Davidson and Oshlack, 2014). For that, raw reads for each sample were mapped onto their respective transcriptome assemblies using Bowtie2 (version 2.3.5) (parameter: –no-mixed –no-discordant –end-to-end –all –score-min L,− 0.1,− 0.1). After that, Corset was used to cluster contigs, filtered redundancies, and exclude any transcripts containing less than 10 mapped reads. The overall quality of the five transcriptomes (before and after redundancy removal) was assessed for completeness, using Benchmarking Universal Single-Copy Orthologs tool (BUSCO version 3.0.2) with the lineage-specific libraries for Eukaryota and Metazoa (Simão et al., 2015) and for structural integrity using TransRate (version 1.0.3) (Smith-Unna et al., 2016) (Figure P5.3).

2.6. Open reading frame prediction and transcriptome annotation

The open reading frames (ORFs) for each non-redundant transcriptome, were produced using Transdecoder (version 5.3.0) (https://transdecoder.github.io/) (Figure P5.3). During the ORF prediction process, the homology and protein searches were performed in UniProtKB/Swiss-Prot (Bateman et al., 2017) and PFAM databases (Punta et al.,

2012) using the Blast-p (version 2.12.0) (Camacho et al., 2009) and hmmscan of hmmer2 package (version 2.4i) (Finn et al., 2011) software, respectively. Next, the Gtf/Gff Analysis Toolkit (AGAT) (version 0.8.0) (Dainat et al., 2020) was applied to produce the structural annotation file (in gff3 format) from the Transdecoder output file (.gff) and transcriptome assembly file (.fasta). In the end, the AGAT tool was used to extract the protein and transcript fasta files with the names properly uniformized and formatted per species. Afterwards, the functional annotation was performed with InterProScan tool (version 5.44.80) and Blast-n/p/x searches in several databases. While the proteins per species were queried against InterPro (Download; 30/03/2019) and protein databases of NCBI (NCBI-RefSeq – Reference Sequence Database (Download; 10/03/2022) (Pruitt et al., 2007) NCBI-nr – non-redundant database of NCBI (Download; 15/12/2021) (Agarwala et al., 2016) with the Blast-p/x tool of DIAMOND software (version version 2.0.13) (Buchfink et al., 2015), the transcripts were searched by Blast-n/x in NCBI-nt and NCBI-nr databases, with Blast-n tool of NCBI and Blast-x tool of DIAMOND software. In the end, all blast (outfmt6 files) and InterProScan (tsv file) outputs were integrated into the gff3 annotation file with the AGAT tool. The putative gene name per sequence was assigned based on the best blast hit (Gene symbol – NCBI Accession Number) and following the ranking: 1- Blast-p Hit in RefSeq database; 2 - Blast-p Hit in NCBI-nr database; 3 - Blast-x Hit in NCBI-nr database; 4 - Blast-n Hit in NCBI-nt database.

3. Data Records

The raw reads for each sample were deposited at the NCBI Sequence Read Archive with the accessions numbers: SRR19261768 (MM), SRR19261764 (UD), SRR19261767 (UP), SRR19261765 (UM), SRR19261766 (UC); the BioSample accessions numbers: SAMN28495338 (MM), SAMN28495283 (UD), SAMN28495235 (UP), SAMN28495263 (UM), SAMN28495214 (UC) and under BioProject PRJNA839062. The remaining information was uploaded to figshare (https://doi.org/10.6084/m9.figshare.19787566.v2). In detailed, the files uploaded to figshare include, the filtered trinity redundant assemblies (_trinity_filtered.fasta), the non-redundant transcriptomes (_transcriptome.fa), transcripts files (_genes.fa), messenger RNA file (_mrna.fa), open reading frames predictions (_cds.fa), open reading frames proteins predictions (_proteins.fa) as well as the annotation files (_annotation_sorted.gff3.gz).

4. Technical Validation

4.1. Raw datasets and pre-assembly processing quality control

The raw sequencing outputs resulted in a total of 131051306 million reads (M) for *M. margaritifera*, 132002266 M for *U. crassus*, 104108396 M for *U. pictorum*, 100704688 M for *U. mancus*, and 112439686 M for *U. delphinus*. Although the initial overall quality of raw data was considerably good (Figure P5.2), the datasets were further improved by quality trimming (Trimmomatic), error-correction (Rcorrector), and decontaminated (Centrifuge) (Figure P5.2,3). The number of reads removed during the pre-assembly processing represented less than 3% of each dataset (Table P5.2) and the overall Phred scores were all above 25 (Figure P5.2a–e).

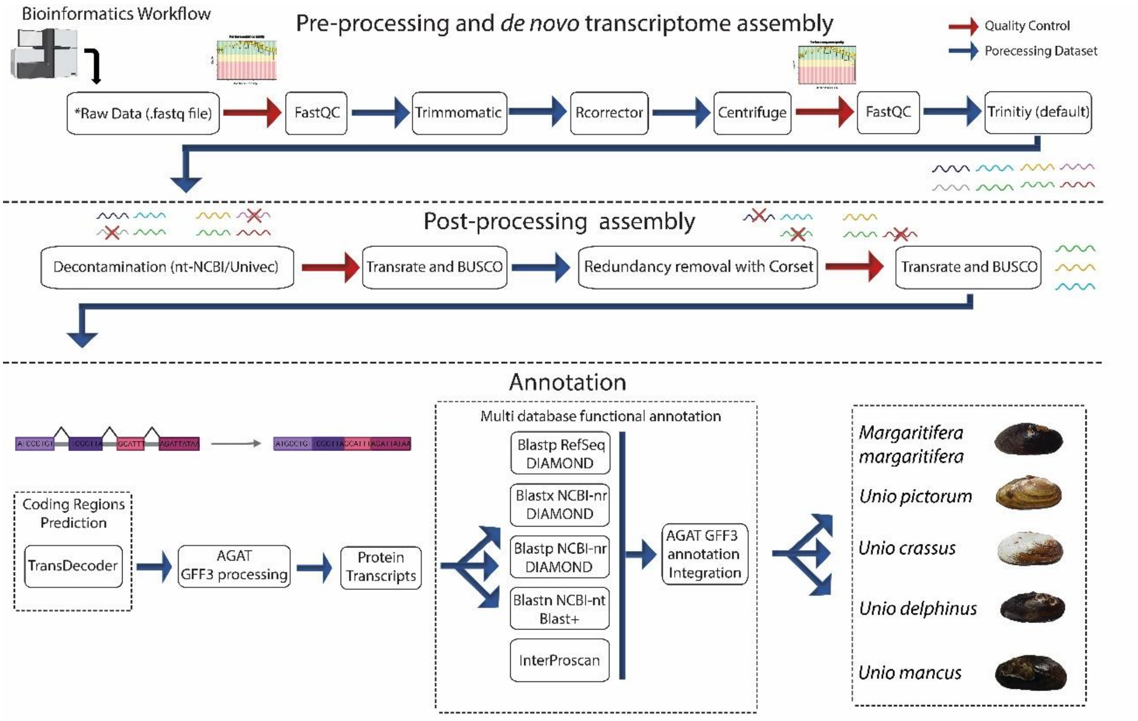
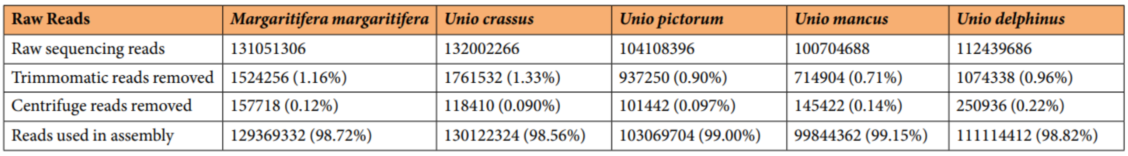


Figure P5. 3 - Bioinformatics pipeline applied for the transcriptome assembly and annotation. Auxiliary representative figures were created with BioRender.com.

Table P5. 2 - Basic statistics of raw sequencing datasets and percentages of removed reads at each step of the preassembly processing strategy.

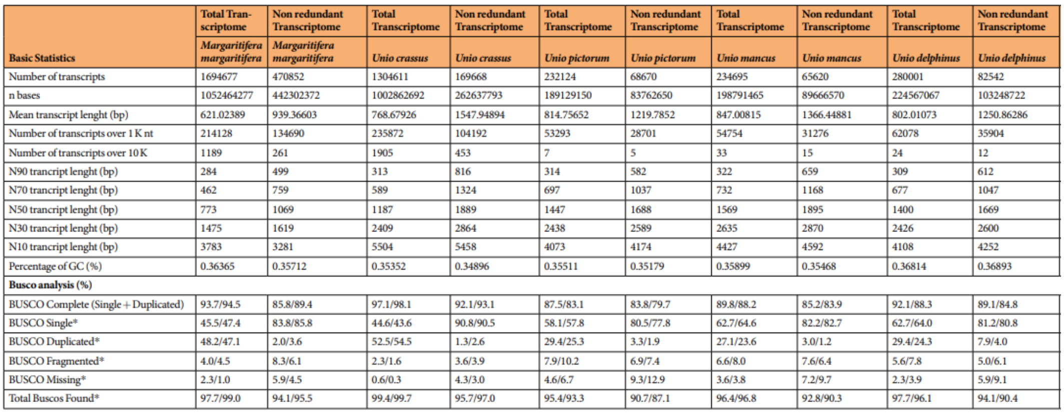


Raw Reads	***Margaritifera margaritifera***	***Unio crassus***	***Unio pictorum***	***Unio mancus***	***Unio delphinus***
Raw sequencing reads	131051306	132002266	104108396	100704688	112439686
Trimmomatic reads removed	1524256 (1.16%)	1761532 (1.33%)	937250 (0.90%)	714904 (0.71%)	1074338 (0.96%)
Centrifuge reads removed	157718 (0.12%)	118410 (0.090%)	101442 (0.097%)	145422 (0.14%)	250936 (0.22%)
Reads used in assembly	129369332 (98.72%)	130122324 (98.56%)	103069704 (99.00%)	99844362 (99.15%)	111114412 (98.82%)

4.2. Transcriptome assembly metrics

The *de novo* transcriptome assemblies were performed using Trinity, with default paraments, which has been successfully applied for other Unionida transcriptome assembly projects (Bertucci et al., 2017; Patnaik et al., 2016; Roznere et al., 2018; X. Wang et al., 2017; Q. Yang et al., 2021). Furthermore, the overall completeness of the transcriptome assemblies was evaluated using Benchmarking Universal Single-Copy Orthologs (BUSCO), by searching the Eukaryota (n:303) and Metazoa (n:978) near-universal single-copy orthologs databases, for all species. The overall metrics for each transcriptome *de novo* assembly, as well as their corresponding BUSCO scores, are presented in Table P5.3. The general assembly metrics of *U. pictorum*, *U. mancus*, and *U. delphinus* are very similar, both in the number of transcripts (~250,000) and N50 values (>1400 bp) (Table P5.3). On the other hand, *M. margaritifera* and *U. crassus* transcriptomes, have a much higher number of assembled transcripts (>1,000,000) and, consequently lower N50 lengths (Table P5.3). However, all these values are within the reported for other Unionida transcriptomes assembly projects (Bertucci et al., 2017; Capt et al., 2018, 2019; Cornman et al., 2014; D. Huang et al., 2019; Luo et al., 2014; Patnaik et al., 2016; Roznere et al., 2018; R. Wang et al., 2015; X. Wang et al., 2017). Furthermore, *M. margaritifera* and *U. crassus* transcriptome assemblies also have a considerably high level of duplicated BUSCO scores, i.e., around 50%, compared with the remaining species which presented values around 30% (Table P5.3). The percentage of total genes found (complete + fragmented) in all BUSCO analyses, for all species, was above 95%, except for the *U. pictorum* transcriptome in the Metazoan lineage-specific profile library, which had a total of 93.3%. These results reveal that despite being produced from a single tissue the initial assemblies were highly efficient in capturing conserved and widely express genes, thus providing a highly complete gill transcriptomic repertoire.

Table P5. 3 - Transrate and Busco scores of redundant and non-redundant gill transcriptome assemblies for each species. *euk/met. Euk: Dataset with 303 genes of Eukaryota library profile. Met: Dataset with 978 genes of Metazoa library profile.



	Total Tran-scriptome	Non redundant Transcriptome	Total Transcriptome	Non redundant Transcriptome	Total Transcriptome	Non redundant Transcriptome	Total Transcriptome	Non redundant Transcriptome	Total Transcriptome	Non redundant Transcriptome
Basic Statistics	***Margaritifera margaritifera***	***Margaritifera margaritifera***	***Unio crassus***	***Unio crassus***	***Unio pictorum***	***Unio pictorum***	***Unio mancus***	***Unio mancus***	***Unio delphinus***	***Unio delphinus***
Number of transcripts	1694677	470852	1304611	169668	232124	68670	234695	65620	280001	82542
n bases	1052464277	442302372	1002862692	262637793	189129150	83762650	198791465	89666570	224567067	103248722
Mean transcript lenght (bp)	621.02389	939.36603	768.67926	1547.94894	814.75652	1219.7852	847.00815	1366.44881	802.01073	1250.86286
Number of transcripts over 1 K nt	214128	134690	235872	104192	53293	28701	54754	31276	62078	35904
Number of transcripts over 10 K	1189	261	1905	453	7	5	33	15	24	12
N90 trancript lenght (bp)	284	499	313	816	314	582	322	659	309	612
N70 trancript lenght (bp)	462	759	589	1324	697	1037	732	1168	677	1047
N50 trancript lenght (bp)	773	1069	1187	1889	1447	1688	1569	1895	1400	1669
N30 trancript lenght (bp)	1475	1619	2409	2864	2438	2589	2635	2870	2426	2600
N10 trancript lenght (bp)	3783	3281	5504	5458	4073	4174	4427	4592	4108	4252
Percentage of GC (%)	0.36365	0.35712	0.35352	0.34896	0.35511	0.35179	0.35899	0.35468	0.36814	0.36893
Busco analysis (%)										
BUSCO Complete (Single + Duplicated)	93.7/94.5	85.8/89.4	97.1/98.1	92.1/93.1	87.5/83.1	83.8/79.7	89.8/88.2	85.2/83.9	92.1/88.3	89.1/84.8
BUSCO Single*	45.5/47.4	83.8/85.8	44.6/43.6	90.8/90.5	58.1/57.8	80.5/77.8	62.7/64.6	82.2/82.7	62.7/64.0	81.2/80.8
BUSCO Duplicated*	48.2/47.1	2.0/3.6	52.5/54.5	1.3/2.6	29.4/25.3	3.3/1.9	27.1/23.6	3.0/1.2	29.4/24.3	7.9/4.0
BUSCO Fragmented*	4.0/4.5	8.3/6.1	2.3/1.6	3.6/3.9	7.9/10.2	6.9/7.4	6.6/8.0	7.6/6.4	5.6/7.8	5.0/6.1
BUSCO Missing*	2.3/1.0	5.9/4.5	0.6/0.3	4.3/3.0	4.6/6.7	9.3/12.9	3.6/3.8	7.2/9.7	2.3/3.9	5.9/9.1
Total Buscos Found*	97.7/99.0	94.1/95.5	99.4/99.7	95.7/97.0	95.4/93.3	90.7/87.1	96.4/96.8	92.8/90.3	97.7/96.1	94.1/90.4

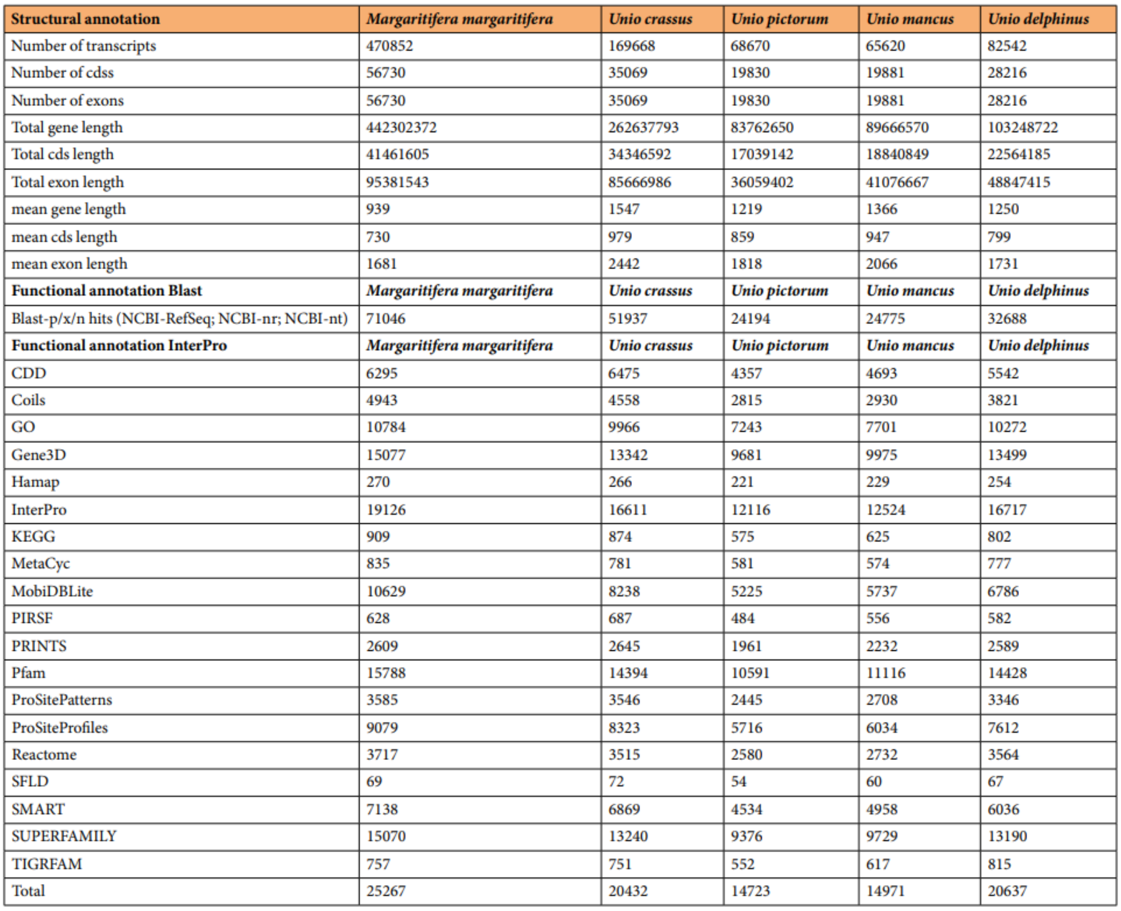
4.4. Post-assembly processing and annotation verification

The newly assembled transcriptomes were after subject to a decontamination process by Blast-n search against NCBI-nt and Univec databases. The Blast-n hits against NCBI-nt, were manually validated based on the reads with a minimum alignment length of 100 bp, an e-value of 1e-5, an identity score of 90% and a match to Mollusca phylum (NCBI: taxid 6447) or without matches at all, were retained. On the other hand, all Blast-n hits against Univec database were considered exogenous and removed. This decontamination approach has been routinely and successfully used by the team (e.g. Machado et al., 2022, 2020) and focuses the analyses on the identification, by homology, of putative contaminations and only excluded them if they are well supported and thus avoiding the exclusion of unambiguous matches.

Subsequently, before proceeding to the annotation, the decontaminated transcriptomes were subjected to redundancy removal using Corset. This software relies on hierarchical clustering of contigs that share read alignments and thus allows an unbiased removal of redundancy without discarding non-coding transcripts from the process (Davidson and Oshlack, 2014). The general transcriptome metrics after redundancy removal are presented in Table P5.3. Corset was extremely efficient in removing the redundancy from the filtered assemblies (Table P5.3). In fact, over 70% of the initial transcripts were removed during the process, suggesting that although Trinity was effective in producing a complete transcriptome assembly, it as has also generated several duplicated transcripts as well as many transcripts with low read support (Table P5.3). These results highlight the importance of using read clustering approach to remove redundancy, rather than simply relying on coding transcripts and selection of the largest isoform. The efficiency of the redundancy removal is also supported by the BUSCO analyses, where duplicated scores were on average 3.5% for Eukaryota (n:303) and 2.66% for Metazoa (n:978) after Corset, in opposition to an average 37.32% for Eukaryota (n:303) and 34.96% for Metazoa (n:978) before redundancy removal (Table P5.3). Furthermore, redundancy removal did not impact the overall completeness of the transcriptome assemblies, which still maintained the total BUSCO scores of over 90% (Table P5.3). In the end, the final gill transcriptomes were significantly reduced, fairly complete and cleared of putative errors introduced during the assembly, thus properly adjusted for annotation.

TransDecoder prediction of transcripts with an assigned ORF, resulted in a total of 56,730 for *M. margaritifera*, 35,069 for *U. crassus*, 19,830 for *U. pictorum*, 19,881 for *U. mancus*, and 28,216 for *U. delphinus* (Table P5.4). These predictions were performed in the non-redundant transcriptomes and were deposited in FigShare (https://doi.org/10.6084/m9.figshare.19787566.v2). Finally, the results of the functional annotation are presented in Table P5.4, where a thorough listing of hits counts from distinct databases used in the functional annotation processes is presented. The number of transcripts functionally annotated was InterProScan:25,267; Blast:71,046 for *M. margaritifera*, InterProScan:20,432; Blast:51,937 for *U. crassus*, InterProScan:14,723; Blast:24,194 for *U. pictorum*, InterProScan:14,971; Blast:24,775 for *U. mancus* and InterProScan:20,637; Blast:32,688 for *U. delphinus* (Table P5.4). These values are within the observed values for other Unionida genomics projects, both in transcriptomes (Bertucci et al., 2017; Capt et al., 2018; D. Huang et al., 2019; Luo et al., 2014; Patnaik et al., 2016; Roznere et al., 2018; X. Wang et al., 2017) and genome (Luo et al., 2014; Renaut et al., 2018; Rogers et al., 2021; Smith, 2021). Particularly for *M. margaritifera*, the number of genes functionally annotated, is very similar to the values obtained for the annotated genome assembly available for the species, i.e., 26,836 transcripts (Gomes-dos-Santos et al., 2021).

Table P5. 4 - Structural and functional annotation statistics for the final gill transcriptome assemblies for each species.



Structural annotation	***Margaritifera margaritifera***	***Unio crassus***	***Unio pictorum***	***Unio mancus***	***Unio delphinus***
Number of transcripts	470852	169668	68670	65620	82542
Number of cdss	56730	35069	19830	19881	28216
Number of exons	56730	35069	19830	19881	28216
Total gene length	442302372	262637793	83762650	89666570	103248722
Total cds length	41461605	34346592	17039142	18840849	22564185
Total exon length	95381543	85666986	36059402	41076667	48847415
mean gene length	939	1547	1219	1366	1250
mean cds length	730	979	859	947	799
mean exon length	1681	2442	1818	2066	1731
Functional annotation Blast	***Margaritifera margaritifera***	***Unio crassus***	***Unio pictorum***	***Unio mancus***	***Unio delphinus***
Blast-p/x/n hits (NCBI-RefSeq; NCBI-nr; NCBI-nt)	71046	51937	24194	24775	32688
Functional annotation InterPro	***Margaritifera margaritifera***	***Unio crassus***	***Unio pictorum***	***Unio mancus***	***Unio delphinus***
CDD	6295	6475	4357	4693	5542
Coils	4943	4558	2815	2930	3821
GO	10784	9966	7243	7701	10272
Gene3D	15077	13342	9681	9975	13499
Hamap	270	266	221	229	254
InterPro	19126	16611	12116	12524	16717
KEGG	909	874	575	625	802
MetaCyc	835	781	581	574	777
MobiDBLite	10629	8238	5225	5737	6786
PIRSF	628	687	484	556	582
PRINTS	2609	2645	1961	2232	2589
Pfam	15788	14394	10591	11116	14428
ProSitePatterns	3585	3546	2445	2708	3346
ProSiteProfiles	9079	8323	5716	6034	7612
Reactome	3717	3515	2580	2732	3564
SFLD	69	72	54	60	67
SMART	7138	6869	4534	4958	6036
SUPERFAMILY	15070	13240	9376	9729	13190
TIGRFAM	757	751	552	617	815
Total	25267	20432	14723	14971	20637

Overall, these results provide evidence of the quality and completeness of the five gill transcriptome assemblies, which represent timely needed genomic resources for this highly threatened group of organisms. Although future studies should also aim to obtain transcriptomic information from other tissues/development stages, these five annotated gill transcriptomes represent a valuable baseline tool to study these organisms and can ultimately help and guide future conservation actions.

Code availability

All software with respective versions and parameters used for producing the resources here presented (i.e., transcriptome assembly, pre and post-assembly processing stages, and transcriptome annotation) are listed in the methods section. Software programs with no parameters associated were used with the default settings.

Acknowledgements

AGS was funded by the Portuguese Foundation for Science and Technology (FCT) under the grant SFRH/ BD/137935/2018, that also supported MLL (2020.03608.CEECIND) and EF (CEECIND/00627/2017). This research was developed under the project EdgeOmics - Freshwater Bivalves at the edge: Adaptation genomics under climate-change scenarios (PTDC/CTA-AMB/3065/2020) funded by FCT through national funds. Additional strategic funding was provided by FCT UIDB/04423/2020 and UIDP/04423/2020.

2.4. Paper 6 – The Crown Pearl: a draft genome assembly of the European freshwater pearl mussel *Margaritifera margaritifera* (Linnaeus, 1758)

Gomes-dos-Santos, A., Lopes-Lima, M., Machado, A.M., Marcos Ramos, A., Usié, A., Bolotov, I.N., Vikhrev, I. v, Breton, S., Castro, L.F.C., da Fonseca, R.R., Geist, J., Österling, M.E., Prié, V., Teixeira, A., Gan, H.M., Simakov, O., Froufe, E., 2021. The Crown Pearl: a draft genome assembly of the European freshwater pearl mussel *Margaritifera margaritifera* (Linnaeus, 1758). DNA Research 28:2. https://doi.org/10.1093/dnares/dsab002

The Crown Pearl: a draft genome assembly of the European freshwater pearl mussel *Margaritifera margaritifera* (Linnaeus, 1758)

André Gomes-dos-Santos[1,2*], Manuel Lopes-Lima[1,3,4*], André M. Machado[1], António Marcos Ramos[5,15], Ana Usié[5,15], Ivan N. Bolotov[6], Ilya V. Vikhrev[6], Sophie Breton[7], L. Filipe C. Castro[1,2], Rute R. da Fonseca[8], Juergen Geist[9], Martin E. Österling[10], Vincent Prié[11], Amílcar Teixeira[12], Han Ming Gan[13], Oleg Simakov[14], Elsa Froufe[1*]

[1] CIIMAR/CIMAR — Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Terminal de Cruzeiros do Porto de Leixões, Avenida General Norton de Matos, S/N, P 4450-208 Matosinhos, Portugal; [2] Department of Biology, Faculty of Sciences, University of Porto, Rua do Campo Alegre 1021/1055, 4169-007 Porto, Portugal; [3] CIBIO/InBIO - Research Center in Biodiversity and Genetic Resources, Universidade do Porto, Campus Agrário de Vairão, Rua Padre Armando Quintas, 4485-661 Vairão, Portugal; [4] IUCN SSC Mollusc Specialist Group, c/o IUCN, David Attenborough Building, Pembroke St., Cambridge, England; [5] Centro de Biotecnologia Agrícola e Agro-alimentar do Alentejo/Instituto Politécnico de Beja (IPBeja), Beja 7801-908 Beja, Portugal; [6] Federal Center for Integrated Arctic Research, Russian Academy of Sciences, Arkhangelsk, Severnoy Dviny emb. 23 163000 Russia; [7] Department of Biological Sciences, University of Montreal, Campus MIL, Montreal, Canada; [8] Center for Macroecology, Evolution and Climate, GLOBE Institute, University of Copenhagen, 2100 Copenhagen, Denmark; [9] Aquatic Systems Biology Unit, Technical University of Munich, TUM School of Life Sciences, Mühlenweg 22, D-85354 Freising, Germany; [10] Department of Environmental and Life Sciences – Biology, Karlstad University, Universitetsgatan 2, 651 88 Karlstad, Sweden; [11] Research Associate, Institute of Systematics, Evolution, Biodiversity (ISYEB), National Museum of Natural History (MNHN), CNRS, SU, EPHE, UA CP 51, 57 rue Cuvier, 75005 Paris, France; [12] Centro de Investigação de Montanha (CIMO), Instituto Politécnico de Bragança, Bragança, Portugal; [13] GeneSEQ Sdn Bhd, Bandar Bukit Beruntung, Rawang 48300 Selangor, Malaysia; [14] Department of Neurosciences and Developmental Biology, University of Vienna, Universitätsring 1, 1010 Wien, Vienna, Austria; [15] MED — Mediterranean Institute for Agriculture, Environment and Development, CEBAL — Centro de Biotecnologia Agrícola e Agro-Alimentar do Alentejo, 7801-908 Beja, Portugal

* Corresponding authors

DOI: https://doi.org/10.1093/dnares/dsab002

Abstract

Since historical times, the inherent human fascination with pearls turned the freshwater pearl mussel *Margaritifera margaritifera* (Linnaeus, 1758) into a highly valuable cultural and economic resource. Although pearl harvesting in *M. margaritifera* is nowadays residual, other human threats have aggravated the species conservation status, especially in Europe. This mussel presents a myriad of rare biological features, e.g. high longevity coupled with low senescence and Doubly Uniparental Inheritance of mitochondrial DNA, for which the underlying molecular mechanisms are poorly known. Here, the first draft genome assembly of *M. margaritifera* was produced using a combination of Illumina Paired-end and Mate-pair approaches. The genome assembly was 2.4 Gb long, possessing 105,185 scaffolds and a scaffold N50 length of 288,726 bp. The *ab initio* gene prediction allowed the identification of 35,119 protein-coding genes. This genome represents an essential resource for studying this species’ unique biological and evolutionary features and ultimately will help to develop new tools to promote its conservation.

Keywords

Margaritifera margaritifera; freshwater mussel: pearls: unionida genome; whole genome

1. Introduction

Pearls are fascinating organic gemstones that have populated the human beauty imaginary for millennia. Legend says that Cleopatra, to display her wealth to her lover Marc Antony, dissolved a pearl in a glass of vinegar and drank it. The human use of pearls or their shell precursor material, nacre, is ancient. The earliest known use of decorative nacre dates to 4200 BC in Egypt, with pearls themselves only becoming popular around 600 BC. Before the arrival of marine pearls to Europe, most were harvested from a common and widespread freshwater bivalve, the freshwater pearl mussel *Margaritifera margaritifera* L. 1758 (Figure P6.1), where generally one pearl is found per 3,000 mussels leading to massive mortality (Hessling, 1859). During the Roman Empire period, pearls were a desirable luxury, so that it is believed that one of the reasons that persuaded Julius Caesar to invade Britain was to access its vast freshwater pearl resources (Strack, 2015). *M.margaritifera* freshwater pearls were extremely valuable being included in many royal family jewels, such as the British, Scottish, Swedish, Austrian, and German crown jewels and even in the Russian city’s coat of arms (Bespalaya et al., 2012; Makhrov et al., 2014; Schlüter and Rätsch, 1999; Strack, 2015). Although over-harvesting represented a serious threat to the species for centuries, there has been a decrease in interest and demand for freshwater pearls in the 20th century (Makhrov et al., 2014). However, the global industrialization process introduced stronger threats to the survival of the species (Geist, 2010; Lopes-Lima et al., 2018a; Moorkens et al., 2018). In fact, *M. margaritifera* belongs to one of the most threatened taxonomic groups on earth, the Margaritiferidae (Lopes-Lima et al., 2018a). The species was once abundant in cool oligotrophic waters throughout most of northwest Europe and northeast North America (Geist, 2010; Lopes-Lima et al., 2018a; Moorkens et al., 2018). However, habitat degradation, fragmentation, and pollution have resulted in massive population declines (Geist, 2010). Consequently, the Red List of Threatened Species from the International Union for Conservation of Nature has classified *M. margaritifera* as Endangered globally and Critically Endangered in Europe (Moorkens et al., 2018; Moorkens, 2018). Besides being able to produce pearls, *M. margaritifera* presents many other remarkable biological characteristics, e.g. is among the most longest-living invertebrates, reaching up to 280 years (Bauer, 1992; Lopes-Lima et al., 2018a); displays very weak signs of senescence, referred as the concept of ‘*negligible senescence*’ (Hassall et al., 2017); has an obligatory parasitic larval stage on salmonid fishes used for nurturing and dispersion (Geist, 2010; Lopes-Lima et al., 2017c); and, like many other bivalves (see Gusman et al., 2016) for a recent enumeration), shows an unusual mitochondrial DNA inheritance system, called Doubly Uniparental Inheritance or DUI (Breton et al., 2007, 2011b). Although these biological features are well described, the molecular mechanisms underlying their regulation and functioning are poorly studied and practically unknown. Thus, a complete genome assembly for *M. margaritifera* is critical for developing the molecular resources required to improve our knowledge of such mechanisms.

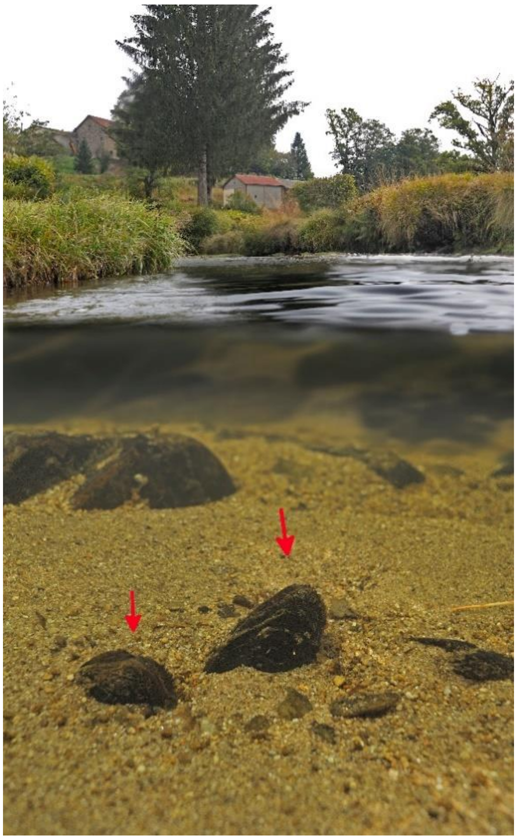


Figure P6. 1 - *Margaritifera margaritifera* specimens in their natural habitat. Red arrows point to two individuals.

To date, several Mollusca genomes are currently available and new assemblies are released every year at an increasing trend (reviewed in Gomes-dos-Santos et al., 2020; Hollenbeck and Johnston, 2018; Takeuchi, 2017) Despite this, to date, only three Unionida mussel genomes have been published, *Venustaconcha ellipsiformis* (Conrad, 1836) (Renaut et al., 2018), *Megalonaias nervosa* (Rafinesque, 1820) (Rogers et al., 2021), and *Potamilus streckersoni* (Smith, Johnson, Inoue, Doyle and Randklev, 2019) (Smith, 2021). Therefore, considering the importance of increasing the availability of genomic resources for Unionida, this study presents the first draft genome assembly of the freshwater pearl mussel *M. margaritifera.* The assembled genome has a total length of 2.4 Gb, a scaffold N50 length of 288,726 bp and 35,119 protein-coding genes were predicted. A Bivalvia phylogeny using whole-genome single copy orthologs was also constructed and the Hox and ParaHox gene complement within Unionida order was here characterized for the first time.

2. Materials and methods

2.1. Sample collection, DNA extraction, and sequencing

One female *M. margaritifera* (Linnaeus, 1758) specimen was collected from the River Tua, Douro basin in the North of Portugal (permit 284/2020/CAPT and fishing permit 26/20 issued by ICNF—Instituto de Conservação da Natureza e das Florestas). The whole individual is stored in 96% ethanol at the Unionoid DNA and Tissue Databank, CIIMAR, University of Porto. Genomic DNA (gDNA) was extracted from the foot tissue using DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) according to the manufacturer’s instructions.

Two distinct NGS libraries and sequencing approaches were implemented i.e. Illumina Paired-end reads (PE) and Illumina long insert size Mate-pair reads (MP). Illumina PE library preparation with standard Illumina adaptors used 100 ng of gDNA sheared to a length of 300–400 bp and was sequenced in an Illumina machine NovaSEQ6000 system located at Deakin Genomics Centre using a run configuration of 2 × 150 bp. Illumina MP library preparation and sequencing were performed by Macrogen Inc., Korea, where a 10 kb insert size Nextera Mate Pair Library was constructed and subsequently sequenced in a NovaSeq6000 S4 using a run configuration of 2 ×150 bp.

2.2. Genome size and heterozygosity estimation

The overall characteristics of the genome were accessed using PE reads. Reads quality was evaluated using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and raw reads were quality trimmed with Trim Galore v.0.4.0 (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/), allowing the trimming of adapter sequences and removal of low-quality reads using Cutadapt (Martin, 2011). Clean reads were used for genome size estimation with Jellyfish v.2.2.10 and GenomeScope2 (Ranallo-Benavidez et al., 2020) using *k*-mers lengths of 25 and 31.

2.3. Genome assembly and quality assessment

Long range Illumina MP quality processing was as described above and both PE and MP cleaned reads were used for whole-genome assembly. The assembly was produced by running Meraculous v.2.2.6 with several distinct *k*-mer sizes (meraculoususing) (Chapman et al., 2011). This allowed determining the optimal *k*-mer size of 101. Genome assembly metrics were estimated using QUAST v5.0.2 (Gurevich et al., 2013). Assembly completeness, heterozygosity, and collapsing of repetitive regions were evaluated through analysis of *k*-mer distribution using PE reads, with *K*-mer Analysis Toolkit (Mapleson et al., 2017). Furthermore, PE reads were aligned to the genome assembly using BBMap (Bushnell and Rood, 2018). BUSCO v. 3.0.2 (Simão et al., 2015) was used to provide a quantitative measure of the assembly completeness, with a curated list of eukaryotic (*n* = 303) and metazoan (*n* = 978) near-universal single-copy orthologous. Finally, in order to inspect the genome for possible contamination, we used BlobTools (Laetsch and Blaxter, 2017) (Additional File 1).

The whole mitochondrial genome was assembled using the PE reads with MitoBim v.1.9.0 (Hahn et al., 2013) and its annotation performed using MITOS2 (Bernt et al., 2013b) web server and manually validated against other Margaritiferidae mitogenomes.

2.4. Repeat sequences, gene models predictions, and transcriptome alignment

Given the generally high composition of repetitive elements in Mollusca genomes (e.g. Ref. Gomes-dos-Santos et al., 2020) they should be identified and masked before proceeding to genome annotation. A *de novo* library of repetitive elements was created for *M. margaritifera* genome assembly, using RepeatModeler v.2.0.1 (Smit and Hubley, 2015a) (excluding sequences <2.5 kb). Soft masking of the genome was performed with RepeatMasker v.4.0.7 (Smit and Hubley, 2015b) combining the *de novo* library with the ‘Bivalvia’ libraries from Dfam_consensus-20170127 and RepBase-20181026.

BRAKER2 pipeline v2.1.5 (Hoff et al., 2019, 2016) was used for gene prediction in the genome. First, all RNA-seq data of *M. margaritifera* (Bertucci et al., 2017; V. L. Gonzalez et al., 2015) available on GenBank were downloaded, assessed with FastQC v.0.11.8, quality trimmed with Trimmomatic v.0.38 (Bolger et al., 2014) (Parameters, LEADING: 5 TRAILING: 5 SLIDINGWINDOW: 4:20 MINLEN: 36) and error corrected with Rcorrector v.1.0.3 (Song and Florea, 2015). Afterwards, the RNA-seq data were aligned to the masked genome assembly, using Hisat2 v.2.2.0 with the default settings (Kim et al., 2015). The complete proteomes of 13 mollusc species, one Chordata (*Ciona intestinalis*) and one Echinodermata (*Strongylocentrotus purpuratus*) were downloaded from distinct public databases (Table P6.S1) and used as additional evidence for gene prediction. The BRAKER2 pipeline was applied with the parameters (–etpmode; –softmasking; –UTR=off; –crf; –cores =30) and following the authors' instructions (Hoff et al., 2019, 2016). The resulting gene predictions (i.e. gff3 file) were renamed, cleaned, and filtered using AGAT v.0.4.0 (Dainat et al., 2020), correcting coordinates of overlapping gene prediction, removing predicted coding sequence regions (CDS) with <100 amino acid (in order to avoid a high rate of false-positive predictions) and removing incomplete gene predictions (i.e. without start and/or stop codons). Functional annotation was conducted by searching for protein domain information using InterProScan v.5.44.80 (Quevillon et al., 2005) and protein blast search using DIAMOND v. 0.9.32 (Buchfink et al., 2015) against SwissProt (Download at 2/07/2020), TREMBL (Download at 2/07/2020), and RefSeq-NCBI (Download at 3/07/2020) (Boeckmann, 2003; Pruitt et al., 2007). BUSCO v. 3.0.2 (Simão et al., 2015) scores for the predicted proteins were accessed using the eukaryotic (*n* = 255) and metazoan (*n* = 954) curated lists of near-universal single-copy orthologous.

Finally, the *M. margaritifera* transcriptome assembly from Bertucci et al., (2017) downloaded from NCBI (BioProject: PRJNA369722) was aligned to the masked genome with pblat_v2.5 (Wang and Kong, 2019), specifying the option '-fine -q=rna' while maintaining the remaining parameters as default. Alignment stats were calculated with isoblat_v0.31 (Ryan, 2013) using default parameters.

2.5. Phylogenetic analyses

For the phylogenetic assessment, the proteomes of 12 molluscan species were downloaded from distinct public databases (Table P6.S2), which included 11 Autobranchia bivalves and 2 outgroup species, i.e. the Cephalopoda *Octopus bimaculoides* and Gastropoda *Biomphalaria glabrata* (Figure P6.3). Single-copy orthologous between these 12 species and *M. margaritifera* were retrieved using OrthoFinder v2.4.0 (Emms and Kelly, 2019), specifying multiple sequence alignment as the method of gene tree inference (-M). The resulting 118 single-copy orthologous sequences were individually aligned using MUSCLE v3.8.31 (Edgar, 2004), with default parameters and subsequently trimmed with TrimAl v.1.2 (Capella-Gutiérrez et al., 2009) specifying a gap threshold of 0.5 (-gt). Trimmed sequences were then concatenated using FASconCAT-G (https://github.com/PatrickKueck/FASconCAT-G). The best molecular evolutionary model was estimated using ProtTest v.3.4.1 (Abascal et al., 2005). Phylogenetic inferences were conducted in IQ-Tree v.1.6.12 (Nguyen et al., 2015) for Maximum-Likelihood analyses (with initial tree searches followed by 10 independent runs and 10,000 ultra-bootstrap replicates) and MrBayes v.3.2.6 (Ronquist et al., 2012) for Bayesian Inference (2 independent runs, 1,000,000 generations, sampling frequency of 1 tree per 1,000 generations). All phylogenetic analyses were applied using the substitution model LG+I + G.

2.6. Hox and ParaHox gene identification and phylogeny

To identify the repertoire Hox and ParaHox genes *in M. margaritifera*, a similarity search by BLASTn (Altschul et al., 1990) of the CDS of *M. margaritifera* genome, was conducted using the annotated homeobox gene set of *Crassostrea gigas* (Barton-Owen et al., 2018; Paps et al., 2015). Candidate CDSs were further validated for the presence of the homeodomain by CD-Search (Lu et al., 2020). Finally, each putative CDS identity was verified by BLASTx and BLASTp (Altschul et al., 1990) searches in Nr-NCBI nr database and phylogenetic analyses. Since the search was conducted in the annotated genome (i.e. scaffolds over 2.5 kb), when genes were not found, a new search was conducted in the remaining scaffolds. At the end, any genes still undetected were search in the Transcriptome assembly of the species (Bioproject: PRJNA369722) (Bertucci et al., 2017). Due to the phylogenetic proximity and for comparative purposes, Hox and ParaHox genes were also searched in the genome assembly of *M.nervosa* (Rogers et al., 2021).

For phylogenetic assessment of Hox and Parahox genes, amino acid sequences of homeodomain of the genes from *M. margaritifera* and *M. nervosa*, were aligned with other Mollusca orthologous (Huan et al., 2020; Y. Li et al., 2020). Molecular evolutionary models and Maximum-Likelihood phylogenetic analyses were obtained using IQ-TREE v.1.6.12 (Kalyaanamoorthy et al., 2017; Nguyen et al., 2015).

3. Results and discussion

3.1. Sequencing results

A total of 494 Gb (~209×) of raw PE and 76 Gb (~32×) of raw MP data were generated, which after trimming and quality filtering were reduced by 0.3% and 10%, respectively (Table P6.S3). GenomeScope2 model fitting of the *k*-mer distribution analysis estimated a genome size between 2.31–2.36 Gb and very low heterozygosity between 0.127–0.105% (Figure P6.2A). Although larger than the genome of *V. ellipsiformis* (Renaut et al., 2018) (i.e. 1.80 Gb), the size estimation of the *M. margaritifera* genome is in line with the recently assembled Unionida mussel *M. nervosa* (Rogers et al., 2021) (i.e. 2.38 Gb). The estimated heterozygosity is the lowest observed within Unionida genomes (Renaut et al., 2018; Rogers et al., 2021) and one of the lowest in Mollusca (Gomes-dos-Santos et al., 2020), which is remarkable considering it refers to a wild individual. This low value is likely a consequence of population bottlenecks during glaciations events, which have been shown to shape the evolutionary history of many freshwater mussels (e.g. Froufe et al., 2016b, 2016c; Renaut et al., 2018) and may also be enhanced by recent human-mediated threats.

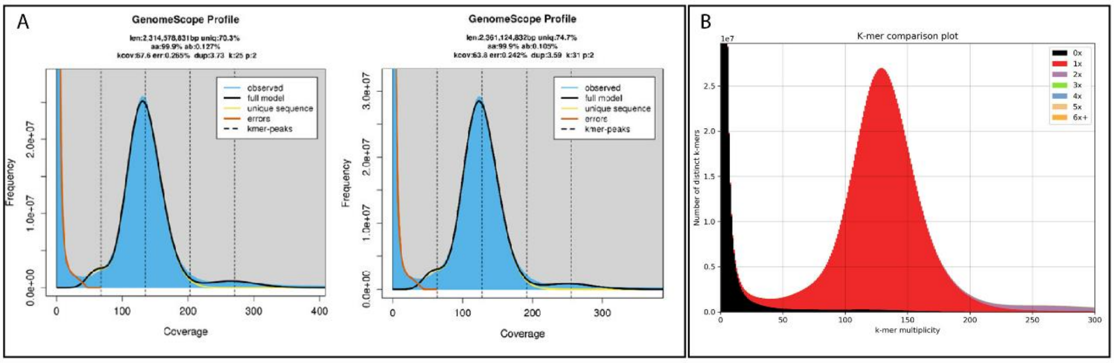


Figure P6. 2 - (A) GenomeScope2 k-mer (25 and 31) distribution displaying estimation of genome size (len), homozygosity (aa), heterozygosity (ab), mean k-mer coverage for heterozygous bases (kcov), read error rate (err), the average rate of read duplications (dup), k-mer size used on the run (k:), and ploidy (p:). (B) *Margaritifera margaritifera* genome assembly assessment using KAT comp tool to compare the Illumina PE k-mer content within the genome assembly. Different colours represent the read k-mer frequency in the assembly.

3.2. *Margaritifera margaritifera de novo* genome assembly

The Meraculous assembly and scaffolding yield a final genome size of 2.47 Gb with a contig N50 of 16,899 bp and a scaffold N50 of 288,726 bp (Table P6.1). Both N50 values are significantly higher than *V. ellipsiformis* genome assembly, i.e., 3,117 and 6,523 bp, respectively (Renaut et al., 2018). Presently, this *M. margaritifera* genome assembly reveals one the highest scaffold N50 of the Unionida genomes currently available (Renaut et al., 2018; Rogers et al., 2021). On the other hand, *M. nervosa* genome assembly contig N50, i.e. 51,552 bp, is higher than *M. margaritifera*, which is expected given the use of Oxford Nanopore ultra-long reads libraries in the assembly produced by Rogers et al., (2021) BUSCOs scores of the final assembly indicate a fairly complete genome assembly (Table P6.1) and although the contiguity is lower when compared with other recent Bivalve genome assemblies, the low percentage of fragmented genes (i.e. 5.9% for Eukaryota and 4.9% for Metazoa) gives further support to the quality of the genome assembly. Similarly, the slight difference observed between the genome size and the initial size estimation is unlikely to be a consequence of erroneous assembly duplication, as duplicated BUSCOs scores are also low (i.e. 1% for Eukaryota and 1.1% for Metazoa). The quality of the genome assembly is further supported by the high percentages of PE reads mapping back to the genome (i.e. 97.75%, Table P6.1), as well as the KAT *k*-mer distribution spectrum (Figure P6.2B), which demonstrates that almost no read information was excluded from the final assembly. Additionally, around 99% of the transcripts of the *M. margaritifera* transcriptome assembly (Bertucci et al., 2017) aligned to the genome assembly (Table P6.S4). Overall, these statistics indicate that the *M. margaritifera* draft genome assembly here presented is fairly complete, non-redundant, and useful resource for various applications.

Table P6. 1 - *Margaritifera margaritifera* genome assembly, read alignment, gene prediction, and annotation general statistics. *All statistics are based on contigs/scaffolds of size ≥1,000bp; # Euk: From a total of 303 genes of Eukaryota library profile; # Met: From a total of 978 genes of Metazoa library profile; + Euk: From a total of 255 genes of Eukaryota library profile; + Met: From a total of 954 genes of Metazoa library profile; #,+ C: Complete; S: Single; D: Duplicated; F: Fragmented; ± All statistics are based on contigs/scaffolds of size ≥2,500bp.

	Contig *	Scaffold *
Total number of Sequences (>= 1,000 bp)	265,718	105,185
Total number of Sequences (>= 10,000 bp)	66,019	15,384
Total number of Sequences (>= 25,000 bp)	18,725	11,583
Total number of Sequences (>= 50,000 bp)	4,284	9,265
Total length (>= 1,000 bp)	2,230,001,992	2,472,078,101
Total length (>= 10,000 bp)	1,523,143,239	2,293,496,118
Total length (>= 25,000 bp)	789,559,702	2,236,013,546
Total length (>= 50,000 bp)	299,796,296	2,152,307,394
N50 length (bp)	16,899	288,726
L50	34,910	2,393
Maximum length (bp)	209,744	2,510,869
GC content, %	35.42	35.42
Clean Paired-end (PE) Reads Alignment Stats		
Pecentage of Mapped PE (%)	-	97.754
Pecentage of Proper pairs PE (%)	-	90.653
Average PE sequence coverage	-	181.968
Pecentage of scaffolds with any coverage (%)	-	100.00
Total BUSCOS for the genome assembly (%)		
# Euk database	-	C:86.8% [S:85.8%, D:1.0%], F:5.9%
# Met database	-	C:84.9% [S:83.8%, D:1.1%], F:4.9%
Gene Prediction and Annotation Stats ±		
Protein coding genes (CDS)	-	35,119
Transcripts (mRNA)	-	40,544
Protein Coding genes Functional Annotated	-	26,836
Transcripts Functional Annotated	-	31,584
Total gene length (bp)	-	902,994,752
Total mRNA length (bp)	-	1,101,526,909
Total CDS length (bp)	-	52,211,391
Total exon length (bp)	-	52,211,391
Total intron length (bp)	-	1,024,450,311
Total BUSCOS for the predicted proteins (%)		
+ Euk database	-	C:90.6%[S:81.2%,D:9.4%], F:3.9%
+ Met database	-	C:92.6%[S:82.3%,D:10.3%], F:3.2%

The whole mitochondrial genome obtained with MitoBim is 16,124bp long and its gene content is the expected for Margaritiferidae female type mitogenomes (Gomes-dos-Santos et al., 2019) with 13 protein-coding genes, 22 transfer RNA, and 2 ribosomal RNA.

3.3. Repeat identification and masking and gene models prediction

The use of the custom repetitive library combined with the RepBase (Bao et al., 2015) ‘Bivalvia’ library, resulted in masking repetitive elements in more than half of the genome assembly, i.e. 59.07% (Table P6.2). Most of the annotated repetitive elements were unclassified (31.86%), followed by DNA elements (16.00%), long interspersed nuclear elements (6.13%), long terminal repeats (3.72%), and short interspersed nuclear elements (0.79%). After masking, gene prediction resulted in the identification of 35,119 protein-coding genes, with an average gene length of 25,712 bp and average CDS length of 1,287 bp (Table P6.S5). Furthermore, 26,836 genes were functionally annotated by similarity to at least one of the three databases used in the annotation (Table P6.1). The number of predicted genes is in accordance to those observed in other bivalves (and Mollusca) genome assemblies, which although highly variable, in average have around 34,949 predicted genes (calculated from Table P1.2 of Gomes-dos-Santos et al., 2021) Although the number of genes predicted within the three Unionida genomes is highly variable, i.e. 123,457 in *V. ellipsiformis*, 49,149 in *M. nervosa*, and 35,119 in *M. margaritifera*, a direct comparison should be taken with caution, given the considerable differences in genome qualities and the different gene predictions strategies applied in the three assemblies.

Table P6. 2 - Statistics of the content of repetitive elements in the *M. margaritifera* genome assembly.

		Number of elements	Length occupied (bp)	Percentage of sequence (%)
		Marmar + Bivalvia	**Marmar + Bivalvia**	**Marmar + Bivalvia**
SINEs:		108,986	17,810,092	0.79%
	ALUs	0	0	0%
	MIRs	51,807	7,321,859	0.33%
LINEs:		395,376	137,422,770	6.13%
	LINE1	7,854	2,661,360	0.12%
	LINE2	108,179	29,801,298	1.33%
	L3/CR1	13,806	3,697,570	0.17%
LTR elements:		174,445	83,417,191	3.72%
	ERVL	0	0	0%
	ERVL-MaLRs	0	0	0%
	ERV_classI	2,849	481,472	0.02%
	ERV_classII	1,072	286,047	0.01%
DNA elements:		1,208,077	358,545,022	16.00%
	hAT-Charlie	22,178	3,778,430	0.17%
	TcMar-Tigger	54,446	15,068,283	0.67%
Unclassified:		3,057,728	713,890,849	31.86%
Total interspersed repeats:			1,311,085,924	58.51%
Small RNA:		51,767	7,672,478	0.34%
Satellites:		24,005	4,250,110	0.19%
Simple repeats:		64,021	8,534,185	0.38%
Low complexity:		970	115,583	0.01%
Total masked			**1,323,560,844**	**59.07%**

3.4. Single copy orthologous phylogeny

Both Maximum-Likelihood and Bayesian Inference phylogenetic trees revealed the same topology with high support for all nodes (Figure P6.3). The phylogeny recovered the reciprocal monophyletic groups Pteriomorphia (represented by Orders Ostreida, Mytilida, Pectinida, and Arcida) and Heteroconchia (represented by Orders Unionida and Venerida). These results are in accordance with recent comprehensive bivalve phylogenetic studies (Bieler et al., 2014; V. L. Gonzalez et al., 2015; Lemer et al., 2016, 2019). The only difference is observed within Pteriomorphia, where two sister clades are present, one composed by Arcida and Pectinida and the other by Mytilida and Osteida (Figure P6.3), while accordingly to the most recent phylogenomic studies, Arcida appears basal to all other Pteriomorphia (V. L. Gonzalez et al., 2015; Lemer et al., 2016, 2019). It is noteworthy that Arcida and Pectinida clade is the less supported in the phylogeny, which together with the fact that many Pteriomorphia clades are missing in this study, should explain these discrepant results. Heteroconchia is divided into monophyletic Palaeoheterodonta and Heterodonta (here only represented by two

Euheterodonta bivalves). As expected, the two Unionida species, i.e. *M. nervosa* and the newly obtained *M. margaritifera*, are placed within Palaeoheterodonta.

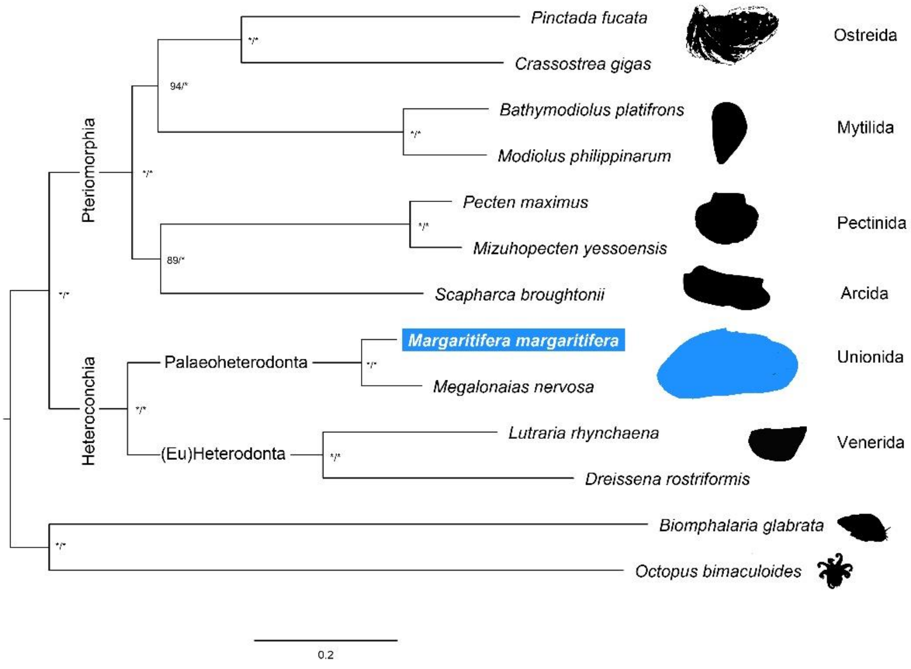


Figure P6. 3 - Maximum Likelihood phylogenetic tree based on concatenated alignments of 118 single-copy orthologous amino acid sequences retrieved by OrthoFinder. *Above the nodes refer to bootstrap and posterior probabilities support values above 99%.

3.5. Hox and ParaHox gene repertoire and phylogeny

Homeobox genes refer to a family of homeodomain-containing transcription factors with important roles in Metazoan development by specifying anterior–posterior axis and segment identity (e.g. Ferrier and Holland, 2001; Holland, 2013). Many of these genes are generally found in tight evolutionary conserved physical clusters (e.g. Castro and Holland, 2003; Pollard and Holland, 2000). Hox genes are typically arranged into tight physical clusters, showing temporal and spatial collinearity (Ferrier and Holland, 2002). Consequently, Hox genes provide useful information for understanding the emergence of morphological novelties, understanding the historical evolution of the species, infer ancestral genomic states of genes/clusters, and even study genome rearrangements, such as whole-genome duplications (e.g. Brooke et al., 1998; Ferrier and Holland, 2001; Holland, 2013). Given the disparate body plans in molluscan classes, the study of Hox cluster composition, organization and gene expression has practically become a standard in Mollusca genome assembly studies (Albertin et al., 2015; R. R. da Fonseca et al., 2020; Y. Li et al., 2017, 2020; Liu et al., 2021; Pérez-Parallé et al., 2016; Simakov et al., 2012; Sun et al., 2017, 2019, 2020; Takeuchi et al., 2016; Varney et al., 2021; S. Wang et al., 2017; Yan et al., 2019; Zhang et al., 2012). Homeobox genes are divided into four classes, of which the Antennapedia (ANTP)-class (Hox, ParaHox, NK, Mega-homeobox, SuperHox) is the best studied, particularly the Hox and ParaHox clusters (Brooke et al., 1998; Y. Li et al., 2020; Pérez-Parallé et al., 2016). The number of genes from these two clusters is relatively well conserved across Lophotrochozoa, with Hox cluster being composed of 11 genes (3 anterior, 6 central, and 2 posterior) and ParaHox cluster composed of 3 genes. Although several structural and compositional differences have been observed within Mollusca ANTP-class (e.g. Bivalvia: Zhang et al., 2012, Cephalopoda: R. R. da Fonseca et al., 2020, Gastropoda: Liu et al., 2021, and Polyplacophora: Varney et al., 2021), most Bivalvia seem to retain the gene composition expected for lophotrochozoans: Hox1, Hox2, Hox3, Hox5, Lox, Antp, Lox4, Lox2, Post2, and Post1 for the Hox cluster and Gsx, Xlox, and Cdx for the ParaHox cluster (S. Wang et al., 2017). Consequently, the identification of these genes on a bivalve genome assembly represent further validation of the genome completeness and overall correctness. Furthermore, to the best of our knowledge, this study reports for the first time the Hox and ParaHox genes were identified Unionida. A single copy of the 3 ParaHox and 10 Hox genes were found in the *M. margaritifera* genome assembly (Table P6.S6). Despite an intensive search, no evidence of the presence of Hox4 was detected. However, the gene was identified in the *M. margaritifera* transcriptome, thus confirming its presence in the species. All genes, apart from *Antp* and Lox5, were scattered in different scaffolds, with Hox5, Post1, and Gsx being present in scaffolds smaller than 2.5 kb (Table P6.S6). Both the small proximity between Antp and Lox5 and the fact that both genes are expressed in the same direction are in accordance with the results observed in other bivalves, including in the phylogenetically closest species (from which Hox cluster has been characterized), i.e. the Venerida clam *Cyclina sinensis* (Gmelin, 1791) (Y. Li et al., 2020). The fact that the remaining genes were scattered in the different scaffolds is likely a consequence of the low contiguity of the genome assembly since the distances between Bivalvia Hox genes within a cluster can be as high as 9.9 Mb (Y. Li et al., 2020). Conversely, three Hox and one ParaHox genes were found in the *M. margaritifera* transcriptome assembly and nine Hox and one ParaHox gene were found in *M. nervosa* genome assembly (Table P6.S6). Finally, to further validate the identity of the identified Hox and ParaHox genes, a phylogenetic analysis using the homeodomains (encoded 60–63 amino acid domain) of several Mollusca species was conducted (Figure P6.4). All Hox and ParaHox genes of *M. margaritifera* (as well as *M. nervosa*) were well positioned within their respective orthologous genes from other Mollusca species (Figure P6.3), thus confirming their identity.

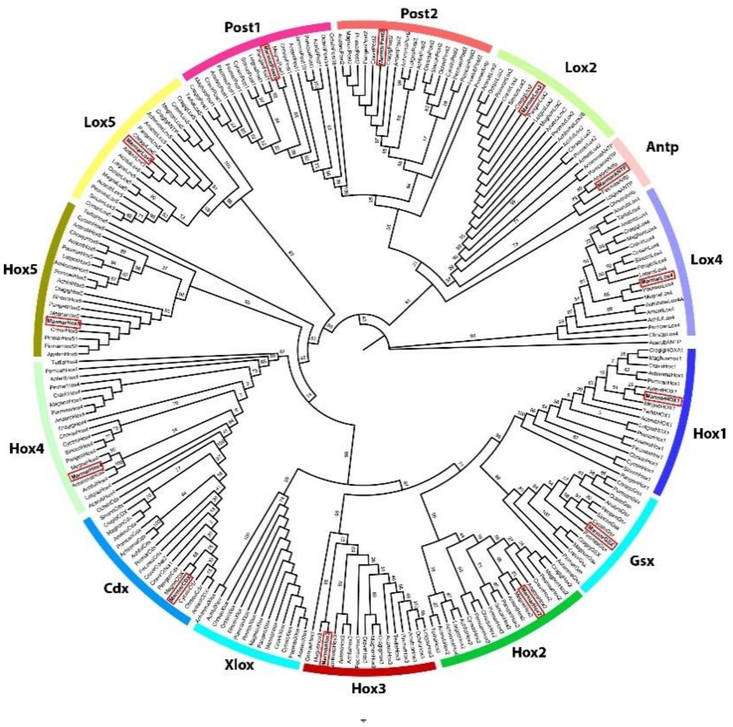


Figure P6. 4 - Hox and ParaHox Maximum Likelihood gene tree constructed using Mollusca homeodomain amino acid sequences. Bootstrap values are presented above the nodes. Red squares highlight the position of *M. margaritifera* Hox and ParaHox.

3.6. Conclusion and future perspectives

Unionida freshwater mussels are a worldwide distributed and diverse group of organisms with 6 recognized families and around 800 described species (Bogan, 2008; Graf and Cummings, 2007). These organisms play fundamental roles in ecosystems, such as water filtration, nutrient cycling, and sediment bioturbation and oxygenation (Howard and Cuffey, 2006; Vaughn, 2017), allowing to maintain and support freshwater communities (Lopes-Lima et al., 2017c). However, as a consequence of several anthropogenic threats, freshwater mussels are experiencing a global-scale decline (Böhm et al., 2021; Lopes-Lima et al., 2017c). *M.margaritifera* belongs to the most threatened of the 6 Unionida families, i.e. Margaritiferidae. Despite all this, our understanding of the genetics of this species is still to date restricted to a few mtDNA markers phylogenetic and restricted phylogeographical studies (Araujo et al., 2017; Bolotov et al., 2016; Lopes-

Lima et al., 2018a; Zanatta et al., 2018) as well as neutral genetic markers (Bouza et al., 2007; Geist and Kuehn, 2005; Zanatta et al., 2018), making the availability of the present genome a timely resource with application in multiple fields. The characterization of genetic features and identification of genomic novelties (such as single genes or gene families, genomic pathways, single-nucleotide polymorphism, among others) may provide guidance understanding molecular and cellular mechanisms of biomineralization in freshwater mussel shells that may facilitate the use of shell material as environmental and metabolic archives (Geist et al., 2005) and even help clarify the formation of new mineralized tissue following extracorporeal shock wave therapy in humans (Sternecker et al., 2018). Being the first representative genome of the family Margaritiferidae, it will help launch both basic and applied genomic-level research on the unique biological and evolutionary features characteristic of this emblematic group.

Data availability

All the raw sequencing data are available from GenBank via the accession numbers SRR13091478, SRR13091479, and SRR13091477. The assembled genomes are available in the assession number JADWMO000000000, under the BioProject PRJNA678877 and BioSample SAMN16815977 (Table P6.S7). The whole mitogenome is available in GenBank under the accession number MW556443. Fasta alignment of homeodomain amino acid sequences from Hox and ParaHox genes used in gene tree construction is available in Additional File 2. The scaffolds in which homeodomains were detected (as described in Table P6.S6) are available as Additional File 3. The repeat masked genome assembly, BRAKER2 prediction statistic and prediction gff files, as well as all predicted genes, transcripts and amino acid sequence files are available at Figshare: 10.6084/m9.figshare.13333841.

Acknowledgments

A.G.-d.-S. was funded by the Portuguese Foundation for Science and Technology (FCT) under the grants SFRH/BD/137935/2018, EF (CEECIND/00627/2017) and MLL (2020.03608.CEECIND). This research was developed under ConBiomics: the missing approach for the Conservation of freshwater Bivalves Project No. NORTE-01-0145-FEDER-030286, co-financed by COMPETE 2020, Portugal 2020 and the European

Union through the ERDF, and by FCT through national funds. Additional strategic funding was provided by FCT UIDB/04423/2020 and UIDP/04423/2020. Authors’ interaction and writing of the article was promoted and facilitated by the COST Action CA18239: CONFREMU—Conservation of freshwater mussels: a pan-European approach.

Supplementary material

Below is the link to the electronic supplementary material.

https://doi.org/10.1093/dnares/dsab002

Additional File 1 BloobTools contamination screening methods’ description and results.

Additional File 2 Fasta alignment of homeodomain amino acid sequences from Hox and ParaHox genes used in gene tree construction. Sequences used include the Hox and ParaHox homeodomains obtained in this study as well as other Mollusca homeodomain sequences retrieved from(Huan et al., 2020; Y. Li et al., 2020).

Additional File 3 Scaffolds fasta sequences in which homeodomains were detected (as described in Table P6.S6).

Table P6.S1 – List of proteomes used for BRAKER2 gene prediction pipeline.

Table P6.S2 – List of proteomes used to retrieve single-copy orthologs in OrthoFinder v2.4.0.

Table P6.S3 – *Margaritifera margaritifera* Ilumina Paire-End (PE) and Mate-Pair (MP) sequencing reads information.

Table P6.S4 – isoblat stats report of the pblat transcriptome alignment to the masked genome assembly.

Table P6.S5 – BRAKER2 gene prediction complete report.

Table P6.S6 – Genomic locations of Hox and ParaHox genes in the genome assemblies of *M. margaritifera* and *M. nervosa* and trancriptome assembly of *M. margaritifera*.

Table P6.S7 - *Margaritifera margaritifera* spcemien sampling descriptors and whole genome sequencing and assembly acessions numbers.

2.5. Manuscript 1 – The Crown Pearl V2: an improved genome assembly of the European freshwater pearl mussel *Margaritifera margaritifera* (Linnaeus, 1758)

André Gomes-dos-Santos [1,2*], Manuel Lopes-Lima [1,3,4*], André M. Machado [1], Thomas Forest [5,6,7], Guillaume Achaz [5,6], Amílcar Teixeira [8], L. Filipe C. Castro [1,2], Elsa Froufe [1*]

[1] CIIMAR/CIMAR — Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Matosinhos, Portugal; [2] Department of Biology, Faculty of Sciences, University of Porto, Porto, Portugal; [3] CIBIO/InBIO - Research Center in Biodiversity and Genetic Resources, Universidade do Porto, Vairão, Portugal; [4] IUCN SSC Mollusc Specialist Group, c/o IUCN, David Attenborough Building, Pembroke St., Cambridge, England; [5] Éco-anthropologie, Muséum National d’Histoire Naturelle, CNRS UMR 7206, Paris, France; [6] SMILE group, Center for Interdisciplinary Research in Biology (CIRB), Collège de France, CNRS UMR 7241, INSERM U 1050, Paris, France; [7] Institut de Systématique Evolution Biodiversité, CNRS MNHN SU EPHE, CP 51, 55 rue Buffon, 75005 Paris, France; [8] Centro de Investigação de Montanha (CIMO), Instituto Politécnico de Bragança, Bragança, Portugal;

* Corresponding authors.

Abstract

Producing contiguous genome assemblies for molluscs is considerably challenging owing to their large sizes, heterozygosity and widespread content of repetitive content. Consequently, the usage of long-read sequencing approaches is fundamental to achieving high contiguity and quality of the genome assemblies. The freshwater pearl mussel *Margaritifera margaritifera* (Linnaeus, 1758) (Mollusca: Bivalvia: Unionida) is one of the most culturally relevant, widespread and threatened species of freshwater mussels. The first genome assembly for this species has been produced recently, however, since the assembly relied solely on short-read approaches the genome is highly fragmented. To overcome this caveat, here, a new improved reference genome assembly is provided for the freshwater pear mussel. The new assembly is produced using a combination of PacBio CLR long reads and Illumina paired-end short reads. The genome assembly is 2.4 Gb long, possessing 1700 scaffolds with a scaffold N50 length of 3.4Mbp. The *ab initio* gene prediction resulted in a total of 48,314 protein-coding genes. This new assembly represents a substantial improvement to the previous genome and is an essential resource for studying this species’ unique biological and evolutionary features that ultimately will help to promote its conservation.

Keywords

Margaritifera margaritifera; freshwater mussel; pearls; unionida genome; whole genome

1 - Background

Initial efforts to sequence molluscs' genomes relied primarily on short-read approaches, which, despite their unarguable accomplishments, constantly result in highly fragmented genome assemblies (Gomes-dos-Santos et al., 2020; Klein et al., 2019; Takeuchi, 2017; Z. Yang et al., 2020). Consequently, long-read sequencing approaches, such as Pacific Bioscience (PacBio) or Nanopore (Oxford Nanopore), are becoming the common ground of emerging molluscan genome assembly projects (Gomes-dos-Santos et al., 2020; Klein et al., 2019; Takeuchi, 2017; Z. Yang et al., 2020). This is further facilitated by the constantly decreasing prices, coupled with increasing sequencing accuracy, of these long-read sequencing approaches (Goodwin et al., 2016). The structural information provided by long-reads is crucial to span large indels or inform about long structural variants (e.g. E. L. Koch et al., 2021; Rhie et al., 2021; Sedlazeck et al., 2018), which is particularly relevant for molluscans that have large, heterozygous and highly repetitive genomes (reviewed in Gomes-dos-Santos et al., 2020). Consequently, long-read-based assemblies have reduced levels of fragmentation, fewer levels of missing and truncated genes and reduced chances of chimerically assembled regions (Rhie et al., 2021; Sedlazeck et al., 2018).

The genome of the freshwater pearl mussel, *Margaritifera margaritifera* (Linnaeus, 1758), provided in (Gomes-dos-Santos et al., 2021) represented a key resource for the study of this highly emblematic species. The robust assembly approach resulted in a considerably complete genome assembly, which was validated with several statistics (see Gomes-dos-Santos et al., 2021 for details). However, the fact that it was assembled using solely short-read sequencing approaches (i.e., Illumina paired-end and mate-pair sequencing), resulted in a genome with hampering contiguity. The subsequent release of the highly contiguous genome assembly of the freshwater mussel *Potamilus streckersoni* (Smith, Johnson, Inoue, Doyle and Randklev, 2019), which relied on PacBio sequencing, demonstrated how long-reads are critical to significantly improve the contiguity of genome assemblies for the group (Smith, 2021).

Aiming to provide a superior genome assembly for the freshwater pearl mussel, *M. margaritifera*, here the genome of a new individual is sequenced using PacBio CLR and Illumina paired-end short reads. This new assembly represents the most contiguous freshwater mussel genome assembly available to date, representing a significant improvement in contiguity and completeness concerning the genome presented in Paper 6.

2 - Methods

2.1 - Animal sampling

One individual of *M. margaritifera* was collected from the Tuela River in Portugal (Table M1.1) and transported alive to the laboratory, where tissues were separated, flash-frozen and stored at −80 °C. The shell and tissues are deposited at CIIMAR tissue and mussels’ collection.

Table M1. 1 - MixS descriptors for the freshwater pearl mussel *Margaritifera margaritifera* specimen used for whole genome sequencing.

Sample	***Margaritifera margaritifera***
Investigation_type	Eukaryote
Lat_lon	41.862414; -6.931596
Geo_loc_name	Portugal
Collection_date	06/07/2021
Env_package	Water
Collector	Amilcar Teixeira
Sex	Undetermined
Maturity	Mature

2.2 - DNA extraction and sequencing

For PacBio sequencing, mantle tissue was sent to Brigham Young University (BYU), where high-molecular-weight DNA extraction was performed and PacBio library construction was achieved following the SMRT bell construction protocol. The library was sequenced on a single-molecule real-time (SMRT) cell of a PacBio Sequel II system v.9.0. Genomic DNA for short-read sequencing was extracted from muscle tissue using the Qiagen MagAttract HMW DNA Kit, following the manufacturer’s instructions. The extracted DNA was sent to Macrogen Inc. for standard Illumina Truseq Nano DNA library preparation and whole genome sequencing of 150 bp paired-end reads on the Illumina Novaseq6000 machine.

2.3 Genome assembly and annotation

The overall pipeline used to obtain the genome assembly and annotation is provided in Figure M1.1.

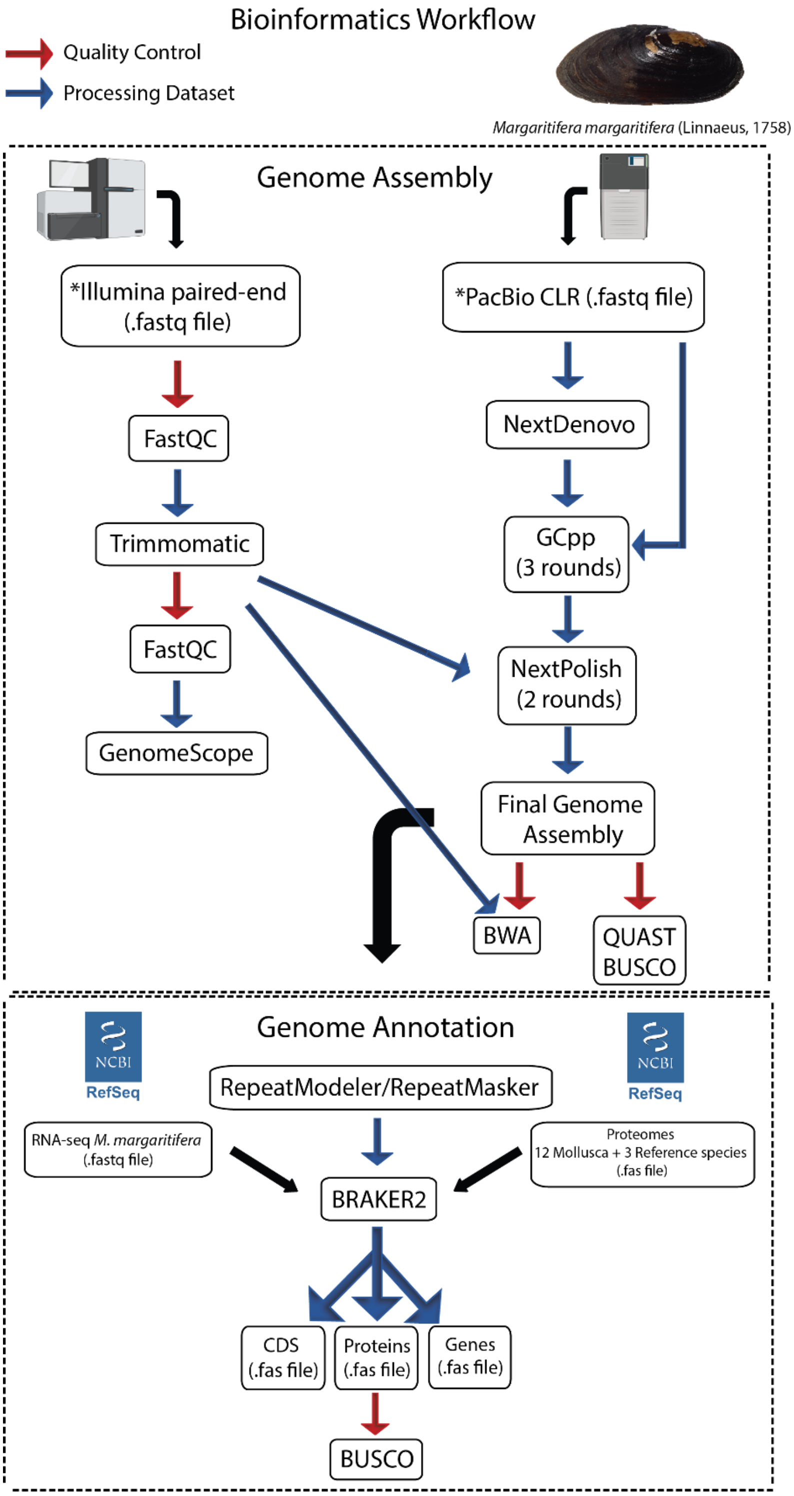


Figure M1. 1 - Bioinformatics pipeline applied for the genome assembly and annotation.

2.3.1 - Genome size and heterozygosity estimation

Prior to the assembly, the characteristics of the genome were accessed with a k-mer frequency spectrum using the paired-end reads. First, the quality of the reads was evaluated using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and the reads were after quality trimmed with Trimmomatic v.0.38 (Bolger et al., 2014), specifying the parameters “LEADING: 5 TRAILING: 5 SLIDINGWINDOW: 5:20 MINLEN: 36”. The quality of the clean reads was validated in FastQC and after used for genome size estimation with Jellyfish v.2.2. and GenomeScope2 (Ranallo-Benavidez et al., 2020) specifying the k-mer length of 21.

2.3.2 - Genome assembly

The primary genome assembly was constructed using the raw PacBio reads with NextDenovo (https://github.com/Nextomics/NextDenovo), with default parameters and specifying an estimated genome size of 2.4Gbp. Polishing of the resulting assembly was performed, first using PacBio reads, with three iterations of GCpp v 2.0.2 (Pacific Biosciences 2019), and after using the clean paired-end reads with two iterations of NextPolish v 1.2.3 (Hu et al., 2019). PacBio read alignments were performed with pbmm2 v 1.4.0 (Pacific Biosciences 2019) and paired-end read alignments were performed with Burrows-Wheeler Aligner (BWA) v.0.7.17 (Li, 2013), both with default parameters.

The general statistic and completeness of the final genome assembly were estimated with QUAST v5.0.2 (Gurevich et al., 2013), BUSCO v5.2.2 (Manni et al., 2021) and using the paired-end reads for read-back mapping, with BWA, and k-mer frequency distribution analysis with the K-mer Analysis Toolkit (Mapleson et al., 2017).

2.3.3 – Masking of repetitive elements, gene models predictions and annotation

To mask repetitive elements, first, a *de novo* library of repeats was created for final genome assembly with RepeatModeler v.2.0.133 (Smit and Hubley, 2015a). Subsequently, the genome was soft masked with RepeatMasker v.4.0.734 (Smit and Hubley, 2015b) combining the *de novo* library with the ‘Bivalvia’ libraries from Dfam_consensus-20170127 and RepBase-20181026.

Gene prediction was performed on the soft masked genome assembly using BRAKER2 pipeline v2.1.5 (Brůna et al., 2021). First, all the available RNA-seq data from *M. margaritifera* from GenBank (Bertucci et al., 2017; V. L. Gonzalez et al., 2015) and (Gomes-dos-Santos et al., 2022) (the same individual used for the genome assembly) was retrieved and quality trimmed with Trimmomatic v.0.3839 (parameters described above). Afterwards, the clean reads were aligned to the masked genome, using Hisat2 v.2.2.0 with the default parameters (Kim et al., 2015). Furthermore, the complete proteomes of 14 mollusc species and three reference species (*Homo sapiens*, *Ciona intestinalis*, *Strongylocentrotus purpuratus*), downloaded from public databases (Table M1.2), were used as additional evidence for gene prediction. The BRAKER2 pipeline was then applied, specifying parameters "–etpmode; –softmasking;". The gene predictions file (gff3) was renamed, cleaned, and filtered using AGAT v.0.8.0 (Dainat et al., 2020), correcting overlapping prediction, removing coding sequence regions (CDS) with <100 amino acid and removing incomplete gene predictions (i.e., without start and/or stop codons). Finally, proteins were extracted from the genome with AGAT and functional annotation was performed using InterProScan v.5.44.80 (Quevillon et al., 2005) and BLASTP searches against the RefSeq database (Pruitt et al., 2007). Homology searches were performed using DIAMOND v.2.0.11.149 (Buchfink et al., 2015), specifying the parameters "-k 1, -b 20, -e 1e-5, --sensitive, --outfmt 6". Finally, BUSCO scores were estimated for the predicted proteins (Manni et al., 2021).

Table M1. 2 - List of proteomes used for BRAKER2 gene prediction pipeline.

Phylum	Class	Order	Species	GenBank/
Mollusca	**Bivalves**			
		Ostreida		
			Crassostrea gigas	GCF_90280
			Crassostrea virginica	GCF_00202
		Pectinida		
			Mizuhopecten yessoensis	GCF_00045
			Pecten maximus	GCF_90265
		Veneroida		
			Dreissena polymorpha	GCA_02053
			Mercenaria mercenaria	GCF_01480
		Unionida		
			Margaritifera margaritifera	GCA_01594
			Megalonaias nervosa	GCA_01661
	Gastropod			
			Biomphalaria glabrata	GCF_00045
			Pomacea canaliculata	GCF_00307
			Gigantopelta aegis	GCF_01609
	Cephalopod			
			Octopus bimaculoides	GCF_00119
			Octopus sinensis	GCF_00634
	Polyplacophora			
			Acanthopleura granulata	GCA_01616
Chordata			*Homo sapiens*	GCF_000001
Chordata			*Ciona intestinalis*	GCF_00022
Echinodermata			*Strongylocentrotus purpuratus*	GCF_00000

3 - Results and discussion

3.1. Sequencing results and genome assembly

The raw sequencing outputs resulted in a total of 103 Gbp of raw PacBio and 203 Gbp of raw paired-end reads. A total of 201 Gbp of paired-end reads were maintained after trimming and quality filtering. Similarly, to the results of (Gomes-dos-Santos et al., 2021), GenomeScope2 estimated genome size was ~2.36 Gb and heterozygosity levels were low, i.e., ~0.163 % (Figure M1.2a).

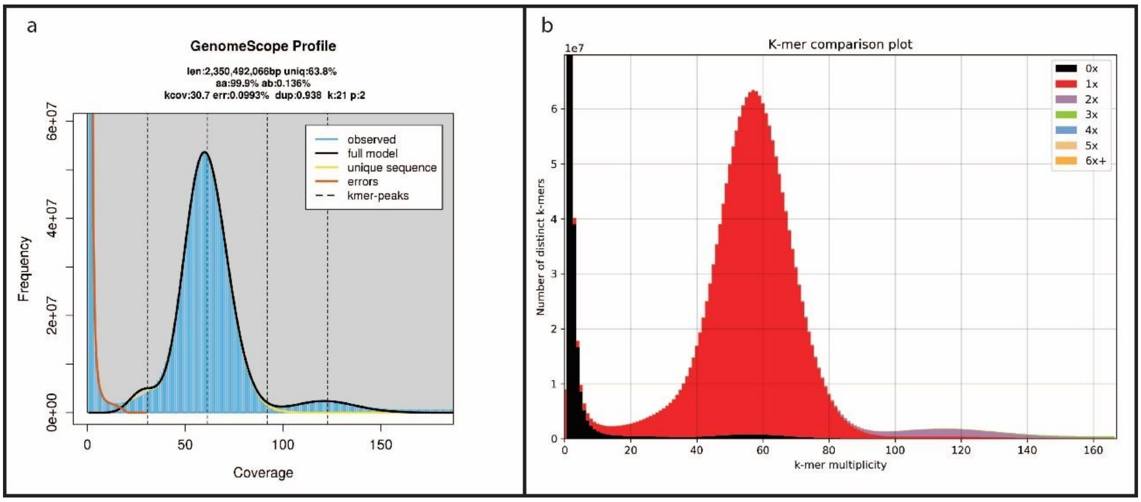


Figure M1. 2 - (a) GenomeScope2 k-mer (21) distribution displaying the estimation of genome size (len), homozygosity (aa), heterozygosity (ab), mean coverage of k-mer for heterozygous bases (kcov), read error rate (err), average rate of read duplications (dup), size of the k-mer used on the run (k:), and ploidy (p:). (b) *Margaritifera margaritifera* genome assembly assessment using KAT comp tool to compare the Illumina paired-end k-mer content within the genome assembly. Different colors represent the read k-mer frequency in the assembly.

The final genome assembly (hereafter referred to as Genome V2) has a total size of 2.45 Gbp, similar to the genome size reported in the previous genome assembly from (Gomes-dos-Santos et al., 2021) (hereafter referred to as Genome V1). Regarding the contiguity, Genome V2 shows a contig N50 of 3.42Mbp (Table M1.3), which represents a ~202-fold increase in contig N50 and ~11-fold increase in scaffold N50 relative to Genome V1 (Table M1.3). Additionally, Genome V2 represents the most contiguous freshwater mussel genome assembly available to date (January 2023) (Renaut et al., 2018; Rogers et al., 2021; Smith, 2021). In fact, Genome V2 shows a ~1.66-fold increase in N50 length regarding the other PacBio-based genome assembly, i.e., from *P. streckersoni*, (Smith, 2021), which is especially impressive considering that the Genome V2 is considerably larger (nearly 400Mbp longer), has more repetitive elements (nearly 7% more) and similar heterozygosity (nearly 0.43% less).

Table M1. 3 - General statistics of the *Margaritifera margaritifera* genome assemblies (V1 and V2) and other published freshwater mussel’s genome assemblies. * Genome V2 is at solely contig level, i.e., has no scaffolds; # Euk: From a total of 303 genes of Eukaryota library profile; # Met: From a total of 978 genes of Metazoa library profile; + Euk: From a total of 255 genes of Eukaryota library profile; + Met: From a total of 954 genes of Metazoa library profile; #,+ C: Complete; S: Single; D: Duplicated; F: Fragmented.

		Genome V2 contig*	Genome V1 contig	Genome V1 scaffold	*Megalonaias nervosa*	*Potamilus streckersoni*
Total number of Sequences (>= 1,000 bp)		1,700	265,718	105,185	90,895	2,366
Total number of Sequences (>= 10,000 bp)		1,700	66,019	15,384	54,764	2,162
Total number of Sequences (>= 25,000 bp)		1,202	18,725	11,583	29,042	1,831
Total number of Sequences (>= 50,000 bp)		1,570	4,284	9,265	12,699	1,641
Total length (>= 1,000 bp)		2,453,571,776	2,230,001,992	2,472,078,101	2,361,438,834	1,776,751,942
Total length (>= 10,000 bp)		2,453,571,776	1,523,143,239	2,293,496,118	2,193,448,794	1,775,453,721
Total length (>= 25,000 bp)		2,453,253,878	789,559,702	2,236,013,546	1,768,523,103	1,769,874,087
Total length (>= 50,000 bp)		2,448,812,075	299,796,296	2,152,307,394	1,194,323,847	1,763,052,140
N50 length (bp)		3,425,502	16,899	288,726	50,662	2,051,244
L50		207	34,910	2,393	12,463	245
Largest contig (bp)		23,800,146	209,744	2,510,869	588,638	10,787,299
GC content, %		35.3	35.42	35.42	35.82	33.79
Clean Paired-end (PE) Reads Alignment Stats						
Percentage of Mapped PE (%)	-	99.69	-	97.75	-	-
Total BUSCOS for the genome assembly (%)						
# Euk database	-	C:99.2% [S:97.6%, D:1.6%], F:0.4%	-	C: 86.8% [S: 85.8%, D:1.0%], F: 5.9%	C:70.6% [S:70.2%, D:0.4%], F:14.9%	C:98.1% [S:97.3%, D:0.8%], F:0.8%
# Met database	-	C:96.9% [S:95.5%, D:1.4%], F:2.0%	-	C: 84.9% (S: 83.8%, D: 1.1%), F: 4.9%	C:71.5% [S:70.1%, D:1.4%], F:14.5%	C:95.0% [S:93.6%, D:1.4%], F:2.3%
Masking Repetitive Regions and Gene Prediction						
Percentage masked bases (%)	-	57.32	-	59.07	25.00	51.03
Number of mRNA	-	48,314	-	40,544	49,149	41,065
Protein coding genes (CDS)	-	48,314	-	35,119	49,149	41,065
Functional annotated genes		35,649	-	31,584	-	-
Total gene length (bp)	-	1,134,996,674	-	902,994,752	-	-
Total BUSCOS for the predicted proteins (%)						
+ Euk database	-	C:97.6% [S:83.9%, D:13.7%], F:2.0%	-	C: 90.6% (S: 81.2%, D: 9.4%), F: 3.9%	-	-
+ Met database	-	C:98.7% [S:84.7%, D:14.0%], F:0.8%	-	C: 92.6% (S: 82.3%, D: 10.3%), F: 3.2%	-	-

Genome V2 also shows a considerable increase in the BUSCOs scores, with nearly no fragmented nor missing hits for both the eukaryotic and metazoan curated lists of near-universal single-copy orthologous (Table M1.3). Short-read back-mapping percentages resulted in almost complete read mapping, 99.69% alignment rate (Table M1.3), and KAT k-mer distribution spectrum revealed that almost all read information was included in the final assembly (Figure M1.2b). Overall, these general statistics validate the high completeness, low redundancy, and quality of the Genome V2.

3.2. Repeat masking, gene models prediction and annotation

RepeatModeler/RepeatMasker masked 57.32% of Genome V2 which is 1.75% less than the values of Genome V1, likely a consequence of the new assembly being able to resolve repetitive regions more accurately. Furthermore, this value was considerably higher than the estimated duplications of GenomeScope, i.e., 36.2%. These differences have been observed in other assemblies of freshwater mussel genomes (Gomes-dos-Santos et al., 2020; Renaut et al., 2018; Smith, 2021) and are likely a consequence of inaccurate estimation of repeat content when applying k-mer frequency spectrum analysis in highly repetitive genomes, using short reads. Similarly, to Genome V1, most repeats are unclassified (27.26%, ~668Mgp), followed by DNA elements (17.18%, ~421Mgp), long terminal repeats (5.95%, ~145Mgp), long interspersed nuclear elements (5.86%, ~143Mgp), and short interspersed nuclear elements (0.75%, ~18Mgp). BRAKER2 gene prediction identified 48,314 CDS, which represents an increase compared with Genome V1, but is closer to the predictions of the other two freshwater mussel assemblies (Table M1.3). This is probably a reflex of the higher contiguity and completeness of Genome V2, evidenced by high BUSCO scores for protein predictions, with almost no missing hits for either of the near-universal single-copy orthologous databases used (Table M1.2). The number of functionally annotated genes was also higher than those of Genome V1, with 4,065 additional genes annotated (Table M1.3). Overall, the numbers of both predicted and annotated genes are within the expected range for bivalves (reviewed in Gomes-dos-Santos et al., 2020), as well as within the records of other freshwater mussel assemblies (Rogers et al., 2021; Smith, 2021).

4. Conclusion

In this report, a new and highly improved genome assembly for the freshwater pearl mussel is presented. This genome assembly, produced using PacBio long-read sequencing, represents the most contiguous freshwater genome assembly available. Unlike other freshwater mussels' genomes, the one presented here has not been scaffolded (i.e., has no gaps of undetermined size), thus representing an ideal framework to employ chromosome anchoring approaches, such as Hi-C sequencing. Therefore, future efforts should aim to use this genome to produce the first freshwater mussel's chromosome-level anchored genome assembly. This improved genome represents already a key resource to start exploring the many biological, ecological, and evolutionary features of this highly threatened group of organisms, for which the availability of genomic resources still falls far behind other molluscs.

Chapter 4 – General Discussion

This thesis aimed to advance the biological study of freshwater mussels of the order Unionida (FMs), with a particular focus on the family Margaritiferidae (margaritiferids). In particular, this thesis aimed to substantially improve the plethora of genomic resources currently available for this taxonomic group of endangered molluscs. This was accomplished using distinct approaches that encompassed several genomics levels with a myriad of applications, described in the previous chapters. These resources and tools represent comprehensive frameworks with practical applications in highly relevant and emerging fields, such as mitogenomics, phylogenomics, population genomics, conservation genomics and adaptative genomics.

3.1 Phylogenomic applications

Accessing relationships among living (and extinct) organisms is perhaps one of the most essential prerequisites of most biological studies. Although important in its own right, this knowledge provides a fundamental framework to infer evolutionary transitions, such as the emergence of phenotypes, comprehend morphological evolution, infer gene origin and divergence, reconstruct demographic changes, and detect molecular adaptation (e.g., Delsuc et al., 2005; Kapli et al., 2020; Telford et al., 2015; Telford and Budd, 2003). The idea of phylogenetic reconstruction was already conveyed by Charles Darwin in The Origin of Species (Darwin, 1859). However, up to the 1970s (and the nucleic sequencing revolution), phylogenetic inferences were mostly based on morphological or ultrastructural characters, which, despite their merits, still resulted in often controversial and disputed evolutionary inferences (Delsuc et al., 2005; Kapli et al., 2020; Telford et al., 2015; Telford and Budd, 2003). The proliferation of DNA sequencing, which provides an increased number of comparable homologous characters, significatively impacted phylogenetic studies with a profound impact on our understanding of the ToL (Delsuc et al., 2005; Field et al., 1988; Fitch and Margoliash, 1967; Halanych, 2004; Telford and Budd, 2003; Woese and Fox, 1977). Although PCR and Sanger sequencing played an unarguable important role in this revolution, the limited number of genes that could be produced using these approaches (at a time and cost-effective rate) often revealed insufficient to obtain firm and statistically supported inferences (Delsuc et al., 2005; Field et al., 1988; Kapli et al., 2020; Sanderson, 2008; Telford et al., 2015). Consequently, entering the genomics era, where hundreds of base pairs from thousands of informative sites can be simultaneously produced, has represented a fundamental shift in the success and way phylogenetics is approached, coining the term phylogenomics (Delsuc et al., 2005; Eisen and Fraser, 2003; Kapli et al., 2020; Telford et al., 2015).

The radical increase in resolution provided by phylogenomic approaches, either based on targeted capturing, mitogenomic, transcriptomic or whole genome-based, has been fundamental in resolving many long-lasting contentious relationships within many groups of organisms (e.g., Combosch et al., 2017; Halanych, 2004; Hughes et al., 2018; James et al., 2020; Kocot et al., 2020; 2016c; Roberts and Kocot, 2021; Uribe et al., 2022). Despite the unarguable potential of large-scale phylogenomics (transcriptome or whole genome-based), these resources are almost inexistent for many lineages of the ToL (Stephan et al., 2022). Moreover, most taxa lack comprehensive taxon sampling, which is essential for proper phylogenetic reconstruction (Kapli et al., 2020; Telford et al., 2015). Consequently, mitogenomic and targeted capturing techniques are often favoured as they allow to increase taxon sampling, reduce sequencing costs and lower the computational burden (a shortcoming when working with whole-genome datasets). However, proper data curation and efficient bioinformatic pipelines for processing genomics outputs are critical in phylogenomic studies. Although a wide coverage of genomic and taxon is important, accurate data assessment and processing is also imperative (Buddenhagen et al., 2016; Lemmon and Lemmon, 2013). Accurate and unbiased phylogenetic reconstruction is a multi-dependent task resulting from an adequate bioinformatics workflow. This requires a set of decisions regarding evolutionary modelling, orthology assessment, and matrices reconstruction to properly balance information from several *loci* with dissimilar underlying evolutionary histories (Bernt et al., 2013a; Buddenhagen et al., 2016; Chen et al., 2007; Edwards et al., 2016; Hosner et al., 2016; Lemmon and Lemmon, 2013; Zhang et al., 2018). Consequently, new approaches for data filtering, assessment, and selection before phylogenetic reconstruction are also needed. In the end, phylogenomic studies should find a balance between having the highest number of phylogenetic informative markers, while ensuring a wide taxon representation and proper data assessment methods, with the sequencing cost generally playing an important role in the decision.

In this thesis we provide a series of novel genomic resources, encompassing several scales of genomics approaches, including mitogenomics (Papers 3, 4), target capturing (Paper 4), transcriptomics (Paper 5) and whole genome (Paper 6 and Manuscript 1). These resources have an enormous potential for phylogenomics reconstruction, not only for margaritiferids (demonstrated in Papers 3-5) but also for FMs and Mollusca.

3.1.1 Phylogenomics in Mollusca

Phylogenomics approaches relying on hundreds of genes have been fundamental in retrieving monophyletic molluscan classes (revised in Papers 1, 2). The results of many recent Mollusca phylogenomic studies are starting to provide a generalized consensus about long contentious inferred relationships within its main lineages (Kocot et al., 2020, 2011; Smith et al., 2011). However, conclusive results are often hampered not only by a generalized lack of genomics resources for the phylum but also by the biased nature of the available resources, which favours the most popular groups of the phylum, that is, gastropods, bivalves and cephalopods (revised in Paper 1). This is even more difficult by the fact that, unlike in many other taxa, phylogenetics based on mitogenomes (the most abundant genomic resource for Mollusca, Paper 1) has constantly failed to infer deeper evolutionary relationships within Mollusca (revised in Paper 2). Consequently, novel genomics datasets, especially whole genome and transcriptomes, are fundamental to providing the taxon representative frameworks to definitively resolve inferences of deeper relationships within the phylum (revised in Papers 1, 2). In Paper 6 the first whole genome assembly for the freshwater pearl mussel is sequenced and its potential for deep phylogenomic application is demonstrated by constructing a simplified Bivalvia phylogeny using a standardized genome single copy orthologue search approach (Emms and Kelly, 2019). Furthermore, a second improved assembly for the species is provided in Manuscript 1 and the transcriptomes of five European FMs species are provided in Paper 5. These resources either solely or combined, have an equally important potential for phylogenomic studies, as previously demonstrated in Mollusca (e.g., V. L. Gonzalez et al., 2015; Kocot et al., 2020; Lemer et al., 2019; Uribe et al., 2022).

3.1.2 Phylogenomics in Freshwater Mussels

When aiming for taxa-specific analysis, where the scarcity of large-scale genomic resources can be even more accentuated (reviewed in Paper 1) the use of alternative phylogenetic approaches, such as mitogenomics or target capturing methods are often the available options. Mitochondrial genomes (or genes) alone or coupled with a few selected nuclear genes have been extremely useful in resolving intrafamilial phylogenies in Mollusca (Paper 1) with particular importance in FMs phylogenetics. Inferring evolutionary relationships within Unionida is particularly challenging, in part due to their high morphological plasticity, with molecular phylogenetics playing a revolutionizing role in phylogenetic studies and systematics for the group (e.g., Araujo et al., 2018; Combosch et al., 2017; Froufe et al., 2019; X. C. Huang et al., 2019; Lopes-Lima et al., 2018a; Lopes-Lima et al., 2017a; Whelan et al., 2011; R. W. Wu et al., 2019). Although revolutionizing, these strategies have not always been sufficient to retrieve unambiguously coherent phylogenies, especially for studying ancient and suprageneric relationships, due to the reduced resolution of limited character sampling and/or the several known caveats of mtDNA (e.g., Combosch et al., 2017; Pfeiffer et al., 2019; Sano et al., 2022). Consequently, the development of the first target capture approach for FMs, i.e., the AHE Unioverse probe dataset (Pfeiffer et al., 2019), represented a promising tool for the phylogenetics and systematics of the group. The utility of this AHE dataset has been demonstrated at distinct evolutionary scales, including at the family level (Pfeiffer et al., 2021, 2019; Smith et al., 2020) (Figure D.1a). However, at a lower taxonomic level, the probes have only been tested for one of the six families of the order Unionida, i.e., family Unionidae (Pfeiffer et al., 2021, 2019; Smith et al., 2020) (Figure D.1a).

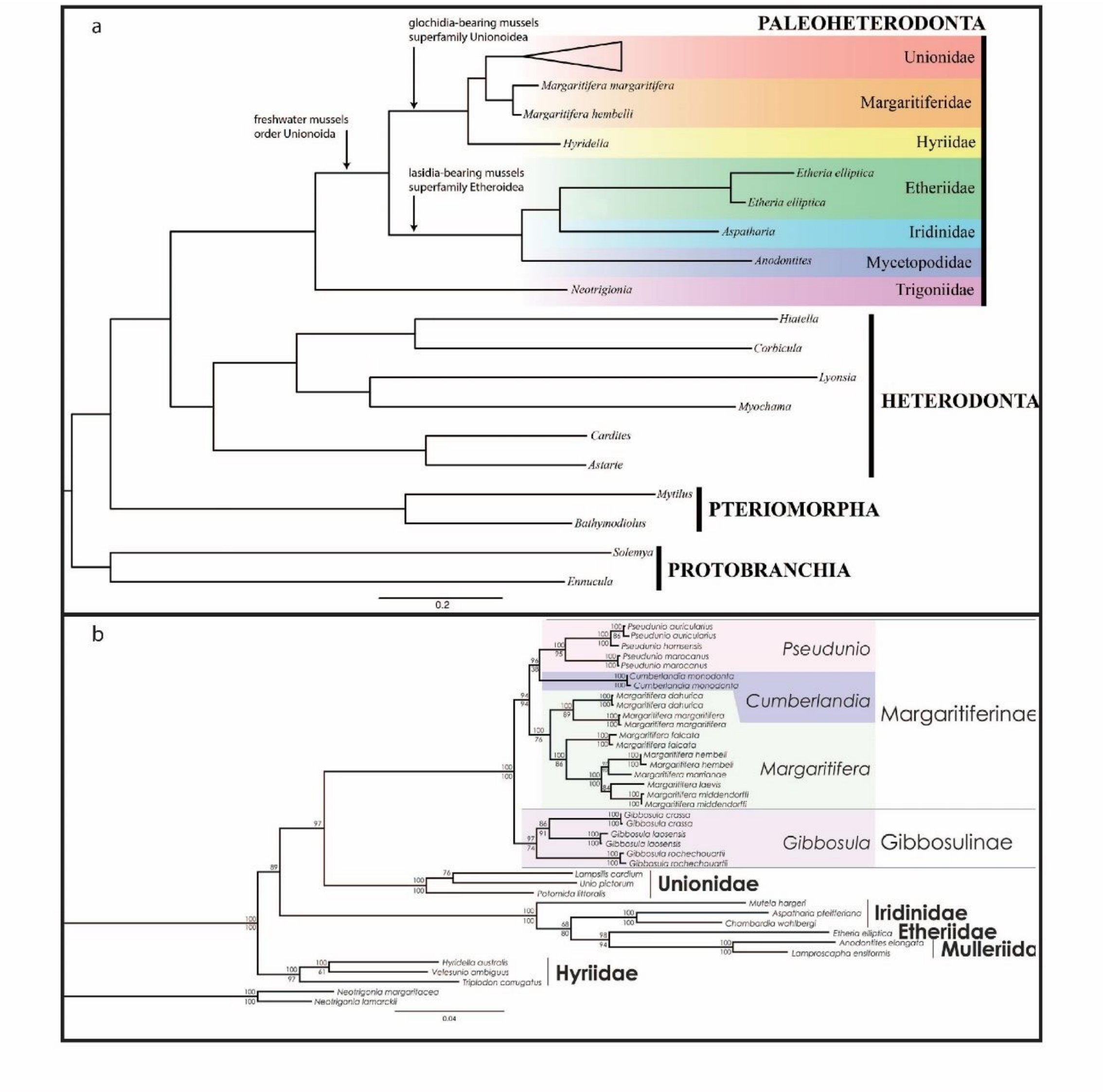


Figure D. 1 - a) Maximum Likelihood reconstruction from a nucleotide supermatrix of 569 AHE loci from Bivalvia. Colour blocks highlight the six different Unionida families, as well as the marine bivalve family Trigoniidae. The clade including all representatives of the family Unionidae is collapsed for clarity. Node support values are not displayed as they were 100 for both bootstrap in Maximum Likelihood and posterior probabilities in Bayesian Inferences analyses. Figure adapted from Pfeiffer et al., 2019; b) Maximum Likelihood reconstruction from concatenated nucleotide alignment of genes COI [3 codons], 16S, 18S, 28S and H3 [3 codons] from Paleoheterodonta. Support values above the branches are posterior probabilities and below the branches are bootstrap supports. Figure adapted from Lopes-Lima et al., 2018a.

Currently, two subfamilies, i.e., Margaritiferinae and Gibbosulinae, and four genera are recognised for margaritiferids, i.e., *Margaritifera* (seven species), *Pseudunio* (three species), *Cumberlandia* (one species), and *Gibbosula* (five species) (Lopes-Lima et al., 2018a) (Figure D.1b). However, the only AHE-based phylogenomics tree reconstruction that contains margaritiferids has only included two species, i.e., *M. margaritifera* and *M.*

hembelli (Pfeiffer et al., 2019) (Figure D.1a). In Paper 4 the first family-level phylogenomic reconstruction for margaritiferids (encompassing 14 out of the 16 recognized species) is produced using a set of AHE loci. These results are further complemented with an equally representative mitophylogenomic reconstruction using several newly sequenced margaritiferids' whole mitogenomes (Papers 3 and 4). The results of Papers 3 and 4 revealed that although both the AHE and the mitogenome phylogenies agreed with the recently established family systematics (Araujo et al., 2017; Huff et al., 2004; Lopes-Lima et al., 2018a), they disagreed concerning the relationships within some genus-level groups.

Mitochondrial genomes have many intrinsic features that make them reliable for phylogenetic inferences (revised Paper 2), however, not always reflect the evolutionary history of the species (Ghiselli et al., 2021; Hurst and Jiggins, 2005; Kern et al., 2020). The consistent, well-supported but disagreeing results between nuclear and mtDNA approaches here presented, suggest a divergent evolutionary history of nuclear and mitochondrial markers. Moreover, given the notably low mitochondrial evolutionary rates observed within margaritiferids (Bolotov et al., 2016), it is unlikely that the results are a reflex of nucleotide substitution saturation. Recently, a similar pattern of mito-nuclear phylogenetic disagreement has been linked to the non-random fixation of mitochondrial haplotypes as a response to environmental variability in gentoo penguins (e.g., Noll et al., 2022). Therefore, this pattern might be an indication of a putative mitogenome selective constraint, thus should be carefully explored in further studies.

On the other hand, mitogenomes of FMs show a myriad of evolutionary novelties that differentiate them from the generally expected patterns of mitochondrial DNA, particularly DUI (Paper 2). Interestingly, the M-type phylogeny provided in Paper 3 seems to support the relationships inferred using the AHE dataset, rather than the F-type mitochondrial phylogeny. Unfortunately, only three M-type mitogenomes have been provided for margaritiferids so far which hinders a deeper interpretation of the results. Increasing the number of M-type mitogenomes for the family is fundamental, not only to provide an additional tool for phylogenomic inference but also to study the many putative functions of this unique yet mysterious pattern of mitochondrial inheritance (further explored in section 3.2 below).

3.1.3 A new pipeline and AHE loci annotation

With the decreasing costs of WGS and given its wide range of applications, it is likely that in a near future, the sequencing cost match those of target capture approaches (Hotaling et al., 2021). However, the underlying design of target-capturing approaches, such as AHE, has already carefully pre-selected several hundred loci with a well-demonstrated potential for phylogenomics (Lemmon and Lemmon, 2013). Therefore, new pipelines that can integrate WGS with target-capturing outputs are now emerging, as they represent timely needed tools (e.g., J. M. Allen et al., 2017; Faircloth, 2016; Knyshov et al., 2021). In Paper 4 a new assembly pipeline that incorporates WGS outputs in target capturing phylogenomics is described, and its potential is demonstrated using the WGS outputs from Paper 6 and the AHE dataset from Paper 4. Although here only tested for margaritiferids, the pipeline was designed without taxa-specific parameters and thus can be easily applied to other taxa. Furthermore, the functional characterization of the target exonic regions of the Unioverse AHE dataset is also generated in Paper 4. This functional characterization will provide a framework to guide data processing, test robustness and identify biases within the data, which will further improve phylogenetic reconstruction, as well as widen the applications of this dataset.

3.2 Mitogenomic of Freshwater Mussels

3.2.1 Doubly Uniparental Inheritance (DUI)

For many metazoan species, sequencing the whole mitogenomes represents the initial frontier to enter the genomic era. This has been largely leveraged by the high availability of cost-effective NGS approaches and bioinformatic tools for whole mitogenome assembly (reviewed in Paper 1). The general goal of many of these mitogenome sequencing projects is often to provide a new robust tool for phylogenetic inferences (reviewed in Paper 1). This arises from the fact that most Metazoa mitogenomes are expected to follow a ‘textbook’ description, both in composition, stability, and inheritance (reviewed in Papers 1 and 2) (Bernt et al., 2013a; Gissi et al., 2008). However, as new mitogenomes are being sequenced, more exceptions to this “textbook” description are being found (Bernt et al., 2013a; Kolesnikov, 2016), with phylum Mollusca representing a “hotspot” of mitogenomic novelties. The current knowledge of all the novelties, applications and challenges of Mollusca mitogenomics is thoroughly reviewed in Paper 2 of this thesis.

Perhaps the most remarkable of all the Molluscan mitogenome characteristics is Doubly Uniparental Inheritance (DUI) of the mitogenome, a pattern found in over 100 Bivalvia species, that contradicts the generally assumed strictly maternal inheritance (Gusman et al., 2016). Since first described, DUI has been the focus of several studies aiming to understand its origin, how is it differentially segregated and what is, if any, its functional relevancy (e.g., Breton et al., 2018; Guerra et al., 2019; Gusman et al., 2016; Zouros, 2013) (reviewed in Paper 2). This phenomenon has been reported in different bivalve lineages, including species from orders Cardiida, Mytilida, Nuculanida, Venerida, and Unionida (Gusman et al., 2016; Smith et al., 2022). However, only within the latter sequences of the F-type lineage (observed in all families) and M-type lineage (reported in ~80 gonochoric species of families Unionidae, Margaritiferidae and Hyriidae) are reciprocally monophyletic (up to ~50% nucleotide divergence from each other) (Guerra et al., 2017, 2019; Gusman et al., 2016). This pattern suggests that DUI was present in the common ancestor of all FMs, therefore the two lineages have evolved separately and without lineage ‘recombination’ (often observed in other bivalves carrying DUI) (Guerra et al., 2017, 2019; Gusman et al., 2016; Smith et al., 2022; Stewart et al., 2009). The well-defined presence of DUI across some groups of FMs, combined with various independent transitions from gonochoric to hermaphroditic as well as the tight linkage of these transitions with lineage-specific ORFans (genes with unknown homology or function), makes the group a great model to study DUI (Breton et al., 2011b; Chase et al., 2018; Guerra et al., 2017, 2019; Gusman et al., 2016; Stewart et al., 2009). Furthermore, a series of family-specific conserved mitogenomes features that help to trace ancestral mitogenome states have been identified within the families of DUI-bearing FMs, including at least four F-type and three M-type conserved gene orders (Froufe et al., 2019; Lopes-Lima et al., 2017a) (Figure D.2a); duplicated and elongated genes (Breton et al., 2009; Chase et al., 2018; Guerra et al., 2017, 2019) (Figure D.2b) and; conserved intergenic regions (Breton et al., 2009; Chase et al., 2018; Guerra et al., 2017, 2019).

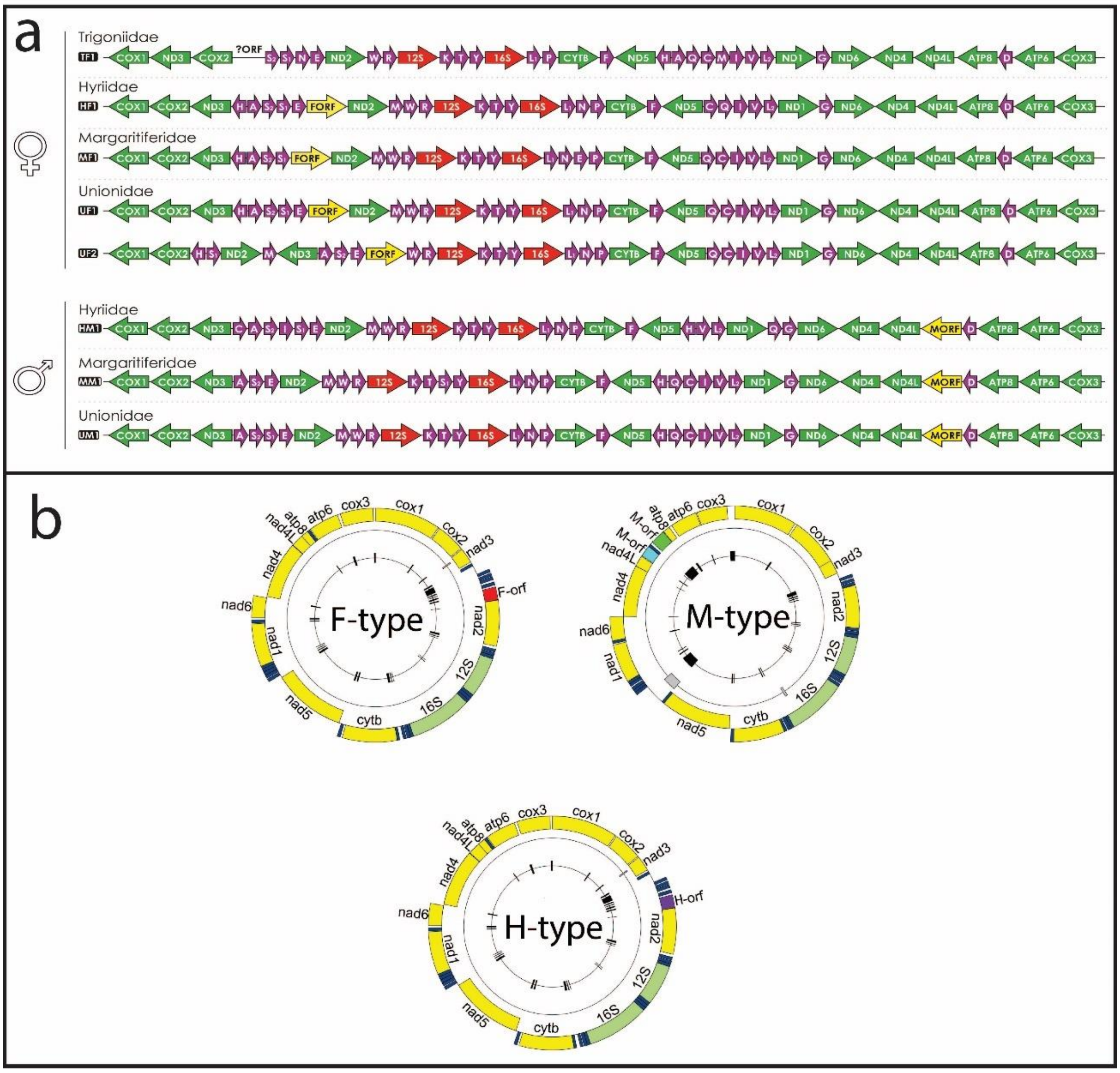


Figure D. 2 - a) Linear representation of all the currently known mitochondrial gene orders for FMs and Trigoniidae family; b) Circular representation of the typical F, M and H-type mitogenomes of margaritiferids. Adapted from Guerra et al., 2019

Among the FMs, margaritiferids possess many peculiarities that made them particularly important models to study DUI, such as: possessing different sexual strategies, including strictly gonochoric or hermaphroditic species, or gonochoric species with occasional hermaphrodism; they also have unique mitogenomic characteristics, including two copies of the M-*orf* and two unique mitogenomic gene orders (Figure D.2) (Breton et al., 2011b; Guerra et al., 2017, 2019). Despite this, before the work presented in this thesis, only two M-type and six F-type mitogenomes were available for the group. Here, several new mitogenome assemblies for the family are generated and presented in Papers 3 and 4. Firstly, the whole mitogenomes of male, female, and hermaphroditic specimens

\of the freshwater pearl mussel *M. margaritifera* are sequenced (Paper 3). This species is the only FM with a transatlantic occurrence and has the particularity that North American individuals seem to be mostly gonochoric (with only a couple records of hermaphroditic individuals), while Iberian individuals show high abundances of hermaphrodites (Breton et al., 2011b; der Schalie, 1970; Grande, 2001). Iberian individuals have only been reported to be either female or hermaphrodites, but never male (Breton et al., 2011b; der Schalie, 1970; Grande, 2001). This interesting pattern seems unique to Iberian individuals, as in other European populations hermaphroditism also seems to occur very rarely (Breton et al., 2011b; der Schalie, 1970; Grande, 2001). Paper 3 shows that similarly to other hermaphroditic species, the M-type lineage is absent in the European individuals of *M. margaritifera.* However, unlike the strictly hermaphroditic congeneric species *Margaritifera falcata* (Gould, 1850), the F-type lineage is still unaltered, i.e., it retains the F-*orf*, rather than the derived H-*orf* (Paper 3) (Figure D.2b). Previous studies on other strict hermaphroditic FMs have shown that although H-*orf* is derived from F-*orf* it has a highly divergent nucleotide sequence, showing repeat motifs insertions (and increased length), suggesting a relaxation of selective pressures (Breton et al., 2011b; Chase et al., 2018; Stewart et al., 2013). These results suggest that only after fully transitioning to hermaphroditism, the gender-specific ORFan starts to diverge, bringing new insights into the putative implication of DUI and the ORFans in sexual determination (Paper 3). Similarly, the other margaritiferid species from the Iberian Peninsula, i.e., *P. auricularius*, seem to share the pattern of occasional hermaphroditism (Grande, 2001). However, the only available mitogenome (F-type) for the species is from a French individual and no sexual characterization has been provided for the sample (Guerra et al., 2019). Furthermore, this pattern is also observed within other DUI-bearing non-margaritiferid FMs, such as *Anodonta anatina*. In this species, different populations have different percentages of individuals with different sexual strategies, including strictly gonochoric populations, populations with either low or high numbers of hermaphrodites, or even strictly hermaphroditic populations (Hinzmann et al., 2013). However, once again, no mitogenome has been sequenced from hermaphroditic individuals. Consequently, further studies should focus on these occasional hermaphroditic species to further explore the absence of the H-*orf*.

On the other hand, the combined results from Papers 3 and 4 provide a series of new F-type whole mitogenomes for the Margaritiferidae family, i.e., *Margaritifera marrianae* Johnson, 1983, *Margaritifera hembeli* (Conrad, 1838), *Margaritifera middendorffi*

(Rosen, 1926), *Margaritifera laevis* (Haas,1910), *Margaritifera falcata* (Gould, 1850), *Pseudunio auricularius* (Spengler, 1793), *Pseudunio homsensis* (Lea, 1864), *Gibbosula laosensis* (Lea, 1863) and *Cumberlandia monodonta* (Say, 1829), increasing the available mitogenomes to 14 out of the 16 currently recognized species of the family. Furthermore, we provide the first M-type mitogenome for *M. margaritifera* in Paper 3. All the new and previously sequenced mitogenomes (n=19) possess the same F-type unique gene order, i.e., MF1 (Lopes-Lima et al., 2017a) (Figure D.2a). Although two species of the genus *Gibbosula* still lack a complete mitogenome, the three already available share this gene order, thus supporting the use of this characteristic as a diagnostic character for the entire family (Lopes-Lima et al., 2017a). The margaritiferid M-type mitogenome (n=3) also shares a unique gene arrangement, i.e., MM1 (Guerra et al., 2017; Lopes-Lima et al., 2017a) (Figure D.2a). Apart from the M-type mitogenomes produced in Paper 3 for *M. margaritifera*, two more have been produced for *C. monodonta* and *P. marocanus* (Guerra et al., 2017; Lopes-Lima et al., 2017a). Previous studies have identified M-type lineage in other species of *Margaritifera and Pseudunio*, although scattered within each genus and (as mentioned above) strongly correlated with the species' reproductive strategy (Breton et al., 2011b; Curole and Kocher, 2005; Guerra et al., 2019; Gusman et al., 2016; Walker et al., 2006). Conversely, the occurrence of the M-type lineage within the genus *Gibbosula* has not been accessed yet. Apart from the distribution (Lopes-Lima et al., 2018a), little is known about the five species recognized within the genus. This is the most morphologically distinct lineage and unarguably the least studied group of margaritiferids, to the extent that until recently some species were assigned to Unionidae instead of Margaritiferidae (X. C. Huang et al., 2018; Lopes-Lima et al., 2018a). Furthermore, no host-fish or sexual reproductive mechanisms have been reported for any species of the genus *Gibbosula* (Lopes-Lima et al., 2018a). Consequently, future studies should aim to infer the presence of the M-type in all *Gibbosula* species and link it to their sexual strategies, which, given their conserved and distinct position in all phylogenies of the family, will provide fundamental insights into the evolution of DUI.

3.2.2 Application of FMs mitogenomes to study adaption

Whole mitogenomes also represent important tools for studying cellular energetic adaptation. Metazoans’ mitochondria by ensuring the production of nearly 95% of the eukaryotic cell energy through oxidative phosphorylation (OXPHOS), play a vital role in cellular bioenergetics (Breton et al., 2014; Letts et al., 2016). This process is controlled by five protein complexes dependent on ~93 genes, of which 13 are encoded in the mitogenome (involved in complexes I, III, IV, and V) and the remaining encoded in the nuclear genome (involved in complexes I, II, III, IV, and V) (Letts et al., 2016; Nicholls and Ferguson, 2002). Given the transverse importance of OXPHOS for cellular maintenance and survival, the proteins involved in this pathway are under high functional constraints (R. R. da Fonseca et al., 2008). Consequently, changes that affect the efficiency of the OXPHOS complex provide strong evidence of key mitochondrial adaptations that promote organism survival in diverse environments (Bennett et al., 2022; Sebastian et al., 2020). In fact, mutations in OXPHOS genes have been linked with a plethora of environmental adaptations, such as hypoxia, high altitude, extreme thermal conditions (cold and warm), dietary availability, realm change (adaptation to living in land) and toxicity, as well as with conditions of high metabolic demand (e.g., Almeida et al., 2015; Breton et al., 2014; Chapdelaine et al., 2020; Fourdrilis et al., 2018; Hraoui et al., 2020; Y. Li et al., 2013; Luo et al., 2013; Noll et al., 2022; Pfenninger et al., 2014; Romero et al., 2016; Sebastian et al., 2020; Toews et al., 2014; Yuan et al., 2020). Moreover, mitochondrial genomes also play an important role in controlling cell ageing and senescence (Lauri et al., 2014). Many species of margaritiferids have adapted to inhabit highly demanding environments, such as *M. margaritifera* which inhabits oligotrophic streams, from temperate Atlantic rivers to cold Arctic rivers; *P. marocanus* and *P. homsensis* found in warmer Mediterranean rivers; and, *M. falcata* that has been found to inhabit mid-level altitude water streams (~2.500m) (Blevins et al., 2016). Margaritiferids are also among the longest-lived invertebrate species, with *M. margaritifera* living over 200 years (Kaliuzhin et al., 2009) and showing very weak signs of senescence (referred to as ‘negligible senescence’) (Hassall et al., 2017). Consequently, margaritiferids represent ideal models to study mitochondrial adaptation. Although the mitogenomes produced in this thesis were only explored for phylogenomics and patterns of DUI evolution, they also represent fundamental tools to explore their putative role of the mitochondria in many of the margaritiferids' interesting biological adaptations.

Finally, given the intimate association of mitochondrial and nuclear genomes in the OXPHOS, a coevolution process is expected. Disruption of this interaction may affect the entire process, with consequences on the organismal fitness. Consequently, studying the coevolutive interaction may provide several important insights into the mitochondrial adaptative process (Biot-Pelletier et al., 2022; Blier et al., 2001; Chapdelaine et al., 2020; Deremiens et al., 2015). The *M. margaritifera* mitogenomes produced in Paper 3 and the whole genome assemblies produced in Paper 6 and Manuscript 1, represent important resources to explore mito-nuclear coevolutionary patterns and their putative implications on the species' biological singularities.

3.3 From Genomes to Phenotypes

3.3.1 Genomes as the first goal of the genomic era

Perhaps the most critical step towards establishing a non-model organism in the genome era is generating the full nuclear and mitochondrial DNA sequence assemblies. The genome encodes the instructions that govern an organism’s appearance, behaviour, and physiology, thus harbouring all the information to access its evolutionary history, describe its biological novelties and estimate its adaptive success (Dunn and Ryan, 2015; Stephan et al., 2022). This highlights the tremendous potential that motivates genome assembly projects while also highlighting that generating a genome *per se* is not the primary motivation for such endeavours. The ultimate goal is to identify and classify genomic features and link them with phenotypic diversity and the historical processes that define an organism (Dunn and Ryan, 2015; Oppen and Coleman, 2022; Stephan et al., 2022). Dunn and Ryan (2015) elegantly described four main goals in studies of animal evolutionary genomics: 1) reconstruct the genome evolutionary history; 2) identify the genomic variations underlying historical phenotypic changes; 3) understand the evolutionary processes that underlie genome changes and; 4) use genome evolution as a proxy to reconstruct other historical patterns. Despite these well-defined goals, the fact is that producing a single genome, by itself, is generally insufficient to answer the aimed biological questions (Richards, 2015; Stephan et al., 2022). First and most importantly, because inter and intra species comparison is a fundamental approach in biological studies, thus comprehensive genome sampling within and among species is fundamental (Dunn and Ryan, 2015; Richards, 2015; Stephan et al., 2022). Secondly, although genomes are “the source code of life”, accurately deciphering the code is a complex task, often depending on alternative approaches, such as transcriptomics, epigenomics, proteomics, association mapping and genome scans that provide clues and define functional regions (Campagna and Toews, 2022; Dunn and Ryan, 2015; Lopez et al., 2019; Stephan et al., 2022). These functional informative approaches are fundamental to unravel the genomic features underlying the many complex aspects of organism biology, especially in non-model species, for which comparable data from closely related species are often limited (Lopez et al., 2019; Stephan et al., 2022). Genome assemblies are rarely provided alone, most often accompanied by a set of RNA-seq data (often from the same individual), which helps to verify the completeness of assemblies, provide structural evidence for accurate gene prediction, and provide tissue-specific gene expression profiles (Dunn and Ryan, 2015; Lopez et al., 2019). Finally, a single-individual genome assembly should be cautiously interpreted as the reference genome of an entire species, as intraspecies (and even intraindividual diversity) won’t be represented within a single assembly (Dunn and Ryan, 2015). Consequently, the concept of producing pangenomes, i.e., genomes encompassing all the genomic regions and variations shared between the individuals of a particular taxon, will likely be a common output of metazoans genome studies, although still extremely rare today (e.g., Calcino et al., 2021; Gerdol et al., 2020).

In the end, although the genome assembly itself is not enough to fully accomplish the four goals of the studies of animal genomics, is still the most fundamental tool towards achieving each goal. The genome by sustaining the “code of life” is the framework in which the characterization of genomic variation will be performed. Therefore, the genome is the “foundation stone”, ultimately defining the entering of the study of species in the genomic era.

3.3.2 Molluscs genome assemblies

Considering the importance of producing genome assemblies for the study of non-model organisms, in Paper 1 of this thesis two sections were dedicated to reviewing the availability, characteristics and challenges of molluscan genome assemblies. The review also provided an overview of the whole genome availability for metazoans, showing the disproportionality of genome sequenced for molluscs. At the time of publication (i.e., data relative to 2019), only 0.04% (n=33) of the molluscan species had their genome sequenced and only three classes were represented, i.e., cephalopods, gastropods, and bivalves. In the publication, an optimistic trend of increasing genome assemblies for molluscs was highlighted, which continued in the following years. The number of molluscan genomes available in 2022 (accessed on 09-Nov-2022 https://www.ncbi.nlm.nih.gov/data-hub/genome/), revealed that a total of 119 new genomes have been generated since Paper 1, now including the first Polyplacophora genome, from *Acanthopleura granulate* (Varney et al., 2021), and the first Solenogastres genome, from *Wirenia argentea* (ASM2580221v1). Although positive, these values are a small fraction of the total known molluscan species highlighting the importance of keeping (and even increasing) this trend through the generation of new assemblies for unsampled taxa.

Another important factor explored in Paper 1, as well as in other recent publications (Klein et al., 2019; McCartney, Mallez, et al., 2019; Takeuchi, 2017; Z. Z. Yang et al., 2020), is the challenging task of molluscan genome assembly projects. In general, molluscan genomes seem to share (with a degree of variation) three characteristics that are known to strongly affect the success of genome assembly software, that is, the large genome size (up to 5.4 Gbp), the high composition of repetitive elements (up to ~70%), and the elevated rate of heterozygosity (up to 3.7%). Initial genome sequencing efforts relied mainly on short-read approaches. Despite their unarguable merits, short-read methods have well-known caveats, particularly when used to assemble large and complex genomes for which they offer little resolution to span long indels or structural variants (Rhie et al., 2021; Sedlazeck et al., 2018). Therefore, when relying solely on short-read sequencing, assemblies can show high levels of fragmentation, with a high rate of missing, truncated, or incorrectly assembled genes. This is evidenced by the general genome statistics of the short-read-based molluscan genome assemblies (reviewed in Paper 1). To overcome this shortcoming, molluscan genome assembly projects started adopting long-read sequencing approaches, such as PacBio or Nanopore long-reads, which are now the dominating methods in molluscan genome projects and have significantly improved the quality of these genomes (Paper 1, McCartney et al., 2019; Takeuchi, 2017; Z. Yang et al., 2020). Besides the generalized implementation of long-read sequencing approaches, many projects are now taking a step further and producing chromosome-level genomes using chromatin crosslinking protocols (Paper 1). In fact, since Paper 1 was released, a total of 36 new chromosome-level mollusc assemblies have been produced (accessed on 09 Nov-2022 https://www.ncbi.nlm.nih.gov/data-hub/genome/).

The first mollusc genome assembly was produced 13 years ago, marking the entry of the phylum in the genome era (Paper 1). Over this period, genome assembly projects have accompanied an astonishing revolution in the fields of DNA sequencing and bioinformatics, resulting in continuously improved genome assemblies. Genomes are works in progress which are continuously improved upon (McCartney et al., 2019). This is transversal to every ToL and is illustrated by the genome that propelled the genomics era in the first place, the human genome, which over the last two decades has been continuously improved over 14 patched releases (announcement of significant improvements in the genome), culminating in the gapless genome assembly presented in 2022 (Nurk et al., 2022). In the end, genome projects aimed to produce the most accurate and representative genome assembly, the success of which depends on many factors, such as sequencing technologies and bioinformatic tools available at the time. However, a new genome represents an unarguably powerful tool to study species biology and is a fundamental tool that will open a completely new way to study them.

3.3.3 Freshwater mussels’ genome assemblies

Whole genome assemblies of FMs species are relatively recent, limited and scattered resources. To this date, only four FMs species have a whole genome assembly available. The first FMs genome was presented in 2018, from *Venustaconcha ellipsiformis* (Renaut et al., 2018), followed by three genomes presented in 2021, from *Megalonaias nervosa* (Rogers et al., 2021), *Potamilus streckersoni* (Smith, 2021) and the freshwater pearl mussel genome, *M. margaritifera*, provided in Paper 6 (further improved in Manuscript 1). Although the latter three suggested an optimistic trend (similar to the trend observed for molluscs as a whole), the fact is that in 2022 no FMs genome has been made publicly available (until November). Furthermore, three of these four genomes belong to species from the same family, i.e., Unionidae. The genomes generated in Paper 6 and Manuscript 1 represent the first genome assemblies from a distinct FMs family, i.e., Margaritiferidae. FMs emerged nearly 300 Mya ago (Bauer, 2001; Graf et al., 2015) and throughout their ancient evolutionary history split into six highly distinct families (Graf and Cummings, 2021; Pfeiffer et al., 2019). As highlighted before in this thesis, comprehensive and well-structured taxa sampling is a fundamental step in comparative genomics (Dunn and Ryan, 2015; Richards, 2015; Stephan et al., 2022), thus expanding genome sequencing to include other families, such as the genomes here presented, is essential to accurately explore the evolutionary history of the group.

The reduced number of FMs genome available results, in part, from the fact that they share two of the three main challenges for genome assembly software (discussed in the subsection above), i.e., FMs have among the largest genomes within Bivalvia (1.8-2.5 Gbp) and, perhaps the most challenging of all, among the highest percentage of repetitive elements of all Bivalvia genomes (37.81% - 59.07% of the genome) (Papers 1 and 6; Smith, 2021). The combination of these two features has highly impacted the contiguity of short-read-based assemblies. This is evident in the genome assemblies of *M. margaritifera* (Paper 6), *M. nervosa* (Rogers et al., 2021), and *V. ellipsiformis* (Renaut et al., 2018). Although both *M. nervosa* and *V. ellipsiformis* also included long-read sequencing, at low coverage (i.e., a hybrid assembly strategy), fragmentation levels were still elevated (Renaut et al., 2018; Rogers et al., 2021). Interestingly, the assembly of the *M. margaritifera* genome had the highest scaffold contiguity of the three, which is impressive considering it was assembled solely with short reads and is the largest (2.4 Gbp) and most repetitive (59.07%) of all FMs genomes available (Paper 6). In fact, in Paper 6 the overall good quality of the genome assembly is demonstrated using several statistics and through the characterization, for the first time, of the Hox and ParaHox gene families. However, the genome assembly is still highly fragmented, preventing more robust analyses, such as the entire characterization of the entire Hox clusters at a macrosyntenic level (Paper 6), which is disrupted in many molluscan lineages (Z. Yang et al., 2020). Consequently, and as shown by the genomes of *P. streckersoni* (Smith, 2021) and the improved *M. margaritifera* genome (Manuscript 1), future FMs genome assemblies should necessarily rely on high-coverage long-read approaches. The improved *M. margaritifera* genome presented in Manuscript 1 shows a nearly 11 times contiguity improvement, accompanied by a substantial improvement in completeness with almost no missing near-universal single-copy orthologous from both eukaryotic and metazoan curated lists (Manni et al., 2021). The potential of this new assembly has not been explored yet, however, its significant increase in contiguity will certainly provide a prime resource in future studies. Furthermore, the genome is also a prime candidate for chromosome anchoring using recently developed and highly efficient strategies, such as chromatin crosslinking protocols (Sedlazeck et al., 2018). This would represent a fundamental step in the genomics of FMs since no chromosome-level genome assembly has been generated. A chromatin crosslinking protocol has already been attempted by Rogers et al., (2021) to anchor the *M. nervosa* genome. This attempt was unsuccessful and although the authors could not determine the reasons, they raised the well-known presence of polysaccharides and polyphenols in molluscs tissues as a possible explanation. Although only a single attempt was carried out, this may suggest another challenge in FMs assemblies and should be further explored in future studies.

3.3.4 “*Lift the curtain*”: Biological Applications of the genome assemblies to study molluscs and FMs

Mollusc genomes have provided an unprecedented opportunity to study the biology and evolution of many of the highly complex lineages of the group (Reviewed in Paper 1; Z. Yang et al., 2020). Genome assemblies, combined with other genomics approaches (especially transcriptomic), are now allowing insights into the molecular mechanism that govern many of defining features of the group (Reviewed in Paper 1; Z. Yang et al., 2020), including: the evolution of diverse body plans through changes in number and expression patterns of Hox genes clusters (e.g., Paper 1; Sun et al., 2019; S. Wang et al., 2017); the evolution of the eye through the characterization of both photoreceptors (e.g., r-opsin, G_o-opsin and c-like opsins) and eye development genes (e.g., Pax2/5/8, Brn3, and Lmx1) (e.g., Li et al., 2017; S. Wang et al., 2017); the evolution and development of the cephalopod neural system, through novel genes recruitment in morphogenetic pathways as well as recombination, duplication, and divergence of coding genes and regulatory regions (e.g., expanded protocadherins and C2H2 superfamily), many previously thought unique of vertebrates (e.g., Albertin et al., 2015; Yoshida et al., 2015); the evolution of muscle and foot through the identification of the repertoire of muscle development genes, as well as genes related to energy production and byssus-related proteins (e.g., Funabara et al., 2013; Li et al., 2017) and; the evolution of the shell, through the identification of several shell formation and regulatory genes involved in the biomineralization toolkit, including matrix-framework formation genes similar to those found in vertebrate bone (collagen-related VWA-containing proteins), as well as the invertebrate-specific chitin-based matrices (e.g., Du et al., 2017; Zhang et al., 2012).

Furthermore, as more genomes are sequenced and studied, insights into the mechanisms that “create” genetic novelties are starting to be understood. The emergence of novel genes is a fundamental step towards evolutionary innovation and may originate through several mechanisms, that include duplication, horizontal gene transfer, gene fusion/fission, exon shuffling, and *de novo* formation (Cai et al., 2008; Ding et al., 2012). Gene duplication through expansion of lineage-specific gene families has been identified in several molluscs and linked with many evolutionary adaptations (e.g., stress response, innate immune systems, eye evolution, freshwater adaptation, toxin production), suggesting a key role for this mechanism throughout the evolutionary history of the phylum (Z. Yang et al., 2020). On the other hand, the increased and diversified number of genomes is beginning to reveal astonishingly unusual and yet transversal patterns in mollusc genomes, such as the widespread distribution of hemizygous regions (significant gene content variation between individuals) (Calcino et al., 2021; Gerdol et al., 2020; Takeuchi et al., 2022).

Compared with other molluscan groups, the genomics of FMs is still in its infancy. Most of the genomics studies that sought to link the biological features of FMs with genetic patterns have been largely based on transcriptomic data. These studies have, among other things, explored transcriptomic responses to climate changes and stress (Luo et al., 2014; Patnaik et al., 2016; Roznere et al., 2018; R. Wang et al., 2015), transcriptomic responses to pollutants (Bertucci et al., 2017; Cornman et al., 2014; Robertson et al., 2017), provide transcriptomic surveys of immunological responses (D. Huang et al., 2019; Q. Yang et al., 2021), identified genes linked to sex determination and DUI (Capt et al., 2018, 2019; Shi et al., 2015), and identified genes linked to shell, nacre and pearl formation (Chen et al., 2019; X. Wang et al., 2017). Most of these studies provide a preliminary characterization and/or survey of genetic features, through broad comparative genetic approaches. The information generated by these studies demonstrates the potential of transcriptomic tools to explore genetic novelties and adaptations in FMs. In Paper 5 of this thesis, the gill transcriptomes of five FMs species were provided, four of which represent the first genomic resource ever sequenced for the respective species. Although these transcriptomes were not explored, they will serve as frameworks for future studies to identify genetic novelties and expression patterns in the targeted species.

Given the very recent history of FMs' genome sequencing, little biological and evolutionary information has been retrieved from them. Only two publications have begun exploring the true potential of the FMs genome assemblies, i.e., in Paper 6, where the Hox and ParaHox genes were identified and characterized for the first time in FMs, and in the publication of the *M. nervosa* genome, which represents the most comprehensive genome studied of a FMs species (Rogers et al., 2021). The astonishing significantly relevant information uncovered in this genome study alone (which was complemented with transcriptomic data) perfectly reflects the underlying leverage resulting from producing a genome assembly. The authors provided an in-depth exploration of the contribution of transposable elements (TE), gene family expansion, and single nucleotide mutations to genetic change, as well as their impact on the adaptative potential of the species (Rogers et al., 2021). The authors also identified several gene family expansions (some with increased rates of amino acid changes and a signature of selection) that they were able to be linked with: detox and stress response, (e.g. cytochrome P450, ABC transporters, and Hsp70), which may be related with the species tolerance to toxicity and thermal fluctuations; anticoagulation action (e.g., von Willebrand factor proteins, Xa-binding genes, and fibrinogen binding proteins), which likely play a role in the survival of the parasitic larval stage when attached to host fish; mitochondrial regulation and electron transport chain (e.g., mitochondria-eating genes and cytochromes), which might play a role in the mitochondrial functioning and DUI; shell formation (e.g., chitin and chitin-binding Peritrophin-A genes); and light sensitivity (e.g., opsin and rhodopsin). These results show that similarly to most molluscs, gene family expansions are a key driver of the adaptive evolution of *M. nervosa*. Furthermore, a survey of TEs revealed a recent burst of TEs activity, with a significant impact on the *M. nervosa* genome size and structure. Transposable elements are important drivers of gene remodelling by promoting gene capturing, ectopic recombination and retrogene formation (Oliver and Greene, 2009). The correlation of enhanced TEs activity with the high rates of the expanded gene families observed in *M. nervosa*, suggests that TEs have a fundamental role in the emergence of genetic novelties in the species. Consequently, given the high content of repetitive elements documented in FMs (discussed above), TEs might represent a key adaptative “tool” for the group as a whole. Moreover, given that before genome annotation, the standard approach of genome projects was to mask highly repetitive sequences (Paper 1), future genome studies should take this into account, as masking might reduce the likelihood of identifying many members of TEs. Finally, using two FMs genomes, the authors were also able to estimate effective population size (Ne) and establishment time of selective sweep on new mutations, providing insights into the species' tempo of evolution and adaptative potential. Estimating these metrics for highly threatened species/populations is extremely important to understand their genetic predisposition to cope with the increasing threats and thus help in conservation and management strategies.

In general, although Rogers et al., (2021) focused on one species, it lifted the curtain of possibilities that emerge from sequencing a single genome. Several genomic novelties were linked with well-known history traits of FMs (e.g., host interaction, shell formation), response to environmental changes (e.g., response to detox and thermal stress), as well as to adaptative potential (i.e., gene family expansion and TEs proliferation), which provide a guided framework that should be further explored in other FMs genomes.

Considering the many contrasting biological and ecological features of *M. nervosa* when compared with *M. margaritifera*, the genomes presented in this thesis (Paper 6 and Manuscript 1) provide a comparative resource to further explore the link of genetic patterns with biological function (Figure D.3). Unlike *M. nervosa,* which can parasite a wide range of fish species (Woody and Holland-Bartels, 2011), *M. margaritifera* has an extremely limited host specificity (Lopes-Lima et al., 2018a). Therefore, by exploring the number and type of expanded anticoagulation action genes in the *M. margaritifera* (especially if complemented with transcriptomics data from glochidia), future studies could further assert their functional role in parasite-host interaction (Figure D.3). Furthermore, *M. margaritifera* is an ecosystem specialist and highly sensitive to habitat disruption (Geist, 2010), while *M. nervosa* seems to be significantly resistant to environmental challenges (Rogers et al., 2021). This might be resulting from a completely distinct repertoire of detox and stress response gene families in both species (Figure D.3). Similarly, the shell formation gene repertoire of pearl-forming FMs with genomics data available, i.e., *M. margaritifera* and *Cristaria plicata* (X. Wang et al., 2017), can be compared with non-pearl forming species (Figure D.3). The apparent impact that TEs have on the *M. nervosa* genome, both in size and structure, should now be explored in the *M. margaritifera* genome as well. Given that *M. margaritifera* is the largest and most repetitive FMs genome sequenced to date (Paper 1 and Manuscript 1), a more profound survey of repetitive elements and TEs will provide a deeper understanding of the role these structural features in the genome evolution of the species and FMs as a whole (Figure D.3).

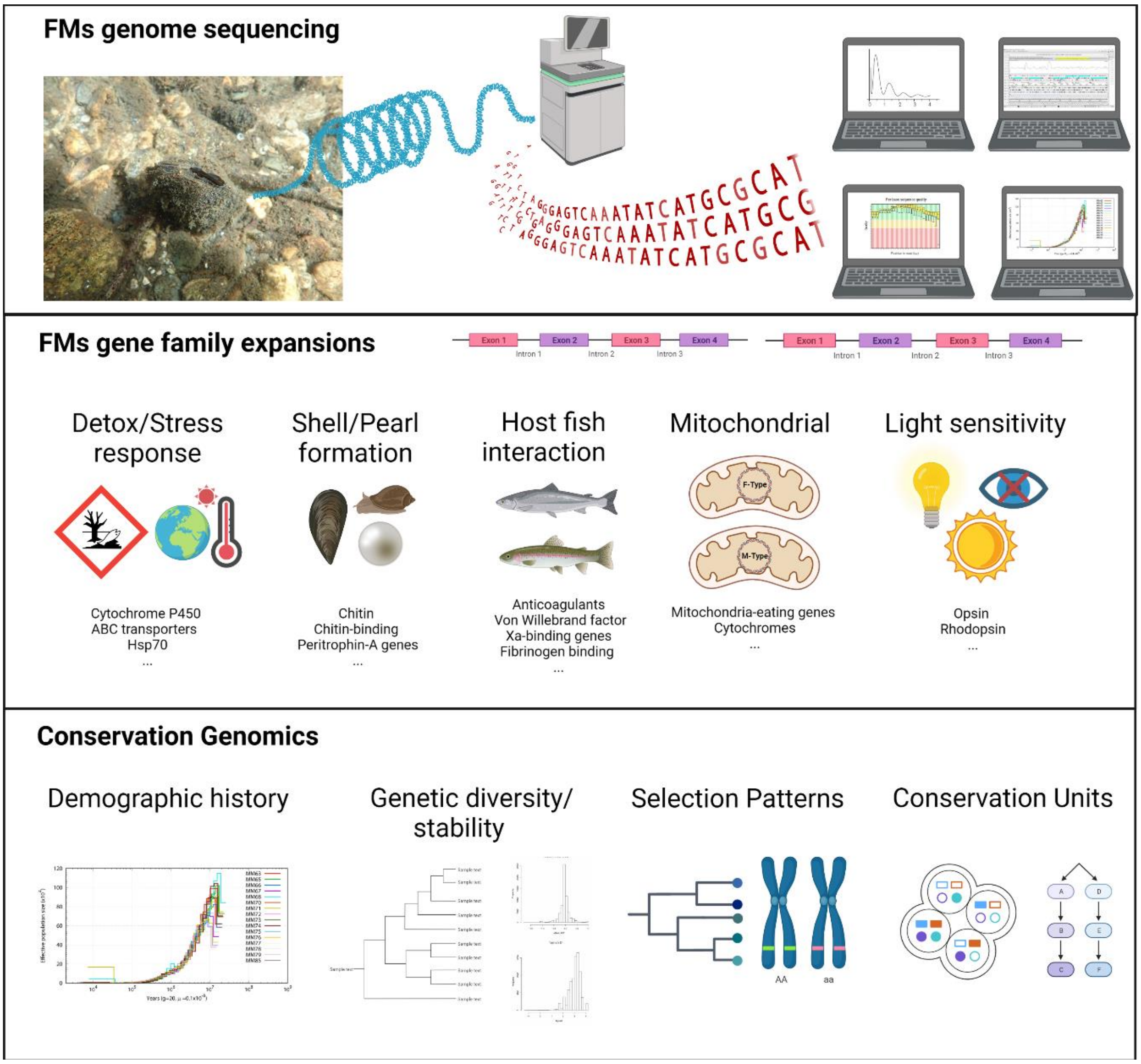


Figure D. 3 - Potencial application of the whole genome assembly of the freshwater pearl mussel *Margaritifera margaritifera*.

Finally, given the global decline of FMs, new genomic resources offer a benchmark to monitor, identify, and classify unities with conservation priority, classify genetic features with conservation importance, and infer the genetic signature of adaptive potential (Oppen and Coleman, 2022; Paez et al., 2022) (Figure D.3). Genomic data, by informing about the health of populations (e.g., size, connectivity, and hybridization), determining the patterns of genetic erosion and disposition for species to persist in the face of environmental change, as well as guiding conservation-targeted genetic manipulations, can significantly increase the success of conservation efforts (Bertorelle et al., 2022; Hohenlohe et al., 2021; Oppen and Coleman, 2022; Paez et al., 2022) (Figure D.3). At the genomic scale, even with relatively reduced sampling, as shown by the *M. nervosa* genome project (Rogers et al., 2021), a wide range of extremely valuable information with conservation relevancy can be acquired. Consequently, increasing the availability of FMs genomes, especially of the most threatened groups (such as margaritiferids), is crucial to increase the knowledge of their biology and ultimately promote their conservation (Figure D.3).

Conclusion

The sequencing revolution of the last two decades has opened the way for the study of all Tree of Life (ToL) “from genes to genomes”. However, the pace at which the genomic revolution is reaching each branch of the ToL is not equitable, with many groups being poorly studied or not studied at all (Stephan et al., 2022). Molluscs, despite being among the oldest, most diverse and most successfully established organisms, only recently entered this new era. The results presented in this thesis, particularly in Chapter 2, provide a thorough review of the progress made by molluscan genomics, summarizing the many fundamental discoveries that emerged from the still-reduced number of studies. Moreover, the work here presented emphasizes the many gaps of knowledge and biased targets of the early stages of molluscan genomics, lifting the curtain of possibilities that are yet to come.

On a more targeted aim, this thesis focused on the most diverse order of freshwater bivalves, the freshwater mussels (FMs) (Graf and Cummings, 2007, 2022). Although FMs are widely dispersed, biologically fascinating and among the most imperil organisms, the number of genomic resources available is astonishingly limited. This is particularly evident for margaritiferids, the most threatened group of FMs. The results provided in this thesis, particularly in Chapter 3, represent a major step in the genomics of margaritiferids by generating a series of novel resources that encompass a large spectrum of genomic applications, including several new whole mitogenomes, a set of target enrichment sequences, a new transcriptome and two whole genome assemblies of the freshwater pearl mussel *Margaritifera margaritifera*. These genome assemblies are the first available for margaritiferid and only the fourth, and fifth, available for FMs. The applications of these newly generated resources are also slightly explored, particularly for phylogenomics and evolutionary genomics studies. Finally, a careful review of the many putative applications of these resources, as well as the way forward in the genomics of FMs, is provided in the last chapter of the thesis.

Overall, the numerous genomics resources generated in this thesis represent a key step towards the study of margaritiferids (and FMs). These results mark the beginning of the long run towards unravelling the many exceptionally fascinating biological and evolutionary traits of this inconspicuous group of organisms.

References

Abalde, S., Tenorio, M.J., Afonso, C.M.L., Uribe, J.E., Echeverry, A.M., Zardoya, R., 2017. Phylogenetic relationships of cone snails endemic to Cabo Verde based on mitochondrial genomes. BMC Evol Biol 17, 1–19. https://doi.org/10.1186/S12862-017-1069-X/FIGURES/5

Abascal, F., Zardoya, R., Posada, D., 2005. ProtTest: Selection of best-fit models of protein evolution. Bioinformatics 21, 2104–2105. https://doi.org/10.1093/bioinformatics/bti263

Adams, K.L., Palmer, J.D., 2003. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. Mol Phylogenet Evol 29, 380–395. https://doi.org/10.1016/S1055-7903(03)00194-5

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N., Sutton, G.G., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X., Brandon, R.C., Rogers, Y.-H.C., Blazej, R.G., Champe, M., Pfeiffer, B.D., Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor, G.L., Miklos, Abril, J.F., Agbayani, A., An, H.-J., Andrews-Pfannkoch, C., Baldwin, D., Ballew, R.M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y., Benos, P. v., Berman, B.P., Bhandari, D., Bolshakov, S., Borkova, D., Botchan, M.R., Bouck, J., Brokstein, P., Brottier, P., Burtis, K.C., Busam, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B., Davies, P., Pablos, B. de, Delcher, A., Deng, Z., Mays, A.D., Dew, I., Dietz, S.M., Dodson, K., Doup, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C., Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Ferriera, S., Fleischmann, W., Fosler, C., Gabrielian, A.E., Garg, N.S., Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Gorrell, J.H., Gu, Z., Guan, P., Harris, M., Harris, N.L., Harvey, D., Heiman, T.J., Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J., Wei, M.-H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z., Kennison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C., Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T.C., McLeod, M.P., McPherson, D., Merkulov, G., Milshina, N. v., Mobarry, C., Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L., Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K., Nusskern, D.R., Pacleb, J.M., Palazzolo, M., Pittman, G.S., Pan, S., Pollard, J., Puri, V., Reese, M.G., Reinert, K., Remington, K., Saunders, R.D.C., Scheeler, F., Shen, H., Shue, B.C., Sidén-Kiamos, I., Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C., Stapleton, M., Strong, R., Sun, E., Svirskas, R., Tector, C., Turner, R., Venter, E., Wang, A.H., Wang, X., Wang, Z.-Y., Wassarman, D.A., Weinstock, G.M., Weissenbach, J., Williams, S.M., Woodage, T., Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.-F., Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H., Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O., Gibbs, R.A., Myers, E.W., Rubin, G.M., Venter, J.C., 2000. The Genome Sequence of *Drosophila melanogaster*. Science (1979) 287, 2185–2195. https://doi.org/10.1126/science.287.5461.2185

Adema, C.M., Hillier, L.W., Jones, C.S., Loker, E.S., Knight, M., Minx, P., Oliveira, G., Raghavan, N., Shedlock, A., do Amaral, L.R., Arican-Goktas, H.D., Assis, J.G., Baba, E.H., Baron, O.L., Bayne, C.J., Bickham-Wright, U., Biggar, K.K., Blouin, M., Bonning, B.C., Botka, C., Bridger, J.M., Buckley, K.M., Buddenborg, S.K., Lima Caldeira, R., Carleton, J., Carvalho, O.S., Castillo, M.G., Chalmers, I.W., Christensens, M., Clifton, S., Cosseau, C., Coustau, C., Cripps, R.M., Cuesta-Astroz, Y., Cummins, S.F., di Stephano, L., Dinguirard, N., Duval, D., Emrich, S., Feschotte, C., Feyereisen, R., FitzGerald, P., Fronick, C., Fulton, L., Galinier, R., Gava, S.G., Geusz, M., Geyer, K.K., Giraldo-Calderón, G.I., de Souza Gomes, M., Gordy, M.A., Gourbal, B., Grunau, C., Hanington, P.C., Hoffmann, K.F., Hughes, D., Humphries, J., Jackson, D.J., Jannotti-Passos, L.K., de Jesus Jeremias, W., Jobling, S., Kamel, B., Kapusta, A., Kaur, S., Koene, J.M., Kohn, A.B., Lawson, D., Lawton, S.P., Liang, D., Limpanont, Y., Liu, S., Lockyer, A.E., Lovato, T.L., Ludolf, F., Magrini, V., McManus, D.P., Medina, M., Misra, M., Mitta, G., Mkoji, G.M., Montague, M.J., Montelongo, C., Moroz, L.L., Munoz-Torres, M.C., Niazi, U., Noble, L.R., Oliveira, F.S., Pais, F.S., Papenfuss, A.T., Peace, R., Pena, J.J., Pila, E.A., Quelais, T., Raney, B.J., Rast, J.P., Rollinson, D., Rosse, I.C., Rotgans, B., Routledge, E.J., Ryan, K.M., Scholte, L.L.S., Storey, K.B., Swain, M., Tennessen, J.A., Tomlinson, C., Trujillo, D.L., Volpi, E. v., Walker, A.J., Wang, T., Wannaporn, I., Warren, W.C., Wu, X.-J., Yoshino, T.P., Yusuf, M., Zhang, S.-M., Zhao, M., Wilson, R.K., 2017. Whole genome analysis of a schistosomiasis-transmitting freshwater snail. Nat Commun 8, 15451. https://doi.org/10.1038/ncomms15451

Adessi, C., Matton, G., Ayala, G., Turcatti, G., Mermod, J.J., Mayer, P., Kawashima, E., 2000. Solid phase DNA amplification: characterisation of primer attachment and amplification mechanisms. Nucleic Acids Res 28, e87–e87. https://doi.org/10.1093/NAR/28.20.E87

Agarwala, R., Barrett, T., Beck, J., Benson, D.A., Bollin, C., Bolton, E., Bourexis, D., Brister, J.R., Bryant, S.H., Canese, K., Charowhas, C., Clark, K., Dicuccio, M., Dondoshansky, I., Federhen, S., Feolo, M., Funk, K., Geer, L.Y., Gorelenkov, V., Hoeppner, M., Holmes, B., Johnson, M., Khotomlianski, V., Kimchi, A., Kimelman, M., Kitts, P., Klimke, W., Krasnov, S., Kuznetsov, A., Landrum, M.J., Landsman, D., Lee, J.M., Lipman, D.J., Lu, Z., Madden, T.L., Madej, T., Marchler-Bauer, A., Karsch-Mizrachi, I., Murphy, T., Orris, R., Ostell, J., O'sullivan, C., Panchenko, A., Phan, L., Preuss, D., Pruitt, K.D., Rodarmer, K., Rubinstein, W., Sayers, E., Schneider, V., Schuler, G.D., Sherry, S.T., Sirotkin, K., Siyan, K., Slotta, D., Soboleva, A., Soussov, V., Starchenko, G., Tatusova, T.A., Todorov, K., Trawick, B.W., Vakatov, D., Wang, Y., Ward, M., Wilbur, W.J., Yaschenko, E., Zbicz, K., 2016. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 44, D7–D19. https://doi.org/10.1093/NAR/GKV1290

Agriculture Organization of the United Nations. Fisheries Department, 2000. The State of World Fisheries and Aquaculture, 2000. Food & Agriculture Org.

Aguilera, F., McDougall, C., Degnan, B.M., Irwin, D., 2017. Co-Option and *De Novo* Gene Evolution Underlie Molluscan Shell Diversity. Mol Biol Evol 34, 779–792. https://doi.org/10.1093/MOLBEV/MSW294

Akasaki, T., Nikaido, M., Tsuchiya, K., Segawa, S., Hasegawa, M., Okada, N., 2006. Extensive mitochondrial gene arrangements in coleoid Cephalopoda and their phylogenetic implications. Mol Phylogenet Evol 38, 648–658. https://doi.org/10.1016/J.YMPEV.2005.10.018

Albertin, C.B., Simakov, O., Mitros, T., Wang, Z.Y., Pungor, J.R., Edsinger-Gonzales, E., Brenner, S., Ragsdale, C.W., Rokhsar, D.S., 2015. The octopus genome and the evolution of cephalopod neural and morphological novelties. Nature 524, 220–224. https://doi.org/10.1038/nature14668

Alda, F., Ludt, W.B., Elías, D.J., McMahan, C.D., Chakrabarty, P., 2021. Comparing Ultraconserved Elements and Exons for Phylogenomic Analyses of Middle American Cichlids: When Data Agree to Disagree. Genome Biol Evol 13. https://doi.org/10.1093/GBE/EVAB161

Allcock, A.L., Cooke, I.R., Strugnell, J.M., 2011. What can the mitochondrial genome reveal about higher-level phylogeny of the molluscan class Cephalopoda? Zool J Linn Soc 161, 573–586. https://doi.org/10.1111/J.1096-3642.2010.00656.X

Allen, J.F., Horner, D.S., Cavalier-Smith, T., Willison, K., Leaver, C.J., Martin, W., 2003. The function of genomes in bioenergetic organelles. Philos Trans R Soc Lond B Biol Sci 358, 19–38. https://doi.org/10.1098/RSTB.2002.1191

Allen, J.M., Boyd, B., Nguyen, N.P., Vachaspati, P., Warnow, T., Huang, D.I., Grady, P.G.S., Bell, K.C., Cronk, Q.C.B., Mugisha, L., Pittendrigh, B.R., Leonardi, M.S., Reed, D.L., Johnson, K.P., 2017. Phylogenomics from Whole Genome Sequences Using aTRAM. Syst Biol 66, 786–798. https://doi.org/10.1093/SYSBIO/SYW105

Allendorf, F.W., Hohenlohe, P.A., Luikart, G., 2010. Genomics and the future of conservation genetics. Nature Reviews Genetics 2010 11:10 11, 697–709. https://doi.org/10.1038/nrg2844

Almeida, D., Maldonado, E., Vasconcelos, V., Antunes, A., 2015. Adaptation of the Mitochondrial Genome in Cephalopods: Enhancing Proton Translocation Channels and the Subunit Interactions. PLoS One 10, e0135405. https://doi.org/10.1371/JOURNAL.PONE.0135405

Al-Nakeeb, K., Petersen, T.N., Sicheritz-Pontén, T., 2017. Norgal: Extraction and de novo assembly of mitochondrial DNA from whole-genome sequencing data. BMC Bioinformatics 18, 1–7. https://doi.org/10.1186/S12859-017-1927-Y/TABLES/2

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J Mol Biol 215, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Amaro, R., Bouza, C., Pardo, B.G., Castro, J., San Miguel, E., Villalba, A., Lois, S., Outeiro, A., Ondina, P., 2016. Identification of novel gender-associated mitochondrial haplotypes in *Margaritifera margaritifera* (Linnaeus, 1758). Zool J Linn Soc 179, 738–750. https://doi.org/10.1111/zoj.12472

Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H.L., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J.H., Staden, R., Young, I.G., 1981. Sequence and organization of the human mitochondrial genome. Nature 1981 290:5806 290, 457–465. https://doi.org/10.1038/290457a0

Andreson, R., Roosaare, M., Kaplinski, L., Laht, S., Kõressaar, T., Lepamets, M., Brauer, A., Kukuškina, V., Remm, M., 2019. Gene content of the fish-hunting cone snail Conus consors. bioRxiv 590695. https://doi.org/10.1101/590695

Andrew, R.L., Bernatchez, L., Bonin, A., Buerkle, C.Alex., Carstens, B.C., Emerson, B.C., Garant, D., Giraud, T., Kane, N.C., Rogers, S.M., Slate, J., Smith, H., Sork, V.L., Stone, G.N., Vines, T.H., Waits, L., Widmer, A., Rieseberg, L.H., 2013. A road map for molecular ecology. Mol Ecol. https://doi.org/10.1111/mec.12319

Andrews, K.R., Good, J.M., Miller, M.R., Luikart, G., Hohenlohe, P.A., 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. Nat Rev Genet. https://doi.org/10.1038/nrg.2015.28

Ansorge, W., Sproat, B., Stegemann, J., Schwager, C., Zenke, M., 1987. Automated DNA sequencing: ultrasensitive detection of fluorescent bands during electrophoresis. Nucleic Acids Res 15, 4593–4602. https://doi.org/10.1093/NAR/15.11.4593

Ansorge, W., Sproat, B.S., Stegemann, J., Schwager, C., 1986. A non-radioactive automated method for DNA sequence determination. J Biochem Biophys Methods 13, 315–323. https://doi.org/10.1016/0165-022X(86)90038-2

Araujo, R., 2011. *Unio delphinus*. The IUCN Red List of Threatened Species. URL https://www.iucnredlist.org/species/195510/8975648 (accessed 5.20.22).

Araujo, R., Buckley, D., Nagel, K.O., García-Jiménez, R., Machordom, A., 2018. Species boundaries, geographic distribution and evolutionary history of the Western palaearctic freshwater mussels unio (Bivalvia: Unionidae). Zool J Linn Soc 182, 275–299. https://doi.org/10.1093/zoolinnean/zlx039

Araujo, R., Schneider, S., Roe, K.J., Erpenbeck, D., Machordom, A., 2017. The origin and phylogeny of Margaritiferidae (Bivalvia, Unionoida): a synthesis of molecular and fossil data. Zool Scr 46, 289–307. https://doi.org/10.1111/zsc.12217

Araujo, R., Toledo, C., van Damme, D., Ghamizi, M., Machordom, A., 2009. *Margaritifera marocana* (Pallary, 1918): a valid species inhabiting Moroccan rivers. Journal of Molluscan Studies 75, 95–101. https://doi.org/10.1093/mollus/eyn043

Arseneau, J.R., Steeves, R., Laflamme, M., 2017. Modified low-salt CTAB extraction of high-quality DNA from contaminant-rich tissues. Mol Ecol Resour 17, 686–693. https://doi.org/10.1111/1755-0998.12616

Arumugam, R., Uli, J.E., Annavi, G., 2019. A Review of the Application of Next Generation Sequencing (NGS) in Wild Terrestrial Vertebrate Research. Annu Res Rev Biol 1–9. https://doi.org/10.9734/arrb/2019/v31i530061

Atkinson, C.L., Vaughn, C.C., Forshay, K.J., Cooper, J.T., 2013. Aggregated filter-feeding consumers alter nutrient limitation: Consequences for ecosystem and community dynamics. Ecology 94, 1359–1369. https://doi.org/10.1890/12-1531.1

Baer, R., Bankier, A.T., Biggin, M.D., Deininger, P.L., Farrell, P.J., Gibson, T.J., Hatfull, G., Hudson, G.S., Satchwell, S.C., Séguin, C., Tuffnell, P.S., Barrell, B.G., 1984. DNA sequence and expression of the B95-8 Epstein—Barr virus genome. Nature 310, 207–211. https://doi.org/10.1038/310207a0

Bailey, C.H., Castellucci, V.F., Koester, J., Chen, M., 1983. Behavioral changes in aging Aplysia: A model system for studying the cellular basis of age-impaired learning, memory, and arousal. Behav Neural Biol 38, 70–81. https://doi.org/10.1016/S0163-1047(83)90399-0

Baillie, J.E.M., Butcher, E.R., 2012. Priceless or Worthless? The world’s most threatened species. Zoological Society of London, United Kingdom. URL http://copa.acguanacaste.ac.cr:8080/handle/11606/655 (accessed 11.4.20).

Bairoch, A., Apweiler, R., 1999. The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1999. Nucleic Acids Res. https://doi.org/10.1093/nar/27.1.49

Bandyopadhyay, P.K., Stevenson, B.J., Cady, M.T., Olivera, B.M., Wolstenholme, D.R., 2006. Complete mitochondrial DNA sequence of a Conoidean gastropod, Lophiotoma (Xenuroturris) cerithiformis: Gene order and gastropod phylogeny. Toxicon 48, 29–43. https://doi.org/10.1016/J.TOXICON.2006.04.013

Bankier, A.T., Beck, S., Bohni, R., Brown, C.M., Cerny, R., Chee, M.S., Hutchison Iii, C.A., Kouzarides, T., Martignetti, J.A., Preddie, E., Satchwell, S.C., Tomlinson, P., Weston, K.M., Barrell, B.G., 1991. The DNA sequence of the human cytomegalovirus genome. DNA Sequence 2, 1–11. https://doi.org/10.3109/10425179109008433

Bao, W., Kojima, K.K., Kohany, O., 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob DNA 6, 11. https://doi.org/10.1186/s13100-015-0041-9

Barghi, N., Concepcion, G.P., Olivera, B.M., Lluisma, A.O., 2016. Structural features of conopeptide genes inferred from partial sequences of the *Conus tribblei* genome. Molecular Genetics and Genomics 291, 411–422. https://doi.org/10.1007/s00438-015-1119-2

Barghi, N., Hermisson, J., Schlötterer, C., 2020. Polygenic adaptation: a unifying framework to understand positive selection. Nat Rev Genet 1–13. https://doi.org/10.1038/s41576-020-0250-z

Barnhart, M.C., Haag, W.R., Roston, W.N., 2015. Adaptations to host infection and larval parasitism in Unionoida. https://doi.org/10.1899/07-093.1 27, 370–394. https://doi.org/10.1899/07-093.1

Barr, C.M., Neiman, M., Taylor, D.R., 2005. Inheritance and recombination of mitochondrial genomes in plants, fungi and animals. New Phytologist 168, 39–50. https://doi.org/10.1111/J.1469-8137.2005.01492.X

Barros, M.H., Tzagoloff, A., 2017. Aep3p-dependent translation of yeast mitochondrial ATP8. Mol Biol Cell 28, 1426–1434. https://doi.org/10.1091/MBC.E16-11-0775/ASSET/IMAGES/LARGE/1426FIG4.JPEG

Barton-Owen, T.B., Szabó, R., Somorjai, I.M.L., Ferrier, D.E.K., 2018. A revised spiralian homeobox gene classification incorporating new polychaete transcriptomes reveals a diverse TALE class and a divergent hox gene. Genome Biol Evol 10, 2151–2167. https://doi.org/10.1093/gbe/evy144

Bateman, A., Martin, M.J., O’Donovan, C., Magrane, M., Alpi, E., Antunes, R., Bely, B., Bingley, M., Bonilla, C., Britto, R., Bursteinas, B., Bye-AJee, H., Cowley, A., da Silva, A., de Giorgi, M., Dogan, T., Fazzini, F., Castro, L.G., Figueira, L., Garmiri, P., Georghiou, G., Gonzalez, D., Hatton-Ellis, E., Li, W., Liu, W., Lopez, R., Luo, J., Lussi, Y., MacDougall, A., Nightingale, A., Palka, B., Pichler, K., Poggioli, D., Pundir, S., Pureza, L., Qi, G., Rosanoff, S., Saidi, R., Sawford, T., Shypitsyna, A., Speretta, E., Turner, E., Tyagi, N., Volynkin, V., Wardell, T., Warner, K., Watkins, X., Zaru, R., Zellner, H., Xenarios, I., Bougueleret, L., Bridge, A., Poux, S., Redaschi, N., Aimo, L., ArgoudPuy, G., Auchincloss, A., Axelsen, K., Bansal, P., Baratin, D., Blatter, M.C., Boeckmann, B., Bolleman, J., Boutet, E., Breuza, L., Casal-Casas, C., de Castro, E., Coudert, E., Cuche, B., Doche, M., Dornevil, D., Duvaud, S., Estreicher, A., Famiglietti, L., Feuermann, M., Gasteiger, E., Gehant, S., Gerritsen, V., Gos, A., Gruaz-Gumowski, N., Hinz, U., Hulo, C., Jungo, F., Keller, G., Lara, V., Lemercier, P., Lieberherr, D., Lombardot, T., Martin, X., Masson, P., Morgat, A., Neto, T., Nouspikel, N., Paesano, S., Pedruzzi, I., Pilbout, S., Pozzato, M., Pruess, M., Rivoire, C., Roechert, B., Schneider, M., Sigrist, C., Sonesson, K., Staehli, S., Stutz, A., Sundaram, S., Tognolli, M., Verbregue, L., Veuthey, A.L., Wu, C.H., Arighi, C.N., Arminski, L., Chen, C., Chen, Y., Garavelli, J.S., Huang, H., Laiho, K., McGarvey, P., Natale, D.A., Ross, K., Vinayaka, C.R., Wang, Q., Wang, Y., Yeh, L.S., Zhang, J., 2017. UniProt: the universal protein knowledgebase. Nucleic Acids Res 45, D158–D169. https://doi.org/10.1093/NAR/GKW1099

Bauer, G., 2001. Ecology and Evolution of the Freshwater Mussels Unionoida. Springer Science & Business Media.

Bauer, G., 1992. Variation in the Life Span and Size of the Freshwater Pearl Mussel. J Anim Ecol 61, 425. https://doi.org/10.2307/5333

Belcaid, M., Casaburi, G., McAnulty, S.J., Schmidbaur, H., Suria, A.M., Moriano-Gutierrez, S., Sabrina Pankey, M., Oakley, T.H., Kremer, N., Koch, E.J., Collins, A.J., Nguyen, H., Lek, S., Goncharenko-Foster, I., Minx, P., Sodergren, E., Weinstock, G., Rokhsar, D.S., McFall-Ngai, M., Simakov, O., Foster, J.S., Nyholm, S. v., 2019. Symbiotic organs shaped by distinct modes of genome evolution in cephalopods. Proc Natl Acad Sci U S A 116, 3030–3035. https://doi.org/10.1073/pnas.1817322116

Benaissa, H., Ghamizi, M., Teixeira, A., Sousa, R., Rassam, H., Varandas, S., Lopes-Lima, M., 2022. Preliminary data on fish hosts and their conservation importance for the critically endangered *Pseudunio marocanus* (Pallary, 1918). Aquat Conserv 32, 229–238. https://doi.org/10.1002/AQC.3756

Bennett, C.F., Latorre-Muro, P., Puigserver, P., 2022. Mechanisms of mitochondrial respiratory adaptation. Nature Reviews Molecular Cell Biology 2022 1–19. https://doi.org/10.1038/s41580-022-00506-6

Bennett, S., 2004. Solexa Ltd. http://dx.doi.org/10.1517/14622416.5.4.433 5, 433–438. https://doi.org/10.1517/14622416.5.4.433

Bennett, S.T., Barnes, C., Cox, A., Davies, L., Brown, C., 2005. Toward the $1000 human genome. http://dx.doi.org/10.1517/14622416.6.4.373 6, 373–382. https://doi.org/10.1517/14622416.6.4.373

Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., Brown, C.G., Hall, K.P., Evers, D.J., Barnes, C.L., Bignell, H.R., Boutell, J.M., Bryant, J., Carter, R.J., Keira Cheetham, R., Cox, A.J., Ellis, D.J., Flatbush, M.R., Gormley, N.A., Humphray, S.J., Irving, L.J., Karbelashvili, M.S., Kirk, S.M., Li, H., Liu, X., Maisinger, K.S., Murray, L.J., Obradovic, B., Ost, T., Parkinson, M.L., Pratt, M.R., Rasolonjatovo, I.M.J., Reed, M.T., Rigatti, R., Rodighiero, C., Ross, M.T., Sabot, A., Sankar, S. v., Scally, A., Schroth, G.P., Smith, M.E., Smith, V.P., Spiridou, A., Torrance, P.E., Tzonev, S.S., Vermaas, E.H., Walter, K., Wu, X., Zhang, L., Alam, M.D., Anastasi, C., Aniebo, I.C., Bailey, D.M.D., Bancarz, I.R., Banerjee, S., Barbour, S.G., Baybayan, P.A., Benoit, V.A., Benson, K.F., Bevis, C., Black, P.J., Boodhun, A., Brennan, J.S., Bridgham, J.A., Brown, R.C., Brown, A.A., Buermann, D.H., Bundu, A.A., Burrows, J.C., Carter, N.P., Castillo, N., Catenazzi, M.C.E., Chang, S., Neil Cooley, R., Crake, N.R., Dada, O.O., Diakoumakos, K.D., Dominguez-Fernandez, B., Earnshaw, D.J., Egbujor, U.C., Elmore, D.W., Etchin, S.S., Ewan, M.R., Fedurco, M., Fraser, L.J., Fuentes Fajardo, K. v., Scott Furey, W., George, D., Gietzen, K.J., Goddard, C.P., Golda, G.S., Granieri, P.A., Green, D.E., Gustafson, D.L., Hansen, N.F., Harnish, K., Haudenschild, C.D., Heyer, N.I., Hims, M.M., Ho, J.T., Horgan, A.M., Hoschler, K., Hurwitz, S., Ivanov, D. v., Johnson, M.Q., James, T., Huw Jones, T.A., Kang, G.D., Kerelska, T.H., Kersey, A.D., Khrebtukova, I., Kindwall, A.P., Kingsbury, Z., Kokko-Gonzales, P.I., Kumar, A., Laurent, M.A., Lawley, C.T., Lee, S.E., Lee, X., Liao, A.K., Loch, J.A., Lok, M., Luo, S., Mammen, R.M., Martin, J.W., McCauley, P.G., McNitt, P., Mehta, P., Moon, K.W., Mullens, J.W., Newington, T., Ning, Z., Ling Ng, B., Novo, S.M., O’Neill, M.J., Osborne, M.A., Osnowski, A., Ostadan, O., Paraschos, L.L., Pickering, L., Pike, Andrew C., Pike, Alger C., Chris Pinkard, D., Pliskin, D.P., Podhasky, J., Quijano, V.J., Raczy, C., Rae, V.H., Rawlings, S.R., Chiva Rodriguez, A., Roe, P.M., Rogers, John, Rogert Bacigalupo, M.C., Romanov, N., Romieu, A., Roth, R.K., Rourke, N.J., Ruediger, S.T., Rusman, E., Sanches-Kuiper, R.M., Schenker, M.R., Seoane, J.M., Shaw, R.J., Shiver, M.K., Short, S.W., Sizto, N.L., Sluis, J.P., Smith, M.A., Ernest Sohna Sohna, J., Spence, E.J., Stevens, K., Sutton, N., Szajkowski, L., Tregidgo, C.L., Turcatti, G., Vandevondele, S., Verhovsky, Y., Virk, S.M., Wakelin, S., Walcott, G.C., Wang, J., Worsley, G.J., Yan, J., Yau, L., Zuerlein, M., Rogers, Jane, Mullikin, J.C., Hurles, M.E., McCooke, N.J., West, J.S., Oaks, F.L., Lundberg, P.L., Klenerman, D., Durbin, R., Smith, A.J., 2008. Accurate whole human genome sequencing using reversible terminator chemistry. Nature 2008 456:7218 456, 53–59. https://doi.org/10.1038/nature07517

Bernt, M., Braband, A., Schierwater, B., Stadler, P.F., 2013a. Genetic aspects of mitochondrial genome evolution. Mol Phylogenet Evol 69, 328–338. https://doi.org/10.1016/J.YMPEV.2012.10.020

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., Pütz, J., Middendorf, M., Stadler, P.F., 2013b. MITOS: Improved de novo metazoan mitochondrial genome annotation. Mol Phylogenet Evol 69, 313–319. https://doi.org/10.1016/J.YMPEV.2012.08.023

Bertorelle, G., Raffini, F., Bosse, M., Bortoluzzi, C., Iannucci, A., Trucchi, E., Morales, H.E., van Oosterhout, C., 2022. Genetic load: genomic estimates and applications in non-model animals. Nature Reviews Genetics 2022 23:8 23, 492–503. https://doi.org/10.1038/s41576-022-00448-x

Bertucci, A., Pierron, F., Thébault, J., Klopp, C., Bellec, J., Gonzalez, P., Baudrimont, M., 2017. Transcriptomic responses of the endangered freshwater mussel *Margaritifera margaritifera* to trace metal contamination in the Dronne River, France. Environmental Science and Pollution Research 24, 27145–27159. https://doi.org/10.1007/S11356-017-0294-6/TABLES/6

Bespalaya, Yu. v., Bolotov, I.N., Makhrov, A.A., Vikhrev, I. v., 2012. Historical geography of pearl fishing in rivers of the Southern White Sea Region (Arkhangelsk Oblast). Regional Research of Russia 2, 172–181. https://doi.org/10.1134/S2079970512020025

Besser, J., Carleton, H.A., Gerner-Smidt, P., Lindsey, R.L., Trees, E., 2018. Next-generation sequencing technologies and their application to the study and control of bacterial infections. Clinical Microbiology and Infection 24, 335–341. https://doi.org/10.1016/J.CMI.2017.10.013

Bettinazzi, S., Nadarajah, S., Dalpé, A., Milani, L., Blier, P.U., Breton, S., 2020. Linking paternally inherited mtDNA variants and sperm performance. Philosophical Transactions of the Royal Society B 375. https://doi.org/10.1098/RSTB.2019.0177

Bettinazzi, S., Plazzi, F., Passamonti, M., 2016. The Complete Female- and Male-Transmitted Mitochondrial Genome of *Meretrix lamarckii*. PLoS One 11, e0153631. https://doi.org/10.1371/JOURNAL.PONE.0153631

Bettinazzi, S., Rodríguez, E., Milani, L., Blier, P.U., Breton, S., 2019. Metabolic remodelling associated with mtDNA: insights into the adaptive value of doubly uniparental inheritance of mitochondria. Proceedings of the Royal Society B 286. https://doi.org/10.1098/RSPB.2018.2708

Bieler, R., Mikkelsen, P.M., Collins, T.M., Glover, E.A., González, V.L., Graf, D.L., Harper, E.M., Healy, J., Kawauchi, G.Y., Sharma, P.P., Staubach, S., Strong, E.E., Taylor, J.D., Tëmkin, I., Zardus, J.D., Clark, S., Guzmán, A., McIntyre, E., Sharp, P., Giribet, G., Bieler, R., Mikkelsen, P.M., Collins, T.M., Glover, E.A., González, V.L., Graf, D.L., Harper, E.M., Healy, J., Kawauchi, G.Y., Sharma, P.P., Staubach, S., Strong, E.E., Taylor, J.D., Tëmkin, I., Zardus, J.D., Clark, S., Guzmán, A., McIntyre, E., Sharp, P., Giribet, G., 2014. Investigating the Bivalve Tree of Life – an exemplar-based approach combining molecular and novel morphological characters. Invertebr Syst 28, 32–115. https://doi.org/10.1071/IS13010

Biot-Pelletier, D., Bettinazzi, S., Gagnon-Arsenault, I., Dubé, A.K., Bédard, C., Nguyen, T.H.M., Fiumera, H.L., Breton, S., Landry, C.R., 2022. Evolutionary trajectories are contingent on mitonuclear interactions. bioRxiv 2022.09.11.507487. https://doi.org/10.1101/2022.09.11.507487

Birky, C.W., 1995. Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. Proceedings of the National Academy of Sciences 92, 11331–11338. https://doi.org/10.1073/PNAS.92.25.11331

Birky, J., 2003. The Inheritance of Genes in Mitochondria and Chloroplasts: Laws, Mechanisms, and Models. https://doi.org/10.1146/annurev.genet.35.102401.090231 35, 125–148. https://doi.org/10.1146/ANNUREV.GENET.35.102401.090231

Blattner, F.R., Plunkett, G., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B., Shao, Y., 1997. The Complete Genome Sequence of *Escherichia coli* K-12. Science (1979) 277, 1453–1462. https://doi.org/10.1126/science.277.5331.1453

Bleidorn, C., Podsiadlowski, L., Bartolomaeus, T., 2006. The complete mitochondrial genome of the orbiniid polychaete *Orbinia latreillii* (Annelida, Orbiniidae) – A novel gene order for Annelida and implications for annelid phylogeny. Gene 370, 96–103. https://doi.org/10.1016/J.GENE.2005.11.018

Blevins, E., Jepsen, S., Brim Box, J. & Nez, D. 2016. *Margaritifera falcata* (errata version published in 2017). The IUCN Red List of Threatened Species 2016: e.T91109639A114128748. https://dx.doi.org/10.2305/IUCN.UK.2016-3.RLTS.T91109639A91109660.en. Accessed on 05 November 2022.

Blier, P.U., Dufresne, F., Burton, R.S., 2001. Natural selection and the evolution of mtDNA-encoded peptides: evidence for intergenomic co-adaptation. Trends in Genetics 17, 400–406. https://doi.org/10.1016/S0168-9525(01)02338-1

Boeckmann, B., 2003. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res 31, 365–370. https://doi.org/10.1093/nar/gkg095

Bogan, A.E., 2008. Global diversity of freshwater mussels (Mollusca, Bivalvia) in freshwater. Hydrobiologia. https://doi.org/10.1007/s10750-007-9011-7

Böhm, M., Dewhurst-Richman, N.I., Seddon, M., Ledger, S.E.H., Albrecht, C., Allen, D., Bogan, A.E., Cordeiro, J., Cummings, K.S., Cuttelod, A., Darrigran, G., Darwall, W., Fehér, Z., Gibson, C., Graf, D.L., Köhler, F., Lopes-Lima, M., Pastorino, G., Perez, K.E., Smith, K., van Damme, D., Vinarski, M. v., von Proschwitz, T., von Rintelen, T., Aldridge, D.C., Aravind, N.A., Budha, P.B., Clavijo, C., van Tu, D., Gargominy, O., Ghamizi, M., Haase, M., Hilton-Taylor, C., Johnson, P.D., Kebapçı, Ü., Lajtner, J., Lange, C.N., Lepitzki, D.A.W., Martínez-Ortí, A., Moorkens, E.A., Neubert, E., Pollock, C.M., Prié, V., Radea, C., Ramirez, R., Ramos, M.A., Santos, S.B., Slapnik, R., Son, M.O., Stensgaard, A.S., Collen, B., 2021. The conservation status of the world’s freshwater molluscs. Hydrobiologia 848, 3231–3254. https://doi.org/10.1007/S10750-020-04385-W/FIGURES/4

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bolotov, I.N., Vikhrev, I. v., Bespalaya, Y. v., Gofarov, M.Y., Kondakov, A. v., Konopleva, E.S., Bolotov, N.N., Lyubas, A.A., 2016. Multi-locus fossil-calibrated phylogeny, biogeography and a subgeneric revision of the Margaritiferidae (Mollusca: Bivalvia: Unionoida). Mol Phylogenet Evol 103, 104–121. https://doi.org/10.1016/J.YMPEV.2016.07.020

Boore, J.L., 2006a. Requirements and Standards for Organelle Genome Databases. https://home.liebertpub.com/omi 10, 119–126. https://doi.org/10.1089/OMI.2006.10.119

Boore, J.L., 2006b. The complete sequence of the mitochondrial genome of *Nautilus macromphalus* (Mollusca: Cephalopoda). BMC Genomics 7, 1–13. https://doi.org/10.1186/1471-2164-7-182/FIGURES/4

Boore, J.L., 2000. Comparative genomics: empirical and analytical approaches to gene order dynamics, map alignment and the evolution of gene families, vol. 1., Computational biology series. The duplication/random loss model for gene rearrangement exemplified by mitochondrial genomes of deuterostome animals 1, 133–147.

Boore, J.L., 1999. Animal mitochondrial genomes. Nucleic Acids Res 27, 1767–1780. https://doi.org/10.1093/NAR/27.8.1767

Boore, J.L., Brown, W.M., 1998. Big trees from little genomes: mitochondrial gene order as a phylogenetic tool. Curr Opin Genet Dev 8, 668–674. https://doi.org/10.1016/S0959-437X(98)80035-X

Boore, J.L., Brown, W.M., 1994. Complete DNA sequence of the mitochondrial genome of the black chiton, *Katharina tunicata*. Genetics 138, 423–443. https://doi.org/10.1093/GENETICS/138.2.423

Boore, J.L., Medina, M., Rosenberg, L.A., 2004. Complete Sequences of the Highly Rearranged Molluscan Mitochondrial Genomes of the Scaphopod *Graptacme eborea* and the Bivalve *Mytilus edulis*. Mol Biol Evol 21, 1492–1503. https://doi.org/10.1093/MOLBEV/MSH090

Bouza, C., Castro, J., Martínez, P., Amaro, R., Fernández, C., Ondina, P., Outeiro, A., San Miguel, E., 2007. Threatened freshwater pearl mussel *Margaritifera margaritifera* L. in NW Spain: low and very structured genetic variation in southern peripheral populations assessed using microsatellite markers. Conservation Genetics 8, 937–948. https://doi.org/10.1007/s10592-006-9248-0

Boyle, P., Rodhouse, P., 2008. Cephalopods: ecology and fisheries. John Wiley & Sons.

Bradnam, K.R., Fass, J.N., Alexandrov, A., Baranay, P., Bechner, M., Birol, I., Boisvert, S., Chapman, J.A., Chapuis, G., Chikhi, R., Chitsaz, H., Chou, W.-C., Corbeil, J., del Fabbro, C., Docking, T.R., Durbin, R., Earl, D., Emrich, S., Fedotov, P., Fonseca, N.A., Ganapathy, G., Gibbs, R.A., Gnerre, S., Godzaridis, É., Goldstein, S., Haimel, M., Hall, G., Haussler, D., Hiatt, J.B., Ho, I.Y., Howard, J., Hunt, M., Jackman, S.D., Jaffe, D.B., Jarvis, E.D., Jiang, H., Kazakov, S., Kersey, P.J., Kitzman, J.O., Knight, J.R., Koren, S., Lam, T.-W., Lavenier, D., Laviolette, F., Li, Y., Li, Z., Liu, B., Liu, Y., Luo, R., MacCallum, I., MacManes, M.D., Maillet, N., Melnikov, S., Naquin, D., Ning, Z., Otto, T.D., Paten, B., Paulo, O.S., Phillippy, A.M., Pina-Martins, F., Place, M., Przybylski, D., Qin, X., Qu, C., Ribeiro, F.J., Richards, S., Rokhsar, D.S., Ruby, J.G., Scalabrin, S., Schatz, M.C., Schwartz, D.C., Sergushichev, A., Sharpe, T., Shaw, T.I., Shendure, J., Shi, Y., Simpson, J.T., Song, H., Tsarev, F., Vezzi, F., Vicedomini, R., Vieira, B.M., Wang, J., Worley, K.C., Yin, S., Yiu, S.-M., Yuan, J., Zhang, G., Zhang, H., Zhou, S., Korf, I.F., 2013. Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. Gigascience 2, 10. https://doi.org/10.1186/2047-217X-2-10

Brauer, A., Kurz, A., Stockwell, T., Baden-Tillson, H., Heidler, J., Wittig, I., Kauferstein, S., Mebs, D., Stöcklin, R., Remm, M., 2012. The Mitochondrial Genome of the Venomous Cone Snail *Conus consors*. PLoS One 7, e51528. https://doi.org/10.1371/JOURNAL.PONE.0051528

Breinholt, J.W., Earl, C., Lemmon, A.R., Lemmon, E.M., Xiao, L., Kawahara, A.Y., 2018. Resolving Relationships among the Megadiverse Butterflies and Moths with a Novel Pipeline for Anchored Phylogenomics. Syst Biol 67, 78–93. https://doi.org/10.1093/sysbio/syx048

Breton, S., Beaupré, H.D., Stewart, D.T., Hoeh, W.R., Blier, P.U., 2007. The unusual system of doubly uniparental inheritance of mtDNA: isn't one enough? Trends in Genetics 23, 465–474. https://doi.org/10.1016/j.tig.2007.05.011

Breton, S., Beaupré, H.D., Stewart, D.T., Piontkivska, H., Karmakar, M., Bogan, A.E., Blier, P.U., Hoeh, W.R., 2009. Comparative Mitochondrial Genomics of Freshwater Mussels (Bivalvia: Unionoida) With Doubly Uniparental Inheritance of mtDNA: Gender-Specific Open Reading Frames and Putative Origins of Replication. Genetics 183, 1575–1589. https://doi.org/10.1534/GENETICS.109.110700

Breton, S., Bouvet, K., Auclair, G., Ghazal, S., Sietman, B.E., Johnson, N., Bettinazzi, S., Stewart, D.T., Guerra, D., 2017. The extremely divergent maternally- and paternally-transmitted mitochondrial genomes are co-expressed in somatic tissues of two freshwater mussel species with doubly uniparental inheritance of mtDNA. PLoS One 12, e0183529. https://doi.org/10.1371/JOURNAL.PONE.0183529

Breton, S., Burger, G., Stewart, D.T., Blier, P.U., 2006. Comparative Analysis of Gender-Associated Complete Mitochondrial Genomes in Marine Mussels (Mytilus spp.). Genetics 172, 1107–1119. https://doi.org/10.1534/GENETICS.105.047159

Breton, S., Capt, C., Guerra, D., Stewart, D., 2018. Sex-Determining Mechanisms in Bivalves, in: Transitions Between Sexual Systems. Springer International Publishing, Cham, pp. 165–192. https://doi.org/10.1007/978-3-319-94139-4_6

Breton, S., Ghiselli, F., Passamonti, M., Milani, L., Stewart, D.T., Hoeh, W.R., 2011a. Evidence for a Fourteenth mtDNA-Encoded Protein in the Female-Transmitted mtDNA of Marine Mussels (Bivalvia: Mytilidae). PLoS One 6, e19365. https://doi.org/10.1371/JOURNAL.PONE.0019365

Breton, S., Milani, L., Ghiselli, F., Guerra, D., Stewart, D.T., Passamonti, M., 2014. A resourceful genome: updating the functional repertoire and evolutionary role of animal mitochondrial DNAs. Trends in Genetics 30, 555–564. https://doi.org/10.1016/J.TIG.2014.09.002

Breton, S., Stewart, D.T., 2015. Atypical mitochondrial inheritance patterns in eukaryotes. https://doi.org/10.1139/gen-2015-0090 58, 423–431. https://doi.org/10.1139/GEN-2015-0090

Breton, S., Stewart, D.T., Hoeh, W.R., 2010. Characterization of a mitochondrial ORF from the gender-associated mtDNAs of *Mytilus spp*. (Bivalvia: Mytilidae): Identification of the "missing" ATPase 8 gene. Mar Genomics 3, 11–18. https://doi.org/10.1016/J.MARGEN.2010.01.001

Breton, S., Stewart, D.T., Shepardson, S., Trdan, R.J., Bogan, A.E., Chapman, E.G., Ruminas, A.J., Piontkivska, H., Hoeh, W.R., 2011b. Novel Protein Genes in Animal mtDNA: A New Sex Determination System in Freshwater Mussels (Bivalvia: Unionoida)? Mol Biol Evol 28, 1645–1659. https://doi.org/10.1093/molbev/msq345

Brooke, N.M., Garcia-Fernàndez, J., Holland, P.W.H., 1998. The ParaHox gene cluster is an evolutionary sister of the Hox gene cluster. Nature 392, 920–922. https://doi.org/10.1038/31933

Brown, G.G., Gadaleta, G., Pepe, G., Saccone, C., Sbisà, E., 1986. Structural conservation and variation in the D-loop-containing region of vertebrate mitochondrial DNA. J Mol Biol 192, 503–511. https://doi.org/10.1016/0022-2836(86)90272-X

Brown, T.A., Cecconi, C., Tkachuk, A.N., Bustamante, C., Clayton, D.A., 2005. Replication of mitochondrial DNA occurs by strand displacement with alternative light-strand origins, not via a strand-coupled mechanism. Genes Dev 19, 2466–2476. https://doi.org/10.1101/GAD.1352105

Brůna, T., Hoff, K.J., Lomsadze, A., Stanke, M., Borodovsky, M., 2021. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. NAR Genom Bioinform 3, 1–11. https://doi.org/10.1093/NARGAB/LQAA108

Brusca, R.C., Brusca, G.J., Haver, N.J., 2003. Invertebrates, Second. ed. Sunderland, MA.

Buchfink, B., Xie, C., Huson, D.H., 2015. Fast and sensitive protein alignment using DIAMOND. Nat Methods 12, 59–60. https://doi.org/10.1038/nmeth.3176

Buddenhagen, C., Lemmon, A.R., Lemmon, E.M., Bruhl, J., Cappa, J., Clement, W.L., Donoghue, M.J., Edwards, E.J., Hipp, A.L., Kortyna, M., Mitchell, N., Moore, A., Prychid, C.J., Segovia-Salcedo, M.C., Simmons, M.P., Soltis, P.S., Wanke, S., Mast, A., 2016. Anchored Phylogenomics of Angiosperms I: Assessing the Robustness of Phylogenetic Estimates. bioRxiv 8, 086298. https://doi.org/10.1101/086298

Burzyński, A., Zbawicka, M., Skibinski, D.O.F., Wenne, R., 2003. Evidence for Recombination of mtDNA in the Marine Mussel *Mytilus trossulus* from the Baltic. Mol Biol Evol 20, 388–392. https://doi.org/10.1093/MOLBEV/MSG058

Bushnell, B., Rood, J., 2018. BBTools. BBMap.

Butler, P.G., Wanamaker, A.D., Scourse, J.D., Richardson, C.A., Reynolds, D.J., 2013. Variability of marine climate on the North Icelandic Shelf in a 1357-year proxy archive based on growth increments in the bivalve *Arctica islandica*. Palaeogeogr Palaeoclimatol Palaeoecol 373, 141–151. https://doi.org/10.1016/j.palaeo.2012.01.016

C. elegans Sequencing Consortium, 1998. Genome Sequence of the Nematode *C. elegans*: A Platform for Investigating Biology. Science (1979) 282, 2012–2018. https://doi.org/10.1126/science.282.5396.2012

Cai, H., Li, Q., Fang, X., Li, J., Curtis, N.E., Altenburger, A., Shibata, T., Feng, M., Maeda, T., Schwartz, J.A., Shigenobu, S., Lundholm, N., Nishiyama, T., Yang, H., Hasebe, M., Li, S., Pierce, S.K., Wang, J., 2019. Data descriptor: A draft genome assembly of the solar-powered sea slug *Elysia chlorotica*. Sci Data 6, 190022. https://doi.org/10.1038/sdata.2019.22

Cai, J., Zhao, R., Jiang, H., Wang, W., 2008. De Novo Origination of a New Protein-Coding Gene in *Saccharomyces cerevisiae*. Genetics 179, 487–496. https://doi.org/10.1534/GENETICS.107.084491

Calcino, A., Baranyi, C., 1, , Wanninger, A., 2020. Heteroplasmy and repeat expansion in the plant-like mitochondrial genome of a bivalve mollusc. bioRxiv 2020.09.23.310516. https://doi.org/10.1101/2020.09.23.310516

Calcino, A.D., de Oliveira, A.L., Simakov, O., Schwaha, T., Zieger, E., Wollesen, T., Wanninger, A., 2019. The quagga mussel genome and the evolution of freshwater tolerance. DNA Research 26, 411–422. https://doi.org/10.1093/dnares/dsz019

Calcino, A.D., Kenny, N.J., Gerdol, M., 2021. Single individual structural variant detection uncovers widespread hemizygosity in molluscs. Philosophical Transactions of the Royal Society B 376. https://doi.org/10.1098/RSTB.2020.0153

Calisi, R.M., MacManes, M.D., 2015. RNAseq-ing a more integrative understanding of animal behavior. Curr Opin Behav Sci 6, 65–68. https://doi.org/10.1016/J.COBEHA.2015.09.007

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: Architecture and applications. BMC Bioinformatics 10, 1–9. https://doi.org/10.1186/1471-2105-10-421/FIGURES/4

Campagna, L., Toews, D.P.L., 2022. The genomics of adaptation in birds. Current Biology 32, R1173–R1186. https://doi.org/10.1016/J.CUB.2022.07.076

Camus, M.F., Clancy, D.J., Dowling, D.K., 2012. Mitochondria, maternal inheritance, and male aging. Current Biology 22, 1717–1721. https://doi.org/10.1016/j.cub.2012.07.018

Cantatore, P., Gadaleta, M.N., Roberti, M., Saccone, C., Wilson, A.C., 1987. Duplication and remoulding of tRNA genes during the evolutionary rearrangement of mitochondrial genomes. Nature 1987 329:6142 329, 853–855. https://doi.org/10.1038/329853a0

Cao, L., Ort, B.S., Mizi, A., Pogson, G., Kenchington, E., Zouros, E., Rodakis, G.C., 2009. The Control Region of Maternally and Paternally Inherited Mitochondrial Genomes of Three Species of the Sea Mussel Genus *Mytilus*. Genetics 181, 1045–1056. https://doi.org/10.1534/GENETICS.108.093229

Cao, L., Pollock, D.D., Stewart, D.T., Piontkivska, H., Karmakar, M., Bogan, A.E., Blier, P.U., Hoeh, W.R., 2004. Differential Segregation Patterns of Sperm Mitochondria in Embryos of the Blue Mussel (*Mytilus edulis*). Genetics 166, 883–894. https://doi.org/10.1534/genetics.166.2.883

Capella-Gutiérrez, S., Silla-Martínez, J.M., Gabaldón, T., 2009. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972–1973. https://doi.org/10.1093/bioinformatics/btp348

Capt, C., Bouvet, K., Guerra, D., Robicheau, B.M., Stewart, D.T., Pante, E., Breton, S., 2020. Unorthodox features in two venerid bivalves with doubly uniparental inheritance of mitochondria. Scientific Reports 2020 10:1 10, 1–13. https://doi.org/10.1038/s41598-020-57975-y

Capt, C., Renaut, S., Ghiselli, F., Milani, L., Johnson, N.A., Sietman, B.E., Stewart, D.T., Breton, S., 2018a. Deciphering the Link between Doubly Uniparental Inheritance of mtDNA and Sex Determination in Bivalves: Clues from Comparative Transcriptomics. Genome Biol Evol 10, 577–590. https://doi.org/10.1093/GBE/EVY019

Capt, C., Renaut, S., Ghiselli, F., Milani, L., Johnson, N.A., Sietman, B.E., Stewart, D.T., Breton, S., 2018b. Deciphering the Link between Doubly Uniparental Inheritance of mtDNA and Sex Determination in Bivalves: Clues from Comparative Transcriptomics. Genome Biol Evol 10, 577–590. https://doi.org/10.1093/GBE/EVY019

Capt, C., Renaut, S., Stewart, D.T., Johnson, N.A., Breton, S., 2019. Putative Mitochondrial Sex Determination in the Bivalvia: Insights From a Hybrid Transcriptome Assembly in Freshwater Mussels. Front Genet 10, 840. https://doi.org/10.3389/FGENE.2019.00840/BIBTEX

Cardoso, P., Erwin, T.L., Borges, P.A.V., New, T.R., 2011. The seven impediments in invertebrate conservation and how to overcome them. Biol Conserv 144, 2647–2655. https://doi.org/10.1016/j.biocon.2011.07.024

Castro, L.F.C., Holland, P.W.H., 2003. Chromosomal mapping of ANTP class homeobox genes in amphioxus: Piecing together ancestral genomes. Evol Dev 5, 459–465. https://doi.org/10.1046/j.1525-142X.2003.03052.x

Chakrabarti, R., Walker, J.M., Chapman, E.G., Shepardson, S.P., Trdan, R.J., Curole, J.P., Watters, G.T., Stewart, D.T., Vijayaraghavan, S., Hoeh, W.R., 2007. Reproductive function for a C-terminus extended, male-transmitted cytochrome c oxidase subunit II protein expressed in both spermatozoa and eggs. FEBS Lett 581, 5213–5219. https://doi.org/10.1016/J.FEBSLET.2007.10.006

Chan, P.P., Lowe, T.M., 2019. tRNAscan-SE: Searching for tRNA genes in genomic sequences. Methods in Molecular Biology 1962, 1–14. https://doi.org/10.1007/978-1-4939-9173-0_1/COVER

Chang, Z., Li, G., Liu, J., Zhang, Y., Ashby, C., Liu, D., Cramer, C.L., Huang, X., 2015. Bridger: A new framework for de novo transcriptome assembly using RNA-seq data. Genome Biol 16, 1–10. https://doi.org/10.1186/s13059-015-0596-2

Chapdelaine, V., Bettinazzi, S., Breton, S., Angers, B., 2020. Effects of mitonuclear combination and thermal acclimation on the energetic phenotype. J Exp Zool A Ecol Integr Physiol 333, 264–270. https://doi.org/10.1002/JEZ.2355

Chapman, J.A., Ho, I., Sunkara, S., Luo, S., Schroth, G.P., Rokhsar, D.S., 2011. Meraculous: De novo genome assembly with short paired-end reads. PLoS One 6, e23501. https://doi.org/10.1371/journal.pone.0023501

Chase, E.E., Robicheau, B.M., Veinot, S., Breton, S., Stewart, D.T., 2018. The complete mitochondrial genome of the hermaphroditic freshwater mussel *Anodonta cygnea* (Bivalvia: Unionidae): In silico analyses of sex-specific ORFs across order Unionoida. BMC Genomics 19, 1–15. https://doi.org/10.1186/S12864-018-4583-3/FIGURES/5

Chen, F., Mackey, A.J., Vermunt, J.K., Roos, D.S., 2007. Assessing Performance of Orthology Detection Strategies Applied to Eukaryotic Genomes. PLoS One 2, e383. https://doi.org/10.1371/JOURNAL.PONE.0000383

Chen, X., Bai, Z., Li, J., 2019. The Mantle Exosome and MicroRNAs of *Hyriopsis cumingii* Involved in Nacre Color Formation. Marine Biotechnology 21, 634–642. https://doi.org/10.1007/S10126-019-09908-8/FIGURES/5

Cheng, R., Zheng, X., Ma, Y., Li, Q., 2013. The Complete Mitochondrial Genomes of Two Octopods *Cistopus chinensis* and *Cistopus taiwanicus*: Revealing the Phylogenetic Position of the Genus *Cistopus* within the Order Octopoda. PLoS One 8, e84216. https://doi.org/10.1371/JOURNAL.PONE.0084216

Chong, J.P., Roe, K.J., 2018. A comparison of genetic diversity and population structure of the endangered scaleshell mussel (*Leptodea leptodon*), the fragile papershell (*Leptodea fragilis*) and their host-fish the freshwater drum (*Aplodinotus grunniens*). Conservation Genetics 19, 425–437. https://doi.org/10.1007/S10592-017-1015-X/TABLES/6

Clark, M.S., Peck, L.S., Arivalagan, J., Backeljau, T., Berland, S., Cardoso, J.C.R., Caurcel, C., Chapelle, G., de Noia, M., Dupont, S., Gharbi, K., Hoffman, J.I., Last, K.S., Marie, A., Melzner, F., Michalek, K., Morris, J., Power, D.M., Ramesh, K., Sanders, T., Sillanpää, K., Sleight, V.A., Stewart-Sinclair, P.J., Sundell, K., Telesca, L., Vendrami, D.L.J., Ventura, A., Wilding, T.A., Yarra, T., Harper, E.M., 2020. Deciphering mollusc shell production: the roles of genetic mechanisms through to ecology, aquaculture and biomimetics. Biological Reviews 95, 1812–1837. https://doi.org/10.1111/BRV.12640

Clary, D.O., Wolstenholme, D.R., 1985. The mitochondrial DNA molecule of *Drosophila yakuba*: Nucleotide sequence, gene organization, and genetic code. Journal of Molecular Evolution 1985 22:3 22, 252–271. https://doi.org/10.1007/BF02099755

Clayton, D.A., 1982. Replication of animal mitochondrial DNA. Cell 28, 693–705. https://doi.org/10.1016/0092-8674(82)90049-6

Cogswell, A.T., Kenchington, E.L.R., Zouros, E., 2011. Segregation of sperm mitochondria in two- and four-cell embryos of the blue mussel *Mytilus edulis*: implications for the mechanism of doubly uniparental inheritance of mitochondrial DNA. https://doi.org/10.1139/g06-036 49, 799–807. https://doi.org/10.1139/G06-036

Cohen, S.N., Chang, A.C.Y., Boyer, H.W., Helling, R.B., 1973. Construction of Biologically Functional Bacterial Plasmids *In Vitro*. Proceedings of the National Academy of Sciences 70, 3240–3244. https://doi.org/10.1073/pnas.70.11.3240

Combosch, D.J., Collins, T.M., Glover, E.A., Graf, D.L., Harper, E.M., Healy, J.M., Kawauchi, G.Y., Lemer, S., McIntyre, E., Strong, E.E., Taylor, J.D., Zardus, J.D., Mikkelsen, P.M., Giribet, G., Bieler, R., 2017. A family-level Tree of Life for bivalves based on a Sanger-sequencing approach. Mol Phylogenet Evol 107, 191–208. https://doi.org/10.1016/J.YMPEV.2016.11.003

Cong, H., Lei, Y., Kong, L., 2020. The mitochondrial genome of the toothed top shell snail *Monodonta labio* (Gastropoda: Trochidae): the first complete sequence in the subfamily monodontinae. http://www.tandfonline.com/action/authorSubmission?journalCode=tmdn20&page=instructions 5, 621–622. https://doi.org/10.1080/23802359.2019.1711221

Cornman, R.S., Robertson, L.S., Galbraith, H., Blakeslee, C., 2014. Transcriptomic Analysis of the Mussel *Elliptio complanata* Identifies Candidate Stress-Response Genes and an Abundance of Novel or Noncoding Transcripts. PLoS One 9, e112420. https://doi.org/10.1371/JOURNAL.PONE.0112420

Cowie, R.H., 2009. Apple snails (Ampullariidae) as agricultural pests: their biology, impacts and management., in: Barker, G.M. (Ed.), Molluscs as Crop Pests. CABI, Wallingford, pp. 145–192. https://doi.org/10.1079/9780851993201.0145

Craig Venter, J., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., Gocayne, J.D., Amanatides, P., Ballew, R.M., Huson, D.H., Wortman, J.R., Zhang, Q., Kodira, C.D., Zheng, X.H., Chen, L., Skupski, M., Subramanian, G., Thomas, P.D., Zhang, J., Gabor Miklos, G.L., Nelson, C., Broder, S., Clark, A.G., Nadeau, J., McKusick, V.A., Zinder, N., Levine, A.J., Roberts, R.J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A.E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T.J., Higgins, M.E., Ji, R.R., Ke, Z., Ketchum, K.A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. v., Milshina, N., Moore, H.M., Naik, A.K., Narayan, V.A., Neelam, B., Nusskern, D., Rusch, D.B., Salzberg, S., Shao, W., Shue, B., Sun, J., Yuan Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M.H., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S.C., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Lai Cheng, M., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y.H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Ni Tint, N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J.F., Guigo, R., Campbell, M.J., Sjolander, K. v., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y.H., Coyne, M., Dahlke, C., Deslattes Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A., Zhu, X., 2001. The sequence of the human genome. Science (1979) 291, 1304–1351. https://doi.org/10.1126/SCIENCE.1058040/SUPPL_FILE/C18_SCIENCE.PDF

Cunha, T.J., Giribet, G., 2019. A congruent topology for deep gastropod relationships. Proceedings of the Royal Society B: Biological Sciences 286, 20182776. https://doi.org/10.1098/rspb.2018.2776

Curole, J.P., Kocher, T.D., 2005. Evolution of a unique mitotype-specific protein-coding extension of the cytochrome c oxidase 5 gene in freshwater mussels (Bivalvia: Unionoida). J Mol Evol 61, 381–389. https://doi.org/10.1007/S00239-004-0192-7/TABLES/3

Curole, J.P., Kocher, T.D., 2002. Ancient Sex-Specific Extension of the Cytochrome c Oxidase II Gene in Bivalves and the Fidelity of Doubly-Uniparental Inheritance. Mol Biol Evol 19, 1323–1328. https://doi.org/10.1093/OXFORDJOURNALS.MOLBEV.A004193

Cuttelod, A., Seddon, M., Neubert, E., 2011. European Red List of Non-marine Molluscs. Luxembourg: Publications Office of the European Union. https://doi.org/10.2779/84538

da Fonseca, R.R., Couto, A., Machado, A.M., Brejova, B., Albertin, C.B., Silva, F., Gardner, P., Baril, T., Hayward, A., Campos, A., Ribeiro, Â.M., Barrio-Hernandez, I., Hoving, H.J., Tafur-Jimenez, R., Chu, C., Frazão, B., Petersen, B., Peñaloza, F., Musacchia, F., Alexander, G.C., Osório, H., Winkelmann, I., Simakov, O., Rasmussen, S., Rahman, M.Z., Pisani, D., Vinther, J., Jarvis, E., Zhang, G., Strugnell, J.M., Castro, L.F.C., Fedrigo, O., Patricio, M., Li, Q., Rocha, S., Antunes, A., Wu, Y., Ma, B., Sanges, R., Vinar, T., Blagoev, B., Sicheritz-Ponten, T., Nielsen, R., Gilbert, M.T.P., 2020. A draft genome sequence of the elusive giant squid, *Architeuthis dux*. Gigascience 9. https://doi.org/10.1093/gigascience/giz152

da Fonseca, R.R., Johnson, W.E., O’Brien, S.J., Ramos, M.J., Antunes, A., 2008. The adaptive evolution of the mammalian mitochondrial genome. BMC Genomics 9, 119. https://doi.org/10.1186/1471-2164-9-119

Dainat, J., Hereñú, D., Pucholt, P., 2020. AGAT: Another Gff Analysis Toolkit to handle annotations in any GTF/GFF format. https://doi.org/10.5281/zenodo.4205393

Danic-Tchaleu, G., Heurtebise, S., Morga, B., Lapègue, S., 2011. Complete mitochondrial DNA sequence of the European flat oyster *Ostrea edulis* confirms Ostreidae classification. BMC Res Notes 4, 1–10. https://doi.org/10.1186/1756-0500-4-400/TABLES/2

Danovaro, R., Dell'Anno, A., Pusceddu, A., Gambi, C., Heiner, I., Møbjerg Kristensen, R., 2010. The first metazoa living in permanently anoxic conditions. BMC Biol 8, 1–10. https://doi.org/10.1186/1741-7007-8-30/COMMENTS

Darwin, C., 1859. On the origin of species by means of natural selection: or, the preservation of favored races in the struggle for life. Appleton, London.

David, K.T., Wilson, A.E., Halanych, K.M., 2019. Sequencing Disparity in the Genomic Era. Mol Biol Evol 36, 1624–1627. https://doi.org/10.1093/molbev/msz117

Davidson, N.M., Oshlack, A., 2014. Corset: Enabling differential gene expression analysis for de novo assembled transcriptomes. Genome Biol 15, 1–14. https://doi.org/10.1186/S13059-014-0410-6/FIGURES/6

Davies, N., Field, D., Amaral-Zettler, L., Clark, M.S., Deck, J., Drummond, A., Faith, D.P., Geller, J., Gilbert, J., Glöckner, F.O., Hirsch, P.R., Leong, J.-A., Meyer, C., Obst, M., Planes, S., Scholin, C., Vogler, A.P., Gates, R.D., Toonen, R., Berteaux-Lecellier, V., Barbier, M., Barker, K., Bertilsson, S., Bicak, M., Bietz, M.J., Bobe, J., Bodrossy, L., Borja, A., Coddington, J., Fuhrman, J., Gerdts, G., Gillespie, R., Goodwin, K., Hanson, P.C., Hero, J.-M., Hoekman, D., Jansson, J., Jeanthon, C., Kao, R., Klindworth, A., Knight, R., Kottmann, R., Koo, M.S., Kotoulas, G., Lowe, A.J., Marteinsson, V.T., Meyer, F., Morrison, N., Myrold, D.D., Pafilis, E., Parker, S., Parnell, J.J., Polymenakou, P.N., Ratnasingham, S., Roderick, G.K., Rodriguez-Ezpeleta, N., Schonrogge, K., Simon, N., Valette-Silver, N.J., Springer, Y.P., Stone, G.N., Stones-Havas, S., Sansone, S.-A., Thibault, K.M., Wecker, P., Wichels, A., Wooley, J.C., Yahara, T., Zingone, A., 2014. The founding charter of the Genomic Observatories Network. Gigascience 3, 2. https://doi.org/10.1186/2047-217X-3-2

Deamer, D., Akeson, M., Branton, D., 2016. Three decades of nanopore sequencing. Nature Biotechnology 2016 34:5 34, 518–524. https://doi.org/10.1038/nbt.3423

Dégletagne, C., Abele, D., Held, C., 2016. A Distinct Mitochondrial Genome with DUI-Like Inheritance in the Ocean Quahog Arctica islandica. Mol Biol Evol 33, 375–383. https://doi.org/10.1093/MOLBEV/MSV224

DeJong, R.J., Emery, A.M., Adema, C.M., 2004. THE MITOCHONDRIAL GENOME OF *BIOMPHALARIA GLABRATA* (GASTROPODA: BASOMMATOPHORA), INTERMEDIATE HOST OF SCHISTOSOMA MANSONI*. https://doi.org/10.1645/GE-284R 90, 991–997. https://doi.org/10.1645/GE-284R

Delsuc, F., Brinkmann, H., Philippe, H., 2005. Phylogenomics and the reconstruction of the tree of life. Nature Reviews Genetics 2005 6:5 6, 361–375. https://doi.org/10.1038/nrg1603

der Schalie, H., 1970. Hermaphroditism among North American freshwater mussels. Malacologia 10, 93–112.

Deremiens, L., Schwartz, L., Angers, A., Glémet, H., Angers, B., 2015. Interactions between nuclear genes and a foreign mitochondrial genome in the redbelly dace *Chrosomus eos*. Comp Biochem Physiol B Biochem Mol Biol 189, 80–86. https://doi.org/10.1016/J.CBPB.2015.08.002

DeSalle, R., Hadrys, H., 2017. Evolutionary Biology and Mitochondrial Genomics: 50 000 Mitochondrial DNA Genomes and Counting, in: ELS. Wiley, pp. 1–35. https://doi.org/10.1002/9780470015902.a0027270

Díaz, S., Settele, J., Brondízio, E.S., Ngo, H.T., Guèze, M., Agard, J., Arneth, A., Balvanera, P., Brauman, K., Butchart, S.H., Chan, K., Garibaldi, L., Ichii, K., Liu, J., Subramanian, S., Midgley, G., Miloslavich, P., Molnár, Z., Obura, D., Pfaff, A., Polasky, S., Purvis, A., Razzaque, J., Reyers, B., Chowdhury, R., Shin, Y., Visseren-Hamakers, I., Willis, K., Zayas, C., 2019. IPBES, 2019: Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services.

Dierckxsens, N., Mardulyn, P., Smits, G., 2016. NOVOPlasty: *de novo* assembly of organelle genomes from whole genome data. Nucleic Acids Res 45, gkw955. https://doi.org/10.1093/nar/gkw955

Ding, Y., Zhou, Q., Wang, W., 2012. Origins of New Genes and Evolution of Their Novel Functions. https://doi.org/10.1146/annurev-ecolsys-110411-160513 43, 345–363. https://doi.org/10.1146/ANNUREV-ECOLSYS-110411-160513

Diz, A.P., Dudley, E., Cogswell, A., Macdonald, B.W., Kenchington, E.L.R., Zouros, E., Skibinski, D.O.F., 2013. Proteomic Analysis of Eggs from *Mytilus edulis* Females

Differing in Mitochondrial DNA Transmission Mode. Molecular & Cellular Proteomics 12, 3068–3080. https://doi.org/10.1074/MCP.M113.031401

Dominguez Del Angel, V., Hjerde, E., Sterck, L., Capella-Gutierrez, S., Notredame, C., Vinnere Pettersson, O., Amselem, J., Bouri, L., Bocs, S., Klopp, C., Gibrat, J.-F., Vlasova, A., Leskosek, B.L., Soler, L., Binzer-Panchal, M., Lantz, H., 2018. Ten steps to get started in Genome Assembly and Annotation. F1000Res 7, 148. https://doi.org/10.12688/f1000research.13598.1

Doucet-Beaupré, H., Breton, S., Chapman, E.G., Blier, P.U., Bogan, A.E., Stewart, D.T., Hoeh, W.R., 2010. Mitochondrial phylogenomics of the Bivalvia (Mollusca): Searching for the origin and mitogenomic correlates of doubly uniparental inheritance of mtDNA. BMC Evol Biol 10, 1–19. https://doi.org/10.1186/1471-2148-10-50/FIGURES/4

Dreyer, H., Steiner, G., 2004. The complete sequence and gene organization of the mitochondrial genome of the gadilid scaphopod *Siphonondentalium lobatum* (Mollusca). Mol Phylogenet Evol 31, 605–617. https://doi.org/10.1016/J.YMPEV.2003.08.007

Du, X., Fan, G., Jiao, Y., Zhang, H., Guo, X., Huang, R., Zheng, Z., Bian, C., Deng, Y., Wang, Q., Wang, Z., Liang, X., Liang, H., Shi, C., Zhao, X., Sun, F., Hao, R., Bai, J., Liu, J., Chen, W., Liang, J., Liu, W., Xu, Z., Shi, Q., Xu, X., Zhang, G., Liu, X., 2017. The pearl oyster *Pinctada fucata martensii* genome and multi-omic analyses provide insights into biomineralization. Gigascience 6. https://doi.org/10.1093/gigascience/gix059

Dudchenko, O., Shamim, M.S., Batra, S., Durand, N.C., Musial, N.T., Mostofa, R., Pham, M., Hilaire, B.G.S., Yao, W., Stamenova, E., Hoeger, M., Nyquist, S.K., Korchina, V., Pletch, K., Flanagan, J.P., Tomaszewicz, A., McAloose, D., Estrada, C.P., Novak, B.J., Omer, A.D., Aiden, E.L., 2018. The Juicebox Assembly Tools module facilitates de novo assembly of mammalian genomes with chromosome-length scaffolds for under $1000. bioRxiv 254797. https://doi.org/10.1101/254797

Dudgeon, D., Arthington, A.H., Gessner, M.O., Kawabata, Z.I., Knowler, D.J., Lévêque, C., Naiman, R.J., Prieur-Richard, A.H., Soto, D., Stiassny, M.L.J., Sullivan, C.A., 2006. Freshwater biodiversity: Importance, threats, status and conservation challenges. Biol Rev Camb Philos Soc. https://doi.org/10.1017/S1464793105006950

Dunisławska, A., Łachmańska, J., Sławińska, A., Siwek, M., 2017. Next generation sequencing in animal science-a review, Animal Science Papers and Reports.

Dunn, C.W., Ryan, J.F., 2015. The evolution of animal genomes. Curr Opin Genet Dev 35, 25–32. https://doi.org/10.1016/J.GDE.2015.08.006

Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26, 2460–2461. https://doi.org/10.1093/bioinformatics/btq461

Edgar, R.C., 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32, 1792–1797. https://doi.org/10.1093/nar/gkh340

Edwards, S. v., Xi, Z., Janke, A., Faircloth, B.C., McCormack, J.E., Glenn, T.C., Zhong, B., Wu, S., Lemmon, E.M., Lemmon, A.R., Leaché, A.D., Liu, L., Davis, C.C., 2016. Implementing and testing the multispecies coalescent model: A valuable paradigm for phylogenomics. Mol Phylogenet Evol 94, 447–462. https://doi.org/10.1016/J.YMPEV.2015.10.027

Egeter, B., Veríssimo, J., Lopes-Lima, M., Chaves, C., Pinto, J., Riccardi, N., Beja, P., Fonseca, N.A., 2022. Speeding up the detection of invasive bivalve species using environmental DNA: A Nanopore and Illumina sequencing comparison. Mol Ecol Resour 22, 2232–2247. https://doi.org/10.1111/1755-0998.13610

Eisen, J.A., Fraser, C.M., 2003. Phylogenomics: Intersection of Evolution and Genomics. Science (1979) 300, 1706–1707. https://doi.org/10.1126/SCIENCE.1086292

Elson, J.L., Lightowlers, R.N., 2006. Mitochondrial DNA clonality in the dock: can surveillance swing the case? Trends in Genetics 22, 603–607. https://doi.org/10.1016/J.TIG.2006.09.004

Emms, D.M., Kelly, S., 2019. OrthoFinder: Phylogenetic orthology inference for comparative genomics. Genome Biol 20, 238. https://doi.org/10.1186/s13059-019-1832-y

Evans, J.D., Brown, S.J., Hackett, K.J.J., Robinson, G., Richards, S., Lawson, D., Elsik, C., Coddington, J., Edwards, O., Emrich, S., Gabaldon, T., Goldsmith, M., Hanes, G., Misof, B., Muñoz-Torres, M., Niehuis, O., Papanicolaou, A., Pfrender, M., Poelchau, M., Purcell-Miramontes, M., Robertson, H.M., Ryder, O., Tagu, D., Torres, T., Zdobnov, E., Zhang, G., Zhou, X., 2013. The i5K Initiative: Advancing Arthropod Genomics for Knowledge, Human Health, Agriculture, and the Environment. Journal of Heredity 104, 595–600. https://doi.org/10.1093/jhered/est050

Faircloth, B.C., 2016. PHYLUCE is a software package for the analysis of conserved genomic loci. Bioinformatics 32, 786–788. https://doi.org/10.1093/BIOINFORMATICS/BTV646

Faircloth, B.C., McCormack, J.E., Crawford, N.G., Harvey, M.G., Brumfield, R.T., Glenn, T.C., 2012. Ultraconserved Elements Anchor Thousands of Genetic Markers Spanning Multiple Evolutionary Timescales. Syst Biol 61, 717–726. https://doi.org/10.1093/SYSBIO/SYS004

Farrington, J.W., Tripp, B.W., Tanabe, S., Subramanian, A., Sericano, J.L., Wade, T.L., Knap, A.H., 2016. Edward D. Goldberg's proposal of "the Mussel Watch": Reflections after 40 years. Mar Pollut Bull 110, 501–510. https://doi.org/10.1016/J.MARPOLBUL.2016.05.074

Farrington, S.J., King, R.W., Baker, J.A., Gibbons, J.G., 2020. Population genetics of freshwater pearl mussel (*Margaritifera margaritifera*) in central Massachusetts and implications for conservation. Aquat Conserv 30, 1945–1958. https://doi.org/10.1002/aqc.3439

Fearnley, I.M., Walker, J.E., 1986. Two overlapping genes in bovine mitochondrial DNA encode membrane components of ATP synthase. EMBO J 5, 2003–2008. https://doi.org/10.1002/J.1460-2075.1986.TB04456.X

Feldmeyer, B., Hoffmeier, K., Pfenninger, M., 2010. The complete mitochondrial genome of *Radix balthica* (Pulmonata, Basommatophora), obtained by low coverage shot gun next generation sequencing. Mol Phylogenet Evol 57, 1329–1333. https://doi.org/10.1016/J.YMPEV.2010.09.012

Fernández, R., Kallal, R.J., Dimitrov, D., Ballesteros, J.A., Arnedo, M.A., Giribet, G., Hormiga, G., 2018. Phylogenomics, Diversification Dynamics, and Comparative Transcriptomics across the Spider Tree of Life. Current Biology 28, 1489-1497.e5. https://doi.org/10.1016/J.CUB.2018.03.064

Fernández-Pérez, J., Froufe, E., Nantón, A., Gaspar, M.B., Méndez, J., 2017. Genetic diversity and population genetic analysis of *Donax vittatus* (Mollusca: Bivalvia) and phylogeny of the genus with mitochondrial and nuclear markers. Estuar Coast Shelf Sci 197, 126–135. https://doi.org/10.1016/J.ECSS.2017.08.032

Fernández-Silva, P., Enriquez, J.A., Montoya, J., 2003. Replication and Transcription of Mammalian Mitochondrial Dna. Exp Physiol 88, 41–56. https://doi.org/10.1113/EPH8802514

Ferreira-Rodríguez, N., Akiyama, Y.B., Aksenova, O. v., Araujo, R., Christopher Barnhart, M., Bespalaya, Y. v., Bogan, A.E., Bolotov, I.N., Budha, P.B., Clavijo, C., Clearwater, S.J., Darrigran, G., Do, V.T., Douda, K., Froufe, E., Gumpinger, C., Henrikson, L., Humphrey, C.L., Johnson, N.A., Klishko, O., Klunzinger, M.W., Kovitvadhi, S., Kovitvadhi, U., Lajtner, J., Lopes-Lima, M., Moorkens, E.A., Nagayama, S., Nagel, K.-O., Nakano, M., Negishi, J.N., Ondina, P., Oulasvirta, P., Prié, V., Riccardi, N., Rudzīte, M., Sheldon, F., Sousa, R., Strayer, D.L., Takeuchi, M., Taskinen, J., Teixeira, A., Tiemann, J.S., Urbańska, M., Varandas, S., Vinarski, M. v., Wicklow, B.J., Zając, T., Vaughn, C.C., 2019. Research priorities for freshwater mussel conservation assessment. Biol Conserv 231, 77–87. https://doi.org/10.1016/J.BIOCON.2019.01.002

Ferrier, D.E.K., Holland, P.W.H., 2002. *Ciona intestinalis* ParaHox genes: Evolution of Hox/ParaHox cluster integrity, developmental mode, and temporal colinearity. Mol Phylogenet Evol 24, 412–417. https://doi.org/10.1016/S1055-7903(02)00204-X

Ferrier, D.E.K., Holland, P.W.H., 2001. Ancient origin of the Hox gene cluster. Nat Rev Genet. https://doi.org/10.1038/35047605

Field, K.G., Olsen, G.J., Lane, D.J., Giovannoni, S.J., Ghiselin, M.T., Raff, E.C., Pace, N.R., Raff, R.A., 1988. Molecular Phylogeny of the Animal Kingdom. Science (1979) 239, 748–753. https://doi.org/10.1126/SCIENCE.3277277

Fiers, W., Contreras, R., Duerinck, F., Haegeman, G., Iserentant, D., Merregaert, J., Min Jou, W., Molemans, F., Raeymaekers, A., van den Berghe, A., Volckaert, G., Ysebaert, M., 1976. Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. Nature 1976 260:5551 260, 500–507. https://doi.org/10.1038/260500a0

Figueras, A., Moreira, R., Sendra, M., Novoa, B., 2019. Genomics and immunity of the Mediterranean mussel *Mytilus galloprovincialis* in a changing environment. Fish Shellfish Immunol 90, 440–445. https://doi.org/10.1016/j.fsi.2019.04.064

Finn, R.D., Clements, J., Eddy, S.R., 2011. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res 39, W29–W37. https://doi.org/10.1093/NAR/GKR367

Fitch, W.M., Margoliash, E., 1967. Construction of phylogenetic trees. Science (1979) 155, 279–284. https://doi.org/10.1126/SCIENCE.155.3760.279/ASSET/D2E24AD2-D8C2-4193-9548-046C303E8AEB/ASSETS/SCIENCE.155.3760.279.FP.PNG

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.-F., Dougherty, B.A., Merrick, J.M., McKenney, K., Sutton, G., FitzHugh, W., Fields, C., Gocayne, J.D., Scott, J., Shirley, R., Liu, L., Glodek, A., Kelley, J.M., Weidman, J.F., Phillips, C.A., Spriggs, T., Hedblom, E., Cotton, M.D., Utterback, T.R., Hanna, M.C., Nguyen, D.T., Saudek, D.M., Brandon, R.C., Fine, L.D., Fritchman, J.L., Fuhrmann, J.L., Geoghagen, N.S.M., Gnehm, C.L., McDonald, L.A., Small, K. v., Fraser, C.M., Smith, H.O., Venter, J.C., 1995. Whole-Genome Random Sequencing and Assembly of *Haemophilus influenzae* Rd. Science (1979) 269, 496–512. https://doi.org/10.1126/science.7542800

Fleming, J.F., Arakawa, K., 2021. Systematics of tardigrada: A reanalysis of tardigrade taxonomy with specific reference to Guil et al. (2019). Zool Scr 50, 376–382. https://doi.org/10.1111/ZSC.12476

Fonseca, M.M., Harris, D.J., Posada, D., 2014. The Inversion of the Control Region in Three Mitogenomes Provides Further Evidence for an Asymmetric Model of Vertebrate mtDNA Replication. PLoS One 9, e106654. https://doi.org/10.1371/JOURNAL.PONE.0106654

Fonseca, M.M., Lopes-Lima, M., Eackles, M.S., King, T.L., Froufe, E., 2016. The female and male mitochondrial genomes of Unio delphinus and the phylogeny of freshwater mussels (Bivalvia: Unionida). Mitochondrial DNA B Resour 1, 954–957. https://doi.org/10.1080/23802359.2016.1241677

Formenti, G., Theissinger, K., Fernandes, C., Bista, I., Bombarely, A., Bleidorn, C., Ciofi, C., Crottini, A., Godoy, J.A., Höglund, J., Malukiewicz, J., Mouton, A., Oomen, R.A., Paez, S., Palsbøll, P.J., Pampoulie, C., Ruiz-López, María J., Svardal, H., Theofanopoulou, C., de Vries, J., Waldvogel, A.M., Zhang, Guojie, Mazzoni, C.J., Jarvis, E.D., Bálint, M., Čiampor, F., Hoglund, J., Palsbøll, P., Ruiz-López, María José, Zhang, Goujie, Jarvis, E., Aghayan, S.A., Alioto, T.S., Almudi, I., Alvarez, N., Alves, P.C., Amorim, I.R., Antunes, A., Arribas, P., Baldrian, P., Berg, P.R., Bertorelle, G., Böhne, A., Bonisoli-Alquati, A., Boštjančić, L.L., Boussau, B., Breton, C.M., Buzan, E., Campos, P.F., Carreras, C., Castro, L.Fi., Chueca, L.J., Conti, E., Cook-Deegan, R., Croll, D., Cunha, M. v., Delsuc, F., Dennis, A.B., Dimitrov, D., Faria, R., Favre, A., Fedrigo, O.D., Fernández, R., Ficetola, G.F., Flot, J.F., Gabaldón, T., Galea Agius, D.R., Gallo, G.R., Giani, A.M., Gilbert, M.T.P., Grebenc, T., Guschanski, K., Guyot, R., Hausdorf, B., Hawlitschek, O., Heintzman, P.D., Heinze, B., Hiller, M., Husemann, M., Iannucci, A., Irisarri, I., Jakobsen, K.S., Jentoft, S., Klinga, P., Kloch, A., Kratochwil, C.F., Kusche, H., Layton, K.K.S., Leonard, J.A., Lerat, E., Liti, G., Manousaki, T., Marques-Bonet, T., Matos-Maraví, P., Matschiner, M., Maumus, F., Mc Cartney, A.M., Meiri, S., Melo-

Ferreira, J., Mengual, X., Monaghan, M.T., Montagna, M., Mysłajek, R.W., Neiber, M.T., Nicolas, V., Novo, M., Ozretić, P., Palero, F., Pârvulescu, L., Pascual, M., Paulo, O.S., Pavlek, M., Pegueroles, C., Pellissier, L., Pesole, G., Primmer, C.R., Riesgo, A., Rüber, L., Rubolini, D., Salvi, D., Seehausen, O., Seidel, M., Secomandi, S., Studer, B., Theodoridis, S., Thines, M., Urban, L., Vasemägi, A., Vella, A., Vella, N., Vernes, S.C., Vernesi, C., Vieites, D.R., Waterhouse, R.M., Wheat, C.W., Wörheide, G., Wurm, Y., Zammit, G., 2022. The era of reference genomes in conservation genomics. Trends Ecol Evol 37, 197–202. https://doi.org/10.1016/J.TREE.2021.11.008/ATTACHMENT/B2798C5E-5A67-4C4E-B9EE-B4A01F77D86A/MMC1.XLSX

Fourdrilis, S., de Frias Martins, A.M., Backeljau, T., 2018. Relation between mitochondrial DNA hyperdiversity, mutation rate and mitochondrial genome evolution in *Melarhaphe neritoides* (Gastropoda: Littorinidae) and other Caenogastropoda. Scientific Reports 2018 8:1 8, 1–12. https://doi.org/10.1038/s41598-018-36428-7

Francino, M.P., Ochman, H., 1997. Strand asymmetries in DNA evolution. Trends in Genetics 13, 240–245. https://doi.org/10.1016/S0168-9525(97)01118-9

Fraser, C.M., Gocayne, J.D., White, O., Adams, M.D., Clayton, R.A., Fleischmann, R.D., Bult, C.J., Kerlavage, A.R., Sutton, G., Kelley, J.M., Fritchman, J.L., Weidman, J.F., Small, K. v., Sandusky, M., Fuhrmann, J., Nguyen, D., Utterback, T.R., Saudek, D.M., Phillips, C.A., Merrick, J.M., Tomb, J.-F., Dougherty, B.A., Bott, K.F., Hu, P.-C., Lucier, T.S., Peterson, S.N., Smith, H.O., Hutchison, C.A., Venter, J.C., 1995. The Minimal Gene Complement of *Mycoplasma genitalium*. Science (1979) 270, 397–404. https://doi.org/10.1126/science.270.5235.397

From, a S., Global, T.H.E., Database, I.S., 2000. 100 of the World ’ S Worst Invasive Alien Species a Selection From the Global. ISSG 12, 12. https://doi.org/10.1614/WT-04-126.1

Froufe, E., Bolotov, I., Aldridge, D.C., Bogan, A.E., Breton, S., Gan, H.M., Kovitvadhi, U., Kovitvadhi, S., Riccardi, N., Secci-Petretto, G., Sousa, R., Teixeira, A., Varandas, S., Zanatta, D., Zieritz, A., Fonseca, M.M., Lopes-Lima, M., 2019. Mesozoic mitogenome rearrangements and freshwater mussel (Bivalvia: Unionoidea) macroevolution. Heredity 2019 124:1 124, 182–196. https://doi.org/10.1038/s41437-019-0242-y

Froufe, E., Gan, H.M., Lee, Y.P., Carneiro, J., Varandas, S., Teixeira, A., Zieritz, A., Sousa, R., Lopes-Lima, M., 2016a. The male and female complete mitochondrial genome sequences of the Endangered freshwater mussel *Potomida littoralis* (Cuvier, 1798) (Bivalvia: Unionidae). Mitochondrial DNA Part A 27, 3571–3572. https://doi.org/10.3109/19401736.2015.1074223

Froufe, E., Gonçalves, D. v, Teixeira, A., Sousa, R., Varandas, S., Ghamizi, M., Zieritz, A., Lopes-Lima, M., 2016b. Who lives where? Molecular and morphometric analyses clarify which *Unio* species (Unionida, Mollusca) inhabit the southwestern Palearctic. Org Divers Evol 16, 597–611. https://doi.org/10.1007/s13127-016-0262-x

Froufe, E., Prié, V., Faria, J., Ghamizi, M., Gonçalves, D. v., Gürlek, M.E., Karaouzas, I., Kebapçi, Ü., Şereflişan, H., Sobral, C., Sousa, R., Teixeira, A., Varandas, S., Zogaris, S., Lopes-Lima, M., 2016c. Phylogeny, phylogeography, and evolution in the Mediterranean region: News from a freshwater mussel (*Potomida*, Unionida). Mol Phylogenet Evol 100, 322–332. https://doi.org/10.1016/J.YMPEV.2016.04.030

Froufe, E., Sobral, C., Teixeira, A., Sousa, R., Varandas, S., Aldridge, D.C., Lopes-Lima, M., 2014. Genetic diversity of the pan-European freshwater mussel *Anodonta anatina* (Bivalvia: Unionoida) based on CO1: New phylogenetic insights and implications for conservation. Aquat Conserv 24, 561–574. https://doi.org/10.1002/aqc.2456

Funabara, D., Watanabe, D., Satoh, N., Kanoh, S., 2013. Genome-Wide Survey of Genes Encoding Muscle Proteins in the Pearl Oyster, *Pinctada fucata*. https://doi.org/10.2108/zsj.30.817 30, 817–825. https://doi.org/10.2108/ZSJ.30.817

Furuhashi, T., Schwarzinger, C., Miksik, I., Smrz, M., Beran, A., 2009. Molluscan shell evolution with review of shell calcification hypothesis. Comp Biochem Physiol B Biochem Mol Biol 154, 351–371. https://doi.org/10.1016/J.CBPB.2009.07.011

Gan, H.M., Schultz, M.B., Austin, C.M., 2014. Integrated shotgun sequencing and bioinformatics pipeline allows ultra-fast mitogenome recovery and confirms substantial gene rearrangements in Australian freshwater crayfishes. BMC Evol Biol 14, 19. https://doi.org/10.1186/1471-2148-14-19

Ganser, A.M., Newton, T.J., Haro, R.J., 2015. Effects of elevated water temperature on physiological responses in adult freshwater mussels. Freshw Biol 60, 1705–1716. https://doi.org/10.1111/FWB.12603

Gao, B., Peng, C., Chen, Q., Zhang, J., Shi, Q., 2018. Mitochondrial genome sequencing of a vermivorous cone snail *Conus quercinus* supports the correlative analysis between phylogenetic relationships and dietary types of *Conus species*. PLoS One 13, e0193053. https://doi.org/10.1371/JOURNAL.PONE.0193053

Garone, C., Minczuk, M., D’Souza, A.R., Minczuk, M., 2018. Mitochondrial transcription and translation: overview. Essays Biochem 62, 309–320. https://doi.org/10.1042/EBC20170102

Garrison, N.L., Johnson, P.D., Whelan, N. v., 2021. Conservation genomics reveals low genetic diversity and multiple parentage in the threatened freshwater mussel, *Margaritifera hembeli*. Conservation Genetics 22, 217–231. https://doi.org/10.1007/s10592-020-01329-8

Gazeau, F., Parker, L.M., Comeau, S., Gattuso, J.-P., O’Connor, W.A., Martin, S., Pörtner, H.-O., Ross, P.M., 2013. Impacts of ocean acidification on marine shelled molluscs. Mar Biol 160, 2207–2245. https://doi.org/10.1007/s00227-013-2219-3

Geist, J., 2011. Integrative freshwater ecology and biodiversity conservation. Ecol Indic 11, 1507–1516. https://doi.org/10.1016/j.ecolind.2011.04.002

Geist, J., 2010. Strategies for the conservation of endangered freshwater pearl mussels (*Margaritifera margaritifera* L.): a synthesis of Conservation Genetics and Ecology. Hydrobiologia 644, 69–88. https://doi.org/10.1007/s10750-010-0190-2

Geist, J., Auerswald, K., Boom, A., 2005. Stable carbon isotopes in freshwater mussel shells: Environmental record or marker for metabolic activity? Geochim Cosmochim Acta 69, 3545–3554. https://doi.org/10.1016/j.gca.2005.03.010

Geist, J., Kuehn, R., 2005. Genetic diversity and differentiation of central European freshwater pearl mussel (*Margaritifera margaritifera* L.) populations: Implications for conservation and management. Mol Ecol 14, 425–439. https://doi.org/10.1111/J.1365-294X.2004.02420.x

Geniza, M., Jaiswal, P., 2017. Tools for building de novo transcriptome assembly. Curr Plant Biol 11–12, 41–45. https://doi.org/10.1016/J.CPB.2017.12.004

Gerdol, M., Moreira, R., Cruz, F., Gómez-Garrido, J., Vlasova, A., Rosani, U., Venier, P., Naranjo-Ortiz, M.A., Murgarella, M., Greco, S., Balseiro, P., Corvelo, A., Frias, L., Gut, M., Gabaldón, T., Pallavicini, A., Canchaya, C., Novoa, B., Alioto, T.S., Posada, D., Figueras, A., 2020. Massive gene presence-absence variation shapes an open pan-genome in the Mediterranean mussel. Genome Biol 21, 275. https://doi.org/10.1186/S13059-020-02180-3/FIGURES/2

Ghiselli, F., Gomes-Dos-Santos, A., Adema, C.M., Lopes-Lima, M., Sharbrough, J., Boore, J.L., 2021. Molluscan mitochondrial genomes break the rules. Philosophical Transactions of the Royal Society B: Biological Sciences. https://doi.org/10.1098/rstb.2020.0159

Ghiselli, F., Iannello, M., Puccio, G., Chang, P.L., Plazzi, F., Nuzhdin, S. v, Passamonti, M., 2018. Comparative transcriptomics in two bivalve species offers different perspectives on the evolution of sex-biased genes. Genome Biol Evol 10, 1389–1402. https://doi.org/10.1093/gbe/evy082

Ghiselli, F., Maurizii, M.G., Reunov, A., Ariño-Bassols, H., Cifaldi, C., Pecci, A., Alexandrova, Y., Bettini, S., Passamonti, M., Franceschini, V., Milani, L., 2019. Natural Heteroplasmy and Mitochondrial Inheritance in Bivalve Molluscs. Integr Comp Biol. https://doi.org/10.1093/icb/icz061

Ghiselli, F., Milani, L., Chang, P.L., Hedgecock, D., Davis, J.P., Nuzhdin, S. v., Passamonti, M., 2012. De Novo Assembly of the Manila Clam *Ruditapes philippinarum* Transcriptome Provides New Insights into Expression Bias, Mitochondrial Doubly Uniparental Inheritance and Sex Determination. Mol Biol Evol 29, 771–786. https://doi.org/10.1093/MOLBEV/MSR248

Ghiselli, F., Milani, L., Guerra, D., Chang, P.L., Breton, S., Nuzhdin, S. v., Passamonti, M., 2013. Structure, Transcription, and Variability of Metazoan Mitochondrial Genome: Perspectives from an Unusual Mitochondrial Inheritance System. Genome Biol Evol 5, 1535–1554. https://doi.org/10.1093/GBE/EVT112

Ghiselli, F., Milani, L., Iannello, M., Procopio, E., Chang, P.L., Nuzhdin, S. v., Passamonti, M., 2017. The complete mitochondrial genome of the grooved carpet shell, *Ruditapes decussatus* (Bivalvia, Veneridae). PeerJ 2017, e3692. https://doi.org/10.7717/PEERJ.3692/SUPP-8

Ghiselli, F., Milani, L., Passamonti, M., 2011. Strict Sex-Specific mtDNA Segregation in the Germ line of the DUI Species *Venerupis philippinarum* (Bivalvia: Veneridae). Mol Biol Evol 28, 949–961. https://doi.org/10.1093/MOLBEV/MSQ271

Ghosh, M., Sharma, N., Singh, A.K., Gera, M., Pulicherla, K.K., Jeong, D.K., 2018. Transformation of animal genomics by next-generation sequencing technologies: a decade of challenges and their impact on genetic architecture. Crit Rev Biotechnol 38, 1157–1175. https://doi.org/10.1080/07388551.2018.1451819

Giani, A.M., Gallo, G.R., Gianfranceschi, L., Formenti, G., 2020. Long walk to genomics: History and current approaches to genome sequencing and assembly. Comput Struct Biotechnol J 18, 9–19. https://doi.org/10.1016/J.CSBJ.2019.11.002

Gissi, C., Iannelli, F., Pesole, G., 2008. Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. Heredity 2008 101:4 101, 301–320. https://doi.org/10.1038/hdy.2008.62

Glanzman, D.L., 2009. Habituation in *Aplysia*: The Cheshire Cat of neurobiology. Neurobiol Learn Mem 92, 147–154. https://doi.org/10.1016/j.nlm.2009.03.005

Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., Louis, E.J., Mewes, H.W., Murakami, Y., Philippsen, P., Tettelin, H., Oliver, S.G., 1996. Life with 6000 Genes. Science (1979) 274, 546–567. https://doi.org/10.1126/science.274.5287.546

Gomes-dos-Santos, A., Froufe, E., Amaro, R., Ondina, P., Breton, S., Guerra, D., Aldridge, D.C., Bolotov, I.N., Vikhrev, I. v., Gan, H.M., Gonçalves, D. v., Bogan, A.E., Sousa, R., Stewart, D., Teixeira, A., Varandas, S., Zanatta, D., Lopes-Lima, M., 2019. The male and female complete mitochondrial genomes of the threatened freshwater pearl mussel *Margaritifera margaritifera* (Linnaeus, 1758) (Bivalvia: Margaritiferidae). Mitochondrial DNA B Resour 4, 1417–1420. https://doi.org/10.1080/23802359.2019.1598794

Gomes-dos-Santos, A., Lopes-Lima, M., Castro, L.F.C., Froufe, E., 2020. Molluscan genomics: the road so far and the way forward. Hydrobiologia. https://doi.org/10.1007/s10750-019-04111-1

Gomes-dos-Santos, A., Lopes-Lima, M., Machado, A.M., Ramos, A.M., Usié, A., Bolotov, I.N., Vikhrev, I. v, Breton, S., C Castro, L.F., da Fonseca, R.R., Geist, J., Österling, M.E., Prié, V., Teixeira, A., Gan, H.M., Simakov, O., Froufe, E., 2021. The Crown Pearl : a draft genome assembly of the European freshwater pearl mussel *Margaritifera margaritifera* (Linnaeus, 1758). DNA Research. https://doi.org/10.1093/dnares/dsab002

Gomes-dos-Santos, A., Machado, A.M., Castro, L.F.C., Prié, V., Teixeira, A., Lopes-Lima, M., Froufe, E., 2022. The gill transcriptome of threatened European freshwater mussels. Scientific Data 2022 9:1 9, 1–10. https://doi.org/10.1038/s41597-022-01613-x

Gómez-Chiarri, M., Warren, W.C., Guo, X., Proestou, D., 2015. Developing tools for the study of molluscan immunity: The sequencing of the genome of the eastern oyster, *Crassostrea virginica*. Fish Shellfish Immunol 46, 2–4. https://doi.org/10.1016/J.FSI.2015.05.004

Gonzalez, D.R., Aramendia, A.C., Davison, A., 2019. Recombination within the *Cepaea nemoralis* supergene is confounded by incomplete penetrance and epistasis. Heredity (Edinb) 123, 153–161. https://doi.org/10.1038/s41437-019-0190-6

Gonzalez, V.L., Andrade, S.C.S., Bieler, R., Collins, T.M., Dunn, C.W., Mikkelsen, P.M., Taylor, J.D., Giribet, G., 2015. A phylogenetic backbone for Bivalvia: an RNA-seq approach. Proceedings of the Royal Society B: Biological Sciences 282, 20142332–20142332. https://doi.org/10.1098/rspb.2014.2332

Goodwin, S., McPherson, J.D., McCombie, W.R., 2016. Coming of age: Ten years of next-generation sequencing technologies. Nat Rev Genet. https://doi.org/10.1038/nrg.2016.49

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., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology 2011 29:7 29, 644–652. https://doi.org/10.1038/nbt.1883

Graf, D.L., 2013. Patterns of Freshwater Bivalve Global Diversity and the State of Phylogenetic Studies on the Unionoida, Sphaeriidae, and Cyrenidae *. https://doi.org/10.4003/006.031.0106 31, 135–153. https://doi.org/10.4003/006.031.0106

Graf, D.L., Cummings, K.S., 2022. The freshwater mussels (Unionoida) of the world (and other less consequential bivalves). http://www.mussel-project.net/.

Graf, Daniel L., Cummings, K.S., 2021. A ‘big data’ approach to global freshwater mussel diversity (Bivalvia: Unionoida), with an updated checklist of genera and species. Journal of Molluscan Studies 87, 34. https://doi.org/10.1093/MOLLUS/EYAA034

Graf, D.L., Cummings, K.S., 2007. Review of the systematics and global diversity of freshwater mussel species (Bivalvia: Unionoida). Journal of Molluscan Studies. https://doi.org/10.1093/mollus/eym029

Graf, D.L., Jones, H., Geneva, A.J., Pfeiffer, J.M., Klunzinger, M.W., 2015. Molecular phylogenetic analysis supports a Gondwanan origin of the Hyriidae (Mollusca: Bivalvia: Unionida) and the paraphyly of Australasian taxa. Mol Phylogenet Evol 85, 1–9. https://doi.org/https://doi.org/10.1016/j.ympev.2015.01.012

Grande, C., 2001. The Gonads Of *Margaritifera Auricularia* (Spengler, 1793) And M. Margaritifera (Linnaeus, 1758) (Bivalvia: Unionoidea). Journal Molluscan Studies 67, 27–36. https://doi.org/10.1093/mollus/67.1.27

Grande, C., Templado, J., Cervera, J.L., Zardoya, R., 2002. The Complete Mitochondrial Genome of the Nudibranch *Roboastra europaea* (Mollusca: Gastropoda) Supports the Monophyly of Opisthobranchs. Mol Biol Evol 19, 1672–1685. https://doi.org/10.1093/OXFORDJOURNALS.MOLBEV.A003990

Grande, C., Templado, J., Zardoya, R., 2008. Evolution of gastropod mitochondrial genome arrangements. BMC Evol Biol 8, 1–15. https://doi.org/10.1186/1471-2148-8-61/FIGURES/4

Guerra, D., Bouvet, K., Breton, S., 2018. Mitochondrial gene order evolution in Mollusca: Inference of the ancestral state from the mtDNA of *Chaetopleura apiculata* (Polyplacophora, Chaetopleuridae). Mol Phylogenet Evol 120, 233–239. https://doi.org/10.1016/J.YMPEV.2017.12.013

Guerra, D., Ghiselli, F., Passamonti, M., 2014. The largest unassigned regions of the male- and female-transmitted mitochondrial DNAs in *Musculista senhousia* (Bivalvia Mytilidae). Gene 536, 316–325. https://doi.org/10.1016/J.GENE.2013.12.005

Guerra, D., Lopes-Lima, M., Froufe, E., Gan, H.M., Ondina, P., Amaro, R., Klunzinger, M.W., Callil, C., Prié, V., Bogan, A.E., Stewart, D.T., Breton, S., 2019. Variability of mitochondrial ORFans hints at possible differences in the system of doubly uniparental inheritance of mitochondria among families of freshwater mussels (Bivalvia: Unionida). BMC Evol Biol 19, 229. https://doi.org/10.1186/s12862-019-1554-5

Guerra, D., Plazzi, F., Stewart, D.T., Bogan, A.E., Hoeh, W.R., Breton, S., 2017. Evolution of sex-dependent mtDNA transmission in freshwater mussels (Bivalvia: Unionida). Sci Rep 7, 1551. https://doi.org/10.1038/s41598-017-01708-1

Guo, X., 2009. Use and exchange of genetic resources in molluscan aquaculture. Rev Aquac 1, 251–259. https://doi.org/10.1111/j.1753-5131.2009.01014.x

Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G., 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29, 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Gusman, A., Azuelos, C., Breton, S., 2017. No evidence of sex-linked heteroplasmy or doubly-uniparental inheritance of mtDNA in five gastropod species. Journal of Molluscan Studies 83, 119–122. https://doi.org/10.1093/MOLLUS/EYW034

Gusman, A., Lecomte, S., Stewart, D.T., Passamonti, M., Breton, S., 2016. Pursuing the quest for better understanding the taxonomic distribution of the system of doubly uniparental inheritance of mtdna. PeerJ 2016, e2760. https://doi.org/10.7717/PEERJ.2760/SUPP-5

Haag, W.R., 2012. North American freshwater mussels: natural history, ecology, and conservation. Cambridge University Press.

Haag, W.R., Rypel, A.L., 2011. Growth and longevity in freshwater mussels: evolutionary and conservation implications. Biological Reviews 86, 225–247. https://doi.org/10.1111/J.1469-185X.2010.00146.X

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., Regev, A., 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc 8, 1494–1512. https://doi.org/10.1038/nprot.2013.084

Hahn, C., Bachmann, L., Chevreux, B., 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. Nucleic Acids Res 41, e129–e129. https://doi.org/10.1093/NAR/GKT371

Halanych, K.M., 2004. The New View of Animal Phylogeny. Annu Rev Ecol Evol Syst 35, 229–256.

Haney, A., Abdelrahman, H., Stoeckel, J.A., 2020. Effects of thermal and hypoxic stress on respiratory patterns of three unionid species: implications for management and conservation. Hydrobiologia 847, 787–802. https://doi.org/10.1007/S10750-019-04138-4/FIGURES/7

Harris, T.D., Buzby, P.R., Babcock, H., Beer, E., Bowers, J., Braslavsky, I., Causey, M., Colonell, J., DiMeo, J., Efcavitch, J.W., Giladi, E., Gill, J., Healy, J., Jarosz, M., Lapen, D., Moulton, K., Quake, S.R., Steinmann, K., Thayer, E., Tyurina, A., Ward, R., Weiss, H., Xie, Z., 2008. Single-molecule DNA sequencing of a viral genome. Science (1979) 320, 106–109. https://doi.org/10.1126/SCIENCE.1150427/SUPPL_FILE/HARRIS.SOM.PDF

Hassall, C., Amaro, R., Ondina, P., Outeiro, A., Cordero-Rivera, A., San Miguel, E., 2017. Population-level variation in senescence suggests an important role for temperature in an endangered mollusc. J Zool 301, 32–40. https://doi.org/10.1111/jzo.12395

Hassanin, A., Léger, N., Deutsch, J., 2005. Evidence for Multiple Reversals of Asymmetric Mutational Constraints during the Evolution of the Mitochondrial Genome of Metazoa, and Consequences for Phylogenetic Inferences. Syst Biol 54, 277–298. https://doi.org/10.1080/10635150590947843

Haszprunar, G., Wanninger, A., 2012. Molluscs. Current Biology 22, R510–R514. https://doi.org/10.1016/j.cub.2012.05.039

Hatzoglou, E., Rodakis, G.C., Lecanidou, R., 1995. Complete sequence and gene organization of the mitochondrial genome of the land snail *Albinaria coerulea*. Genetics 140, 1353–1366. https://doi.org/10.1093/GENETICS/140.4.1353

He, C.B., Wang, J., Gao, X.G., Song, W.T., Li, H.J., Li, Y.F., Liu, W.D., Su, H., 2011. The complete mitochondrial genome of the hard clam *Meretrix meretrix*. Mol Biol Rep 38, 3401–3409. https://doi.org/10.1007/S11033-010-0449-8/FIGURES/4

Heather, J.M., Chain, B., 2016. The sequence of sequencers: The history of sequencing DNA. Genomics 107, 1–8. https://doi.org/10.1016/J.YGENO.2015.11.003

Hessling, T. von, 1859. Die Perlnmuscheln und thre Perlen (Naturwissen-schaftlich und geschichtlich mit, Beruecksichtigung der Perlgewaesser Baerns). Forgotten Books, Leipzig.

Higgs, P.G., Jameson, D., Jow, H., Rattray, M., 2003. The Evolution of tRNA-Leu Genes in Animal Mitochondrial Genomes. J Mol Evol 57, 435–445. https://doi.org/10.1007/S00239-003-2494-6/FIGURES/6

Hinchliff, C.E., Smith, S.A., Allman, J.F., Burleigh, J.G., Chaudhary, R., Coghill, L.M., Crandall, K.A., Deng, J., Drew, B.T., Gazis, R., Gude, K., Hibbett, D.S., Katz, L.A., Dail Laughinghouse, H., McTavish, E.J., Midford, P.E., Owen, C.L., Ree, R.H., Rees, J.A., Soltisc, D.E., Williams, T., Cranston, K.A., 2015. Synthesis of phylogeny and taxonomy into a comprehensive tree of life. Proc Natl Acad Sci U S A 112, 12764–12769. https://doi.org/10.1073/PNAS.1423041112/SUPPL_FILE/PNAS.1423041112.SD02.CSV

Hinzmann, M., Lopes-Lima, M., Teixeira, A., Varandas, S., Sousa, R., Lopes, A., Froufe, E., Machado, J., 2013. Reproductive Cycle and Strategy of *Anodonta anatina* (L., 1758): Notes on Hermaphroditism. J Exp Zool A Ecol Genet Physiol 319, 378–390. https://doi.org/10.1002/JEZ.1801

Hipp, A.L., Eaton, D.A.R., Cavender-Bares, J., Fitzek, E., Nipper, R., Manos, P.S., 2014. A Framework Phylogeny of the American Oak Clade Based on Sequenced RAD Data. PLoS One 9, e93975. https://doi.org/10.1371/JOURNAL.PONE.0093975

Hoeh, W.R., Stewart, D.T., Sutherland, B.W., Zouros, E., 1996. MULTIPLE ORIGINS OF GENDER-ASSOCIATED MITOCHONDRIAL DNA LINEAGES IN BIVALVES (MOLLUSCA: BIVALVIA). Evolution (N Y) 50, 2276–2286. https://doi.org/10.1111/j.1558-5646.1996.tb03616.x

Hoff, K.J., Lange, S., Lomsadze, A., Borodovsky, M., Stanke, M., 2016. BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS: Table 1. Bioinformatics 32, 767–769. https://doi.org/10.1093/bioinformatics/btv661

Hoff, K.J., Lomsadze, A., Borodovsky, M., Stanke, M., 2019. Whole-genome annotation with BRAKER, in: Methods in Molecular Biology. Humana Press Inc., pp. 65–95. https://doi.org/10.1007/978-1-4939-9173-0_5

Hoffmann, R.J., Boore, J.L., Brown, W.M., 1992. A novel mitochondrial genome organization for the blue mussel, *Mytilus edulis*. Genetics 131, 397–412. https://doi.org/10.1093/GENETICS/131.2.397

Hohenlohe, P.A., Funk, W.C., Rajora, O.P., 2021. Population genomics for wildlife conservation and management. Mol Ecol 30, 62–82. https://doi.org/10.1111/MEC.15720

Holland, P.W.H., 2013. Evolution of homeobox genes. Wiley Interdiscip Rev Dev Biol. https://doi.org/10.1002/wdev.78

Hollenbeck, C.M., Johnston, I.A., 2018. Genomic Tools and Selective Breeding in Molluscs. Front Genet 9, 253. https://doi.org/10.3389/fgene.2018.00253

Holley, R.W., Apgar, J., Everett, G.A., Madison, J.T., Marquisee, M., Merrill, S.H., Penswick, J.R., Zamir, A., 1965. Structure of a Ribonucleic Acid. Science (1979) 147, 1462–1465. https://doi.org/10.1126/SCIENCE.147.3664.1462

Hosner, P.A., Faircloth, B.C., Glenn, T.C., Braun, E.L., Kimball, R.T., 2016. Avoiding Missing Data Biases in Phylogenomic Inference: An Empirical Study in the Landfowl (Aves: Galliformes). Mol Biol Evol 33, 1110–1125. https://doi.org/10.1093/MOLBEV/MSV347

Hotaling, S., Kelley, J.L., Frandsen, P.B., 2021. Toward a genome sequence for every animal: Where are we now? Proceedings of the National Academy of Sciences 118, e2109019118. https://doi.org/10.1073/PNAS.2109019118

Houston, D.D., Satler, J.D., Stack, T.K., Carroll, H.M., Bevan, A.M., Moya, A.L., Alexander, K.D., 2022. A phylogenomic perspective on the evolutionary history of the stonefly genus *Suwallia* (Plecoptera: Chloroperlidae) revealed by ultraconserved genomic elements. Mol Phylogenet Evol 166, 107320. https://doi.org/10.1016/J.YMPEV.2021.107320

Houston, R.D., Bean, T.P., Macqueen, D.J., Gundappa, M.K., Jin, Y.H., Jenkins, T.L., Selly, S.L.C., Martin, S.A.M., Stevens, J.R., Santos, E.M., Davie, A., Robledo, D., 2020. Harnessing genomics to fast-track genetic improvement in aquaculture. Nat Rev Genet 21, 389–409. https://doi.org/10.1038/s41576-020-0227-y

Howard, J.K., Cuffey, K.M., 2006. The functional role of native freshwater mussels in the fluvial benthic environment. Freshw Biol 51, 460–474. https://doi.org/10.1111/j.1365-2427.2005.01507.x

Hu, J., Fan, J., Sun, Z., Liu, S., 2019. NextPolish: a fast and efficient genome polishing tool for long-read assembly. Bioinformatics 36, 2253–2255. https://doi.org/10.1093/bioinformatics/btz891

Huan, P., Wang, Q., Tan, S., Liu, B., 2020. Dorsoventral decoupling of Hox gene expression underpins the diversification of molluscs. Proc Natl Acad Sci U S A 117, 503–512. https://doi.org/10.1073/pnas.1907328117

Huang, D., Shen, J., Li, J., Bai, Z., 2019. Integrated transcriptome analysis of immunological responses in the pearl sac of the triangle sail mussel (*Hyriopsis cumingii*) after mantle implantation. Fish Shellfish Immunol 90, 385–394. https://doi.org/10.1016/j.fsi.2019.05.012

Huang, X.C., Rong, J., Liu, Y., Zhang, M.H., Wan, Y., Ouyang, S., Zhou, C.H., Wu, X.P., 2013. The Complete Maternally and Paternally Inherited Mitochondrial Genomes of the Endangered Freshwater Mussel *Solenaia carinatus* (Bivalvia: Unionidae) and Implications for Unionidae Taxonomy. PLoS One 8, e84352. https://doi.org/10.1371/JOURNAL.PONE.0084352

Huang, X.C., Su, J.H., Ouyang, J.X., Ouyang, S., Zhou, C.H., Wu, X.P., 2019. Towards a global phylogeny of freshwater mussels (Bivalvia: Unionida): Species delimitation of Chinese taxa, mitochondrial phylogenomics, and diversification patterns. Mol Phylogenet Evol 130, 45–59. https://doi.org/10.1016/J.YMPEV.2018.09.019

Huang, X.C., Wu, R.-W., An, C.-T., Xie, G.-L., Su, J.-H., Ouyang, S., Zhou, C.-H., Wu, X.-P., 2018. Reclassification of *Lamprotula rochechouartii* as *Margaritifera rochechouartii comb. nov.* (Bivalvia: Margaritiferidae) revealed by time-calibrated multi-locus phylogenetic analyses and mitochondrial phylogenomics of Unionoida. Mol Phylogenet Evol 120, 297–306. https://doi.org/10.1016/J.YMPEV.2017.12.017

Huff, S.W., Campbell, D., Gustafson, D.L., Lydeard, C., Altaba, C.R., Giribet, G., 2004. INVESTIGATIONS INTO THE PHYLOGENETIC RELATIONSHIPS OF FRESHWATER PEARL MUSSELS (BIVALVIA: MARGARITIFERIDAE) BASED ON MOLECULAR DATA: IMPLICATIONS FOR THEIR TAXONOMY AND BIOGEOGRAPHY. Journal of Molluscan Studies 70, 379–388. https://doi.org/10.1093/MOLLUS/70.4.379

Hughes, L.C., Ortí, G., Huang, Y., Sun, Y., Baldwin, C.C., Thompson, A.W., Arcila, D., Betancur-R, R., Li, C., Becker, L., Bellora, N., Zhao, X., Li, X., Wang, M., Fang, C., Xie, B., Zhou, Z., Huang, H., Chen, S., Venkatesh, B., Shi, Q., 2018. Comprehensive phylogeny of ray-finned fishes (Actinopterygii) based on transcriptomic and genomic data. Proc Natl Acad Sci U S A 115, 6249–6254. https://doi.org/10.1073/pnas.1719358115

Hurst, G.D.D., Jiggins, F.M., 2005. Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. Proceedings of the Royal Society B: Biological Sciences 272, 1525–1534. https://doi.org/10.1098/RSPB.2005.3056

Hutchison, C.A., 2007. DNA sequencing: bench to bedside and beyond. Nucleic Acids Res 35, 6227–6237. https://doi.org/10.1093/NAR/GKM688

Ilves, K.L., Torti, D., López-Fernández, H., 2018. Exon-based phylogenomics strengthens the phylogeny of Neotropical cichlids and identifies remaining conflicting clades (Cichliformes: Cichlidae: Cichlinae). Mol Phylogenet Evol 118, 232–243. https://doi.org/10.1016/J.YMPEV.2017.10.008

Inoue, K., Monroe, E.M., Elderkin, C.L., Berg, D.J., 2014. Phylogeographic and population genetic analyses reveal Pleistocene isolation followed by high gene flow in a wide ranging, but endangered, freshwater mussel. Heredity (Edinb) 112, 282–290.

Irisarri, I., Eernisse, D.J., Zardoya, R., 2014. Molecular phylogeny of Acanthochitonina (Mollusca: Polyplacophora: Chitonida): three new mitochondrial genomes, rearranged gene orders and systematics. J Nat Hist 48, 2825–2853. https://doi.org/10.1080/00222933.2014.963721

Irisarri, I., Uribe, J.E., Eernisse, D.J., Zardoya, R., 2020. A mitogenomic phylogeny of chitons (Mollusca: Polyplacophora). BMC Evol Biol 20, 22. https://doi.org/10.1186/s12862-019-1573-2

Jackman, S.D., Vandervalk, B.P., Mohamadi, H., Chu, J., Yeo, S., Hammond, S.A., Jahesh, G., Khan, H., Coombe, L., Warren, R.L., Birol, I., 2017. ABySS 2.0 : Supplementary material. Genome Res 27, 768–777. https://doi.org/10.1101/gr.214346.116.Freely

Jackson, D.A., Symons, R.H., Berg, P., 1972. Biochemical Method for Inserting New Genetic Information into DNA of Simian Virus 40: Circular SV40 DNA Molecules Containing Lambda Phage Genes and the Galactose Operon of *Escherichia coli*. Proceedings of the National Academy of Sciences 69, 2904–2909. https://doi.org/10.1073/pnas.69.10.2904

Jain, M., Koren, S., Miga, K.H., Quick, J., Rand, A.C., Sasani, T.A., Tyson, J.R., Beggs, A.D., Dilthey, A.T., Fiddes, I.T., Malla, S., Marriott, H., Nieto, T., O'Grady, J., Olsen, H.E., Pedersen, B.S., Rhie, A., Richardson, H., Quinlan, A.R., Snutch, T.P., Tee, L., Paten, B., Phillippy, A.M., Simpson, J.T., Loman, N.J., Loose, M., 2018a. Nanopore sequencing and assembly of a human genome with ultra-long reads. Nature Biotechnology 2018 36:4 36, 338–345. https://doi.org/10.1038/nbt.4060

Jain, M., Olsen, H.E., Turner, D.J., Stoddart, D., Bulazel, K. v., Paten, B., Haussler, D., Willard, H.F., Akeson, M., Miga, K.H., 2018b. Linear assembly of a human centromere on the Y chromosome. Nature Biotechnology 2018 36:4 36, 321–323. https://doi.org/10.1038/nbt.4109

James, T.Y., Stajich, J.E., Hittinger, C.T., Rokas, A., 2020. Toward a Fully Resolved Fungal Tree of Life. Annu Rev Microbiol 74, 291–313. https://doi.org/10.1146/annurev-micro-022020-051835

Ji, H., Xu, X., Jin, X., Yin, H., Luo, J., Liu, G., Zhao, Q., Chen, Z., Bu, W., Gao, S., 2019. Using high-resolution annotation of insect mitochondrial DNA to decipher tandem repeats in the control region. RNA Biol 16, 830–837. https://doi.org/10.1080/15476286.2019.1591035/SUPPL_FILE/KRNB_A_1591035_SM5647.TXT

Jiang, L., Ge, C., Liu, W., Wu, C., Zhu, A., 2015. Complete mitochondrial genome of the *Loligo duvaucelii*. https://doi.org/10.3109/19401736.2015.1046164 27, 2723–2724. https://doi.org/10.3109/19401736.2015.1046164

Johnson, P.D., Brown, K.M., 2011. The importance of microhabitat factors and habitat stability to the threatened Louisiana pearl shell, *Margaritifera hembeli* (Conrad). https://doi.org/10.1139/z99-196 78, 271–277. https://doi.org/10.1139/Z99-196

Johnston, P.R., Quijada, L., Smith, C.A., Baral, H.O., Hosoya, T., Baschien, C., Pärtel, K., Zhuang, W.Y., Haelewaters, D., Park, D., Carl, S., López-Giráldez, F., Wang, Z., Townsend, J.P., 2019. A multigene phylogeny toward a new phylogenetic classification of Leotiomycetes. IMA Fungus 10, 1–22. https://doi.org/10.1186/S43008-019-0002-X/FIGURES/6

Jou, W.M., Haegeman, G., Ysebaert, M., Fiers, W., 1972. Nucleotide Sequence of the Gene Coding for the Bacteriophage MS2 Coat Protein. Nature 1972 237:5350 237, 82–88. https://doi.org/10.1038/237082a0

Jung, J., Kim, J.I., Jeong, Y.S., Yi, G., 2018. AGORA: organellar genome annotation from the amino acid and nucleotide references. Bioinformatics 34, 2661–2663. https://doi.org/10.1093/BIOINFORMATICS/BTY196

Kaiser, A.D., Wu, R., 1968. Structure and Function of DNA Cohesive Ends. Cold Spring Harb Symp Quant Biol 33, 729–734. https://doi.org/10.1101/SQB.1968.033.01.083

Kajander, O.A., Rovio, A.T., Majamaa, K., Poulton, J., Spelbrink, J.N., Holt, I.J., Karhunen, P.J., Jacobs, H.T., 2000. Human mtDNA sublimons resemble rearranged mitochondrial genomes found in pathological states. Hum Mol Genet 9, 2821–2835. https://doi.org/10.1093/HMG/9.19.2821

Kaliuzhin, S., Beletsky, V., Miguel, E.S., Popkovitch, E., Fernández, C., Neves, R.J., Amaro, R., Longa, A., Johnson, T., Ziuganov, V., 2009. Life Span Variation of the Freshwater Pearl Shell: A Model Species for Testing Longevity Mechanisms in Animals. AMBIO: A Journal of the Human Environment 29, 102–105. https://doi.org/10.1579/0044-7447-29.2.102

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., Jermiin, L.S., 2017. ModelFinder: Fast model selection for accurate phylogenetic estimates. Nat Methods 14, 587–589. https://doi.org/10.1038/nmeth.4285

Kambara, H., Nishikawa, T., Katayama, Y., Yamaguchi, T., 1988. Optimization of Parameters in a DNA Sequenator Using Fluorescence Detection. Bio/Technology 1988 6:7 6, 816–821. https://doi.org/10.1038/nbt0788-816

Kanehisa, M., Goto, S., 2000. Yeast Biochemical Pathways. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28, 27–30. https://doi.org/10.1093/nar/28.1.27

Kapli, P., Yang, Z., Telford, M.J., 2020. Phylogenetic tree building in the genomic age. Nat Rev Genet 1–17. https://doi.org/10.1038/s41576-020-0233-0

Karatayev, A.Y., Burlakova, L.E., Padilla, D.K., 2015. Zebra versus quagga mussels: a review of their spread, population dynamics, and ecosystem impacts. Hydrobiologia 746, 97–112. https://doi.org/10.1007/s10750-014-1901-x

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol Biol Evol 30, 772–780. https://doi.org/10.1093/molbev/mst010

Kawashima, Y., Nishihara, H., Akasaki, T., Nikaido, M., Tsuchiya, K., Segawa, S., Okada, N., 2013. The complete mitochondrial genomes of deep-sea squid (*Bathyteuthis abyssicola*), bob-tail squid (*Semirossia patagonica*) and four giant cuttlefish (*Sepia apama*, *S. latimanus*, *S. lycidas* and *S. pharaonis*), and their application to the phylogenetic analysis of Decapodiformes. Mol Phylogenet Evol 69, 980–993. https://doi.org/10.1016/J.YMPEV.2013.06.007

Kay, E.A., 1995. The conservation biology of molluscs. International Union for Conservation of Nature and Natural Resources, Gland, Switzerland and Cambridge, UK.

Kenchington, E., MacDonald, B., Cao, L., Tsagkarakis, D., Zouros, E., 2002. Genetics of Mother-Dependent Sex Ratio in Blue Mussels (*Mytilus spp.*) and Implications for Doubly

Uniparental Inheritance of Mitochondrial DNA. Genetics 161, 1579–1588. https://doi.org/10.1093/GENETICS/161.4.1579

Kenchington, E.L., Hamilton, L., Cogswell, A., Zouros, E., 2009. Paternal mtDNA and Maleness Are Co-Inherited but Not Causally Linked in Mytilid Mussels. PLoS One 4, e6976. https://doi.org/10.1371/JOURNAL.PONE.0006976

Kenny, N.J., Namigai, E.K.O., Marlétaz, F., Hui, J.H.L., Shimeld, S.M., 2015. Draft genome assemblies and predicted microRNA complements of the intertidal lophotrochozoans *Patella vulgata* (Mollusca, Patellogastropoda) and *Spirobranchus (Pomatoceros) lamarcki* (Annelida, Serpulida). Mar Genomics 24, 139–146. https://doi.org/10.1016/J.MARGEN.2015.07.004

Kern, E.M.A., Kim, T., Park, J.-K., 2020. The Mitochondrial Genome in Nematode Phylogenetics. Front Ecol Evol 8, 250. https://doi.org/10.3389/fevo.2020.00250

Kijas, J., Hamilton, M., Botwright, N., King, H., McPherson, L., Krsinich, A., McWilliam, S., 2019. Genome Sequencing of Blacklip and Greenlip Abalone for Development and Validation of a SNP Based Genotyping Tool. Front Genet 9, 687. https://doi.org/10.3389/fgene.2018.00687

Kim, B.M., Kang, S., Ahn, D.H., Jung, S.H., Rhee, H., Yoo, J.S., Lee, J.E., Lee, S., Han, Y.H., Ryu, K. bin, Cho, S.J., Park, H., An, H.S., 2018. The genome of common long-arm octopus *Octopus minor*. Gigascience 7, 1–7. https://doi.org/10.1093/gigascience/giy119

Kim, D., Langmead, B., Salzberg, S.L., 2015. HISAT: A fast spliced aligner with low memory requirements. Nat Methods 12, 357–360. https://doi.org/10.1038/nmeth.3317

Kim, D., Song, L., Breitwieser, F.P., Salzberg, S.L., 2016. Centrifuge: rapid and sensitive classification of metagenomic sequences. Genome Res 26, 1721–1729. https://doi.org/10.1101/GR.210641.116

Klein, A.H., Ballard, K.R., Storey, K.B., Motti, C.A., Zhao, M., Cummins, S.F., 2019. Multi-omics investigations within the Phylum Mollusca, Class Gastropoda: from ecological application to breakthrough phylogenomic studies. Brief Funct Genomics 18, 377–394. https://doi.org/10.1093/bfgp/elz017

Klopfenstein, D. v., Zhang, L., Pedersen, B.S., Ramírez, F., Vesztrocy, A.W., Naldi, A., Mungall, C.J., Yunes, J.M., Botvinnik, O., Weigel, M., Dampier, W., Dessimoz, C., Flick, P., Tang, H., 2018. GOATOOLS: A Python library for Gene Ontology analyses. Scientific Reports 2018 8:1 8, 1–17. https://doi.org/10.1038/s41598-018-28948-z

Knyshov, A., Gordon, E.R.L., Weirauch, C., 2021. New alignment-based sequence extraction software (ALiBaSeq) and its utility for deep level phylogenetics. PeerJ 9, e11019. https://doi.org/10.7717/PEERJ.11019/SUPP-6

Koch, E.L., Morales, H.E., Larsson, J., Westram, A.M., Faria, R., Lemmon, A.R., Lemmon, E.M., Johannesson, K., Butlin, R.K., 2021. Genetic variation for adaptive traits is associated with polymorphic inversions in *Littorina saxatilis*. Evol Lett 5, 196–213. https://doi.org/10.1002/EVL3.227

Koch, L., Potenski, C., Trenkmann, M., 2021. Sequencing moves to the twenty-first century. Nature Research 2021.

Kocot, K.M., 2013. Recent Advances and Unanswered Questions in Deep Molluscan Phylogenetics *. Am Malacol Bull 31, 195–208. https://doi.org/10.4003/006.031.0112

Kocot, K.M., Aguilera, F., McDougall, C., Jackson, D.J., Degnan, B.M., 2016a. Sea shell diversity and rapidly evolving secretomes: Insights into the evolution of biomineralization. Front Zool. https://doi.org/10.1186/s12983-016-0155-z

Kocot, K.M., Cannon, J.T., Todt, C., Citarella, M.R., Kohn, A.B., Meyer, A., Santos, S.R., Schander, C., Moroz, L.L., Lieb, B., Halanych, K.M., 2011. Phylogenomics reveals deep molluscan relationships. Nature 477, 452–456. https://doi.org/10.1038/nature10382

Kocot, K.M., Jeffery, N.W., Mulligan, K., Halanych, K.M., Gregory, T.R., 2016b. Genome size estimates for Aplacophora, Polyplacophora and Scaphopoda: Small solenogasters and sizeable scaphopods. Journal of Molluscan Studies 82, 216–219. https://doi.org/10.1093/mollus/eyv054

Kocot, K.M., Poustka, A.J., Stöger, I., Halanych, K.M., Schrödl, M., 2020. New data from Monoplacophora and a carefully-curated dataset resolve molluscan relationships. Sci Rep 10, 101. https://doi.org/10.1038/s41598-019-56728-w

Kocot, K.M., Struck, T.H., Merkel, J., Waits, D.S., Todt, C., Brannock, P.M., Weese, D.A., Cannon, J.T., Moroz, L.L., Lieb, B., Halanych, K.M., 2016c. Phylogenomics of Lophotrochozoa with Consideration of Systematic Error. Syst Biol 66, syw079. https://doi.org/10.1093/sysbio/syw079

Kocot, K.M., Todt, C., Mikkelsen, N.T., Halanych, K.M., 2019a. Phylogenomics of Aplacophora (Mollusca, Aculifera) and a solenogaster without a foot. Proceedings of the Royal Society B: Biological Sciences 286, 20190115. https://doi.org/10.1098/rspb.2019.0115

Kocot, K.M., Wollesen, T., Varney, R.M., Schwartz, M.L., Steiner, G., Wanninger, A., 2019b. Complete mitochondrial genomes of two scaphopod molluscs. Mitochondrial DNA B Resour 4, 3161–3162. https://doi.org/10.1080/23802359.2019.1666689

Kolesnikov, A.A., 2016. The mitochondrial genome. The nucleoid. Biochemistry (Moscow). https://doi.org/10.1134/S0006297916100047

Kondo, T., Kobayashi, O., 2005. Revision of the Genus *Margaritifera* (Bivalvia: Margaritiferidae) of Japan, with Description of a New Species. Venus (Journal of the Malacological Society of Japan) 64, 135–140. https://doi.org/10.18941/VENUS.64.3-4_135

Kong, L., Li, Y., Kocot, K.M., Yang, Y., Qi, L., Li, Q., Halanych, K.M., 2020. Mitogenomics reveals phylogenetic relationships of *Arcoida* (Mollusca, Bivalvia) and multiple independent expansions and contractions in mitochondrial genome size. Mol Phylogenet Evol 150, 106857. https://doi.org/10.1016/J.YMPEV.2020.106857

Kriventseva, E. v., Kuznetsov, D., Tegenfeldt, F., Manni, M., Dias, R., Simão, F.A., 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 Res 47, D807–D811. https://doi.org/10.1093/NAR/GKY1053

Kück, P., Longo, G.C., 2014. FASconCAT-G: Extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. Front Zool 11, 1–8. https://doi.org/10.1186/s12983-014-0081-x

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol 33, 1870–1874. https://doi.org/10.1093/molbev/msw054

Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G., Azevedo, V., Bertero, M.G., Bessières, P., Bolotin, A., Borchert, S., Borriss, R., Boursier, L., Brans, A., Braun, M., Brignell, S.C., Bron, S., Brouillet, S., Bruschi, C. v., Caldwell, B., Capuano, V., Carter, N.M., Choi, S.-K., Codani, J.-J., Connerton, I.F., Cummings, N.J., Daniel, R.A., Denizot, F., Devine, K.M., Düsterhöft, A., Ehrlich, S.D., Emmerson, P.T., Entian, K.D., Errington, J., Fabret, C., Ferrari, E., Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A., Galleron, N., Ghim, S.-Y., Glaser, P., Goffeau, A., Golightly, E.J., Grandi, G., Guiseppi, G., Guy, B.J., Haga, K., Haiech, J., Harwood, C.R., Hénaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M.-F., Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y., Klaerr-Blanchard, M., Klein, C., Kobayashi, Y., Koetter, P., Koningstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A., Lardinois, S., Lauber, J., Lazarevic, V., Lee, S.-M., Levine, A., Liu, H., Masuda, S., Mauël, C., Médigue, C., Medina, N., Mellado, R.P., Mizuno, M., Moestl, D., Nakai, S., Noback, M., Noone, D., O’Reilly, M., Ogawa, K., Ogiwara, A., Oudega, B., Park, S.-H., Parro, V., Pohl, T.M., Portetelle, D., Porwollik, S., Prescott, A.M., Presecan, E., Pujic, P., Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M., Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadaie, Y., Sato, T., Scanlan, E., Schleich, S., Schroeter, R., Scoffone, F., Sekiguchi, J., Sekowska, A., Seror, S.J., Serror, P., Shin, B.-S., Soldo, B., Sorokin, A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K., Takeuchi, M., Tamakoshi, A., Tanaka, T., Terpstra, P., Tognoni, A., Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassarotti, A., Viari, A., Wambutt, R., Wedler, E., Wedler, H., Weitzenegger, T., Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K., Yoshida, K., Yoshikawa, H.-F., Zumstein, E., Yoshikawa, H., Danchin, A., 1997. The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. Nature 390, 249–256. https://doi.org/10.1038/36786

Kurabayashi, A., Ueshima, R., 2000. Complete Sequence of the Mitochondrial DNA of the Primitive Opisthobranch Gastropod *Pupa strigosa*: Systematic Implication of the Genome Organization. Mol Biol Evol 17, 266–277. https://doi.org/10.1093/OXFORDJOURNALS.MOLBEV.A026306

Kyriakou, E., Kravariti, L., Vasilopoulos, T., Zouros, E., Rodakis, G.C., 2015. A protein binding site in the M mitochondrial genome of *Mytilus galloprovincialis* may be responsible for its paternal transmission. Gene 562, 83–94. https://doi.org/10.1016/J.GENE.2015.02.047

Kyriakou, E., Zouros, E., Rodakis, G.C., 2010. The atypical presence of the paternal mitochondrial DNA in somatic tissues of male and female individuals of the blue mussel species *Mytilus galloprovincialis*. BMC Res Notes 3, 1–6. https://doi.org/10.1186/1756-0500-3-222/METRICS

la Roche, J., Snyder, M., Cook, D.I., Fuller, K., Zouros, E., 1990. Molecular characterization of a repeat element causing large-scale size variation in the mitochondrial DNA of the sea scallop *Placopecten magellanicus*. Mol Biol Evol 7, 45–64. https://doi.org/10.1093/OXFORDJOURNALS.MOLBEV.A040586

Laetsch, D.R., Blaxter, M.L., 2017. BlobTools: Interrogation of genome assemblies. F1000Res 6, 1287. https://doi.org/10.12688/f1000research.12232.1

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., Fitzhugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., Levine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J.P., Miranda, C., Morris, W., Naylor, J., Raymond, Christina, Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J.C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, J.D., Marra, M.A., Mardis, E.R., Fulton, L.A., Chinwalla, A.T., Pepin, K.H., Gish, W.R., Chissoe, S.L., Wendl, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R.A., Muzny, D.M., Scherer, S.E., Bouck, J.B., Sodergren, E.J., Worley, K.C., Rives, C.M., Gorrell, J.H., Metzker, M.L., Naylor, S.L., Kucherlapati, R.S., Nelson, D.L., Weinstock, G.M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Smith, D.R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Hong, M.L., Dubois, J., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R.W., Federspiel, N.A., Abola, A.P., Proctor, M.J., Roe, B.A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W.R., de La Bastide, M., Dedhia, N., Blöcker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J.A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D.G., Burge, C.B., Cerutti, L., Chen, H.C., Church, D., Clamp, M., Copley, R.R., Doerks, T., Eddy, S.R., Eichler, E.E., Furey, T.S., Galagan, J., Gilbert, J.G.R., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L.S., Jones, T.A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W.J., Kitts, P., Koonin, E. v., Korf, I., Kulp, D., Lancet, D., Lowe, T.M., McLysaght, A., Mikkelsen, T., Moran, J. v., Mulder, N., Pollara, V.J., Ponting, C.P., Schuler, G., Schultz, J., Slater, G., Smit, A.F.A., Stupka, E., Szustakowki, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y.I., Wolfe, K.H., Yang, S.P., Yeh, R.F., Collins, F., Guyer, M.S., Peterson, J., Felsenfeld, A., Wetterstrand, K.A., Myers, R.M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D.R., Olson, M. v., Kaul, R., Raymond, Christopher, Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G.A., Athanasiou, M., Schultz, R., Patrinos, A., Morgan, M.J., 2001. Initial sequencing and analysis of the human genome. Nature 2001 409:6822 409, 860–921. https://doi.org/10.1038/35057062

Lanfear, R., Calcott, B., Kainer, D., Mayer, C., Stamatakis, A., 2014. Selecting optimal partitioning schemes for phylogenomic datasets. BMC Evol Biol 14, 1–14. https://doi.org/10.1186/1471-2148-14-82/TABLES/3

Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B., 2017. Partitionfinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol Biol Evol 34, 772–773. https://doi.org/10.1093/molbev/msw260

Larsson, A., 2014. AliView: A fast and lightweight alignment viewer and editor for large datasets. Bioinformatics 30, 3276–3278. https://doi.org/10.1093/bioinformatics/btu531

Lauri, A., Pompilio, G., Capogrossi, M.C., 2014. The mitochondrial genome in aging and senescence. Ageing Res Rev 18, 1–15. https://doi.org/10.1016/J.ARR.2014.07.001

Lavrov, D. v., Pett, W., 2016. Animal Mitochondrial DNA as We Do Not Know It: mt-Genome Organization and Evolution in Nonbilaterian Lineages. Genome Biol Evol 8, 2896–2913. https://doi.org/10.1093/GBE/EVW195

Leebens-Mack, J.H., Barker, M.S., Carpenter, E.J., Deyholos, M.K., Gitzendanner, M.A., Graham, S.W., Grosse, I., Li, Z., Melkonian, M., Mirarab, S., Porsch, M., Quint, M., Rensing, S.A., Soltis, D.E., Soltis, P.S., Stevenson, D.W., Ullrich, K.K., Wickett, N.J., DeGironimo, L., Edger, P.P., Jordon-Thaden, I.E., Joya, S., Liu, T., Melkonian, B., Miles, N.W., Pokorny, L., Quigley, C., Thomas, P., Villarreal, J.C., Augustin, M.M., Barrett, M.D., Baucom, R.S., Beerling, D.J., Benstein, R.M., Biffin, E., Brockington, S.F., Burge, D.O., Burris, J.N., Burris, K.P., Burtet-Sarramegna, V., Caicedo, A.L., Cannon, S.B., Çebi, Z., Chang, Y., Chater, C., Cheeseman, J.M., Chen, T., Clarke, N.D., Clayton, H., Covshoff, S., Crandall-Stotler, B.J., Cross, H., dePamphilis, C.W., Der, J.P., Determann, R., Dickson, R.C., di Stilio, V.S., Ellis, S., Fast, E., Feja, N., Field, K.J., Filatov, D.A., Finnegan, P.M., Floyd, S.K., Fogliani, B., García, N., Gâteblé, G., Godden, G.T., Goh, F. (Qi Y., Greiner, S., Harkess, A., Heaney, J.M., Helliwell, K.E., Heyduk, K., Hibberd, J.M., Hodel, R.G.J., Hollingsworth, P.M., Johnson, M.T.J., Jost, R., Joyce, B., Kapralov, M. v., Kazamia, E., Kellogg, E.A., Koch, M.A., von Konrat, M., Könyves, K., Kutchan, T.M., Lam, V., Larsson, A., Leitch, A.R., Lentz, R., Li, F.W., Lowe, A.J., Ludwig, M., Manos, P.S., Mavrodiev, E., McCormick, M.K., McKain, M., McLellan, T., McNeal, J.R., Miller, R.E., Nelson, M.N., Peng, Y., Ralph, P., Real, D., Riggins, C.W., Ruhsam, M., Sage, R.F., Sakai, A.K., Scascitella, M., Schilling, E.E., Schlösser, E.M., Sederoff, H., Servick, S., Sessa, E.B., Shaw, A.J., Shaw, S.W., Sigel, E.M., Skema, C., Smith, A.G., Smithson, A., Stewart, C.N., Stinchcombe, J.R., Szövényi, P., Tate, J.A., Tiebel, H., Trapnell, D., Villegente, M., Wang, C.N., Weller, S.G., Wenzel, M., Weststrand, S., Westwood, J.H., Whigham, D.F., Wu, S., Wulff, A.S., Yang, Y., Zhu, D., Zhuang, C., Zuidof, J., Chase, M.W., Pires, J.C., Rothfels, C.J., Yu, J., Chen, C., Chen, L., Cheng, S., Li, J., Li, R., Li, X., Lu, H., Ou, Y., Sun, X., Tan, X., Tang, J., Tian, Z., Wang, F., Wang, J., Wei, X., Xu, X., Yan, Z., Yang, F., Zhong, X., Zhou, F., Zhu, Y., Zhang, Y., Ayyampalayam, S., Barkman, T.J., Nguyen, N. phuong, Matasci, N., Nelson, D.R., Sayyari, E., Wafula, E.K., Walls, R.L., Warnow, T., An, H., Arrigo, N., Baniaga, A.E., Galuska, S., Jorgensen, S.A., Kidder, T.I., Kong, H., Lu-Irving, P., Marx, H.E., Qi, X., Reardon, C.R., Sutherland, B.L., Tiley, G.P., Welles, S.R., Yu, R., Zhan, S., Gramzow, L., Theißen, G., Wong, G.K.S., 2019. One thousand plant transcriptomes and the phylogenomics of green plants. Nature 2019 574:7780 574, 679–685. https://doi.org/10.1038/s41586-019-1693-2

Lehner, B., Grill, G., 2013. Global river hydrography and network routing: Baseline data and new approaches to study the world's large river systems. Hydrol Process 27, 2171–2186. https://doi.org/10.1002/hyp.9740

Lemer, S., Bieler, R., Giribet, G., 2019. Resolving the relationships of clams and cockles: Dense transcriptome sampling drastically improves the bivalve tree of life. Proceedings of the Royal Society B: Biological Sciences 286, 20182684. https://doi.org/10.1098/rspb.2018.2684

Lemer, S., González, V.L., Bieler, R., Giribet, G., 2016. Cementing mussels to oysters in the pteriomorphian tree: a phylogenomic approach. Proceedings of the Royal Society B: Biological Sciences 283, 20160857. https://doi.org/10.1098/rspb.2016.0857

Lemmon, A.R., Emme, S.A., Lemmon, E.M., 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. Syst Biol 61, 727–744. https://doi.org/10.1093/sysbio/sys049

Lemmon, E.M., Lemmon, A.R., 2013. High-throughput genomic data in systematics and phylogenetics. Annu Rev Ecol Evol Syst. https://doi.org/10.1146/annurev-ecolsys-110512-135822

Letts, J.A., Fiedorczuk, K., Sazanov, L.A., 2016. The architecture of respiratory supercomplexes. Nature 2016 537:7622 537, 644–648. https://doi.org/10.1038/nature19774

Levene, H.J., Korlach, J., Turner, S.W., Foquet, M., Craighead, H.G., Webb, W.W., 2003. Zero-mode waveguides for single-molecule analysis at high concentrations. Science (1979) 299, 682–686. https://doi.org/10.1126/SCIENCE.1079700/SUPPL_FILE/LEVENE.SOM.PDF

Levinson, G., Gutman, G.A., 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol Biol Evol 4, 203–221. https://doi.org/10.1093/OXFORDJOURNALS.MOLBEV.A040442

Lewin, H.A., Robinson, G.E., Kress, W.J., Baker, W.J., Coddington, J., Crandall, K.A., Durbin, R., Edwards, S. v., Forest, F., Gilbert, M.T.P., Goldstein, M.M., Grigoriev, I. v., Hackett, K.J., Haussler, D., Jarvis, E.D., Johnson, W.E., Patrinos, A., Richards, S., Castilla-Rubio, J.C., van Sluys, M.A., Soltis, P.S., Xu, X., Yang, H., Zhang, G., 2018. Earth BioGenome Project: Sequencing life for the future of life. Proc Natl Acad Sci U S A 115, 4325–4333. https://doi.org/10.1073/PNAS.1720115115/SUPPL_FILE/PNAS.1720115115.SAPP.PDF

Li, A., Dai, H., Guo, X., Zhang, Z., Zhang, K., Wang, C., Wang, X., Wang, W., Chen, H., Li, X., Zheng, H., Li, L., Zhang, G., 2021. Genome of the estuarine oyster provides insights into climate impact and adaptive plasticity. Communications Biology 2021 4:1 4, 1–12. https://doi.org/10.1038/s42003-021-02823-6

Li, C., Liu, X., Liu, B., Ma, B., Liu, F., Liu, G., Shi, Q., Wang, C., 2018. Draft genome of the Peruvian scallop *Argopecten purpuratus*. Gigascience 7. https://doi.org/10.1093/gigascience/giy031

Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.

Li, Y., Nong, W., Baril, T., Yip, H.Y., Swale, T., Hayward, A., Ferrier, D.E.K., Hui, J.H.L., 2020. Reconstruction of ancient homeobox gene linkages inferred from a new high-quality assembly of the Hong Kong oyster (*Magallana hongkongensis*) genome. BMC Genomics 21, 713. https://doi.org/10.1186/s12864-020-07027-6

Li, Y., Ren, Z., Shedlock, A.M., Wu, J., Sang, L., Tersing, T., Hasegawa, M., Yonezawa, T., Zhong, Y., 2013. High altitude adaptation of the schizothoracine fishes (Cyprinidae) revealed by the mitochondrial genome analyses. Gene 517, 169–178. https://doi.org/10.1016/J.GENE.2012.12.096

Li, Y., Steenwyk, J.L., Chang, Y., Wang, Y., James, T.Y., Stajich, J.E., Spatafora, J.W., Groenewald, M., Dunn, C.W., Hittinger, C.T., Shen, X.X., Rokas, A., 2021. A genome-scale phylogeny of the kingdom Fungi. Current Biology 31, 1653-1665.e5. https://doi.org/10.1016/J.CUB.2021.01.074

Li, Yuli, Sun, X., Hu, X., Xun, X., Zhang, J., Guo, X., Jiao, W., Zhang, Lingling, Liu, W., Wang, J., Li, J., Sun, Y., Miao, Y., Zhang, X., Cheng, T., Xu, G., Fu, X., Wang, Y., Yu, X., Huang, X., Lu, W., Lv, J., Mu, C., Wang, D., Li, X., Xia, Y., Li, Yajuan, Yang, Z., Wang, F., Zhang, Lu, Xing, Q., Dou, H., Ning, X., Dou, J., Li, Yangping, Kong, D., Liu, Y., Jiang, Z., Li, R., Wang, S., Bao, Z., 2017. Scallop genome reveals molecular adaptations to semi-sessile life and neurotoxins. Nat Commun 8, 1721. https://doi.org/10.1038/s41467-017-01927-0

Liu, C., Ren, Y., Li, Z., Hu, Q., Yin, L., Wang, H., Qiao, X., Zhang, Y., Xing, L., Xi, Y., Jiang, F., Wang, S., Huang, C., Liu, B., Liu, H., Wan, F., Qian, W., Fan, W., 2021. Giant African snail genomes provide insights into molluscan whole-genome duplication and aquatic–terrestrial transition. Mol Ecol Resour 21, 478–494. https://doi.org/10.1111/1755-0998.13261

Liu, C., Zhang, Y., Ren, Y., Wang, H., Li, S., Jiang, F., Yin, L., Qiao, X., Zhang, G., Qian, W., Liu, B., Fan, W., 2018. The genome of the golden apple snail *Pomacea canaliculata* provides insight into stress tolerance and invasive adaptation. Gigascience 7. https://doi.org/10.1093/gigascience/giy101

Liu, Y.G., Kurokawa, T., Sekino, M., Tanabe, T., Watanabe, K., 2013. Complete mitochondrial DNA sequence of the ark shell *Scapharca broughtonii*: An ultra-large metazoan mitochondrial genome. Comp Biochem Physiol Part D Genomics Proteomics 8, 72–81. https://doi.org/10.1016/J.CBD.2012.12.003

Lobachev, K.S., Shor, B.M., Tran, H.T., Taylor, W., Keen, J.D., Resnick, M.A., Gordenin, D.A., 1998. Factors Affecting Inverted Repeat Stimulation of Recombination and Deletion in *Saccharomyces cerevisiae*. Genetics 148, 1507–1524. https://doi.org/10.1093/GENETICS/148.4.1507

Lopes-Lima, M., Bolotov, I.N., Do, V.T., Aldridge, D.C., Fonseca, M.M., Gan, H.M., Gofarov, M.Y., Kondakov, A. v., Prié, V., Sousa, R., Varandas, S., Vikhrev, I. v., Teixeira, A., Wu, R.W., Wu, X., Zieritz, A., Froufe, E., Bogan, A.E., 2018a. Expansion and systematics redefinition of the most threatened freshwater mussel family, the Margaritiferidae. Mol Phylogenet Evol 127, 98–118. https://doi.org/10.1016/j.ympev.2018.04.041

Lopes-Lima, M., Burlakova, L.E., Karatayev, A.Y., Mehler, K., Seddon, M., Sousa, R., 2018b. Conservation of freshwater bivalves at the global scale: diversity, threats and research needs. Hydrobiologia 810, 1–14. https://doi.org/10.1007/S10750-017-3486-7/FIGURES/9

Lopes-Lima, M., Fonseca, M.M., Aldridge, D.C., Bogan, A.E., Gan, H.M., Ghamizi, M., Sousa, R., Teixeira, A., Varandas, S., Zanatta, D., Zieritz, A., Froufe, E., 2017a. The first Margaritiferidae male (M-type) mitogenome: mitochondrial gene order as a potential character for determining higher-order phylogeny within Unionida (Bivalvia). Journal of Molluscan Studies 83, 249–252. https://doi.org/10.1093/mollus/eyx009

Lopes-Lima, M., Froufe, E., Do, V.T., Ghamizi, M., Mock, K.E., Kebapçı, Ü., Klishko, O., Kovitvadhi, S., Kovitvadhi, U., Paulo, O.S., Pfeiffer, J.M., Raley, M., Riccardi, N., Şereflişan, H., Sousa, R., Teixeira, A., Varandas, S., Wu, X., Zanatta, D.T., Zieritz, A., Bogan, A.E., 2017b. Phylogeny of the most species-rich freshwater bivalve family (Bivalvia: Unionida: Unionidae): Defining modern subfamilies and tribes. Mol Phylogenet Evol 106, 174–191. https://doi.org/10.1016/J.YMPEV.2016.08.021

Lopes-Lima, M., Kebapçı, U., van Damme, D., 2014b. *Unio crassus* (Thick Shelled River Mussel) [WWW Document]. The IUCN Red List of Threatened Species. URL https://www.iucnredlist.org/species/22736/42465628 (accessed 5.20.22).

Lopes-Lima, M., Riccardi, N., Urbanska, M., Köhler, F., Vinarski, M., Bogan, A.E., Sousa, R., 2021. Major shortfalls impairing knowledge and conservation of freshwater molluscs. Hydrobiologia 2021 848:12 848, 2831–2867. https://doi.org/10.1007/S10750-021-04622-W

Lopes-Lima, M., Seddon, M.B., 2014. *Unio mancus*. The IUCN Red List of Threatened Species. URL https://www.iucnredlist.org/species/22737/42466471 (accessed 5.20.22).

Lopes-Lima, M., Sousa, R., Geist, J., Aldridge, D.C., Araujo, R., Bergengren, J., Bespalaya, Y., Bódis, E., Burlakova, L., van Damme, D., Douda, K., Froufe, E., Georgiev, D., Gumpinger, C., Karatayev, A., Kebapçi, Ü., Killeen, I., Lajtner, J., Larsen, B.M., Lauceri, R., Legakis, A., Lois, S., Lundberg, S., Moorkens, E., Motte, G., Nagel, K.O., Ondina, P., Outeiro, A., Paunovic, M., Prié, V., von Proschwitz, T., Riccardi, N., Rudzīte, M., Rudzītis, M., Scheder, C., Seddon, M., Şereflişan, H., Simić, V., Sokolova, S., Stoeckl, K., Taskinen, J., Teixeira, A., Thielen, F., Trichkova, T., Varandas, S., Vicentini, H., Zajac, K., Zajac, T., Zogaris, S., 2017c. Conservation status of freshwater mussels in Europe: state of the art and future challenges. Biological Reviews 92, 572–607. https://doi.org/10.1111/brv.12244

Lopes-Lima, Manuel, Teixeira, A., Froufe, E., Lopes, A., Varandas, S., Sousa, R., 2014a. Biology and conservation of freshwater bivalves: Past, present and future perspectives. Hydrobiologia. https://doi.org/10.1007/s10750-014-1902-9

Lopez, J. v., Bracken-Grissom, H., Collins, A.G., Collins, T., Crandall, K., Distel, D., Dunn, C., Giribet, G., Haddock, S., Knowlton, N., Martindale, M., Medina, M., Messing, C., O’Brien, S.J., Paulay, G., Putnam, N., Ravasi, T., Rouse, G.W., Ryan, J.F., Schulze, A., WÃ¶rheide, G., Adamska, M., Bailly, X., Breinholt, J., Browne, W.E., Diaz, M.C., Evans, N., Flot, J.F., Fogarty, N., Johnston, M., Kamel, B., Kawahara, A.Y., Laberge, T., Lavrov, D., Michonneau, F., Moroz, L.L., Oakley, T., Osborne, K., Pomponi, S.A., Rhodes, A., Rodriguez-Lanetty, M., Santos, S.R., Satoh, N., Thacker, R.W., van de Peer, Y., Voolstra, C.R., Welch, D.M., Winston, J., Zhou, X., 2014. The Global Invertebrate Genomics Alliance (GIGA): Developing Community Resources to Study Diverse Invertebrate Genomes. Journal of Heredity. https://doi.org/10.1093/jhered/est084

Lopez, J. v., Kamel, B., Medina, M., Collins, T., Baums, I.B., 2019. Multiple Facets of Marine Invertebrate Conservation Genomics. https://doi.org/10.1146/annurev-animal-020518-115034 7, 473–497. https://doi.org/10.1146/ANNUREV-ANIMAL-020518-115034

Lowe, T.M., Chan, P.P., 2016. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res 44, W54–W57. https://doi.org/10.1093/NAR/GKW413

Lu, S., Wang, J., Chitsaz, F., Derbyshire, M.K., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Marchler, G.H., Song, J.S., Thanki, N., Yamashita, R.A., Yang, M., Zhang, D., Zheng, C., Lanczycki, C.J., Marchler-Bauer, A., 2020. CDD/SPARCLE: The conserved domain database in 2020. Nucleic Acids Res 48, D265–D268. https://doi.org/10.1093/nar/gkz991

Luckey, J.A., Drossman, H., Kostichka, A.J., Mead, D.A., D’cunha, J., Norris, T.B., Smith, L.M., 1990. High speed DNA sequencing by capillary electrophoresis. Nucleic Acids Res 18, 4417–4421. https://doi.org/10.1093/NAR/18.15.4417

Ludwig, A., May, B., Debus, L., Jenneckens, I., 2000. Heteroplasmy in the mtDNA Control Region of Sturgeon (*Acipenser*, *Huso* and *Scaphirhynchus*). Genetics 156, 1933–1947. https://doi.org/10.1093/GENETICS/156.4.1933

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Yunjie, Tang, J., Wu, G., Zhang, H., Shi, Y., Liu, Yong, Yu, C., Wang, B., Lu, Y., Han, C., Cheung, D.W., Yiu, S.-M., Peng, S., Xiaoqian, Z., Liu, G., Liao, X., Li, Y., Yang, H., Wang, Jian, Lam, T.-W., Wang, Jun, 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience 1, 18. https://doi.org/10.1186/2047-217X-1-18

Luo, Y., Li, C., Landis, A.G., Wang, G., Stoeckel, J., Peatman, E., 2014. Transcriptomic Profiling of Differential Responses to Drought in Two Freshwater Mussel Species, the Giant Floater *Pyganodon grandis* and the *Pondhorn Uniomerus tetralasmus*. PLoS One 9, e89481. https://doi.org/10.1371/JOURNAL.PONE.0089481

Luo, Y., Yang, X., Gao, Y., 2013. Mitochondrial DNA response to high altitude: A new perspective on high-altitude adaptation. http://dx.doi.org/10.3109/19401736.2012.760558 24, 313–319. https://doi.org/10.3109/19401736.2012.760558

Luo, Y.J., Satoh, N., Endo, K., 2015. Mitochondrial gene order variation in the brachiopod *Lingula anatina* and its implications for mitochondrial evolution in lophotrochozoans. Mar Genomics 24, 31–40. https://doi.org/10.1016/J.MARGEN.2015.08.005

Lydeard, C., Cowie, R.H., Ponder, W.F., Bogan, A.E., Bouchet, P., Clark, S.A., Cummings, K.S., Frest, T.J., Gargominy, O., Herbert, D.G., Hershler, R., Perez, K.E., Roth, B., Seddon, M., Strong, E.E., Thompson, F.G., 2006. The Global Decline of Nonmarine

Mollusks. Bioscience 54, 321. https://doi.org/https://doi.org/10.1641/0006-3568(2004)054[0321:TGDONM]2.0.CO;2

Macey, J.R., Larson, A., Ananjeva, N.B., Fang, Z., Papenfuss, T.J., 1997. Two novel gene orders and the role of light-strand replication in rearrangement of the vertebrate mitochondrial genome. Mol Biol Evol 14, 91–104. https://doi.org/10.1093/OXFORDJOURNALS.MOLBEV.A025706

Machado, A.M., Fernández-Boo, S., Nande, M., Pinto, R., Costas, B., Castro, L.F.C., 2022. The male and female gonad transcriptome of the edible sea urchin, *Paracentrotus lividus*: Identification of sex-related and lipid biosynthesis genes. Aquac Rep 22, 100936. https://doi.org/10.1016/J.AQREP.2021.100936

Machado, A.M., Muñoz-Merida, A., Fonseca, E., Veríssimo, A., Pinto, R., Felício, M., da Fonseca, R.R., Froufe, E., Castro, L.F.C., 2020. Liver transcriptome resources of four commercially exploited teleost species. Sci Data 7, 1–9. https://doi.org/10.1038/s41597-020-0565-9

Maddison, W.P., Evans, S.C., Hamilton, C.A., Bond, J.E., Lemmon, A.R., Lemmon, E.M., 2017. A genome-wide phylogeny of jumping spiders (Araneae, Salticidae), using anchored hybrid enrichment. ZooKeys 695: 89-101 695, 89–101. https://doi.org/10.3897/ZOOKEYS.695.13852

Makhrov, A., Bespalaya, J., Bolotov, I., Vikhrev, I., Gofarov, M., Alekseeva, Y., Zotin, A., 2014. Historical geography of pearl harvesting and current status of populations of freshwater pearl mussel *Margaritifera margaritifera* (L.) in the western part of Northern European Russia. Hydrobiologia. https://doi.org/10.1007/s10750-013-1546-1

Manni, M., Berkeley, M.R., Seppey, M., Simão, F.A., Zdobnov, E.M., 2021. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. Mol Biol Evol 38, 4647–4654. https://doi.org/10.1093/molbev/msab199

Mapleson, D., Accinelli, G.G., Kettleborough, G., Wright, J., Clavijo, B.J., 2017. KAT: A K-mer analysis toolkit to quality control NGS datasets and genome assemblies. Bioinformatics 33, 574–576. https://doi.org/10.1093/bioinformatics/btw663

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z., Dewell, S.B., Du, L., Fierro, J.M., Gomes, X. v., Godwin, B.C., He, W., Helgesen, S., Ho, C.H., Irzyk, G.P., Jando, S.C., Alenquer, M.L.I., Jarvie, T.P., Jirage, K.B., Kim, J.B., Knight, J.R., Lanza, J.R., Leamon, J.H., Lefkowitz, S.M., Lei, M., Li, J., Lohman, K.L., Lu, H., Makhijani, V.B., McDade, K.E., McKenna, M.P., Myers, E.W., Nickerson, E., Nobile, J.R., Plant, R., Puc, B.P., Ronan, M.T., Roth, G.T., Sarkis, G.J., Simons, J.F., Simpson, J.W., Srinivasan, M., Tartaro, K.R., Tomasz, A., Vogt, K.A., Volkmer, G.A., Wang, S.H., Wang, Y., Weiner, M.P., Yu, P., Begley, R.F., Rothberg, J.M., 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 2005 437:7057 437, 376–380. https://doi.org/10.1038/nature03959

Marin, F., 2020. Mollusc shellomes: Past, present and future. J Struct Biol 212, 107583. https://doi.org/10.1016/J.JSB.2020.107583

Márquez, E.J., Castro, E.R., Alzate, J.F., 2014. Mitochondrial genome of the endangered marine gastropod *Strombus gigas* Linnaeus, 1758 (Mollusca: Gastropoda). http://dx.doi.org/10.3109/19401736.2014.953118 27, 1516–1517. https://doi.org/10.3109/19401736.2014.953118

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 17, 10. https://doi.org/10.14806/ej.17.1.200

Masonbrink, R.E., Purcell, C.M., Boles, S.E., Whitehead, A., Hyde, J.R., Seetharam, A.S., Severin, A.J., 2019. An annotated genome for *Haliotis rufescens* (Red Abalone) and resequenced green, pink, pinto, black, and white abalone species. Genome Biol Evol 11, 431–438. https://doi.org/10.1093/gbe/evz006

Masta, S.E., Longhorn, S.J., Boore, J.L., 2009. Arachnid relationships based on mitochondrial genomes: Asymmetric nucleotide and amino acid bias affects phylogenetic analyses. Mol Phylogenet Evol 50, 117–128. https://doi.org/10.1016/J.YMPEV.2008.10.010

Maxam, A.M., Gilbert, W., 1977. A new method for sequencing DNA. Proceedings of the National Academy of Sciences 74, 560–564. https://doi.org/10.1073/PNAS.74.2.560

Mayer, P., Farinelli, L., Kawashima, E.H., 1998. Method of nucleic acid amplification. WO1998044151A1.

Mayjonade, B., Gouzy, J., Donnadieu, C., Pouilly, N., Marande, W., Callot, C., Langlade, N., Muños, S., 2017. Extraction of high-molecular-weight genomic DNA for long-read sequencing of single molecules. Biotechniques 62, xv. https://doi.org/10.2144/000114503

Maynard, B.T., Kerr, L.J., McKiernan, J.M., Jansen, E.S., Hanna, P.J., 2005. Mitochondrial DNA sequence and gene organization in Australian backup abalone *Haliotis rubra* (Leach). Marine Biotechnology 7, 645–658. https://doi.org/10.1007/S10126-005-0013-Z/FIGURES/3

McCarthy, T.W., Chou, H., Brendel, V.P., 2019. SRAssembler: Selective Recursive local Assembly of homologous genomic regions. BMC Bioinformatics 20, 371. https://doi.org/10.1186/s12859-019-2949-4

McCartney, M.A., Auch, B., Kono, T., Mallez, S., Zhang, Y., Obille, A., Becker, A., Abrahante, J.E., Garbe, J., Badalamenti, J.P., Herman, A., Mangelson, H., Liachko, I., Sullivan, S., Sone, E.D., Koren, S., Silverstein, K.A.T., Beckman, K.B., Gohl, D.M., 2022. The genome of the zebra mussel, *Dreissena polymorpha*: a resource for comparative genomics, invasion genetics, and biocontrol. G3 Genes|Genomes|Genetics 12. https://doi.org/10.1093/G3JOURNAL/JKAB423

McCartney, M.A., Mallez, S., Gohl, D.M., 2019. Genome projects in invasion biology. Conservation Genetics 20, 1201–1222. https://doi.org/10.1007/s10592-019-01224-x

McComish, B.J., Hills, S.F.K., Biggs, P.J., Penny, D., 2010. Index-Free De Novo Assembly and Deconvolution of Mixed Mitochondrial Genomes. Genome Biol Evol 2, 410–424. https://doi.org/10.1093/GBE/EVQ029

McGettigan, P.A., 2013. Transcriptomics in the RNA-seq era. Curr Opin Chem Biol 17, 4–11. https://doi.org/10.1016/J.CBPA.2012.12.008

Meek, M.H., Larson, W.A., 2019. The future is now: Amplicon sequencing and sequence capture usher in the conservation genomics era. Mol Ecol Resour 19, 795–803. https://doi.org/10.1111/1755-0998.12998

Mehrmohamadi, M., Sepehri, M.H., Nazer, N., Norouzi, M.R., 2021. A Comparative Overview of Epigenomic Profiling Methods. Front Cell Dev Biol 9, 1990. https://doi.org/10.3389/FCELL.2021.714687/XML/NLM

Meng, G., Li, Y., Yang, C., Liu, S., 2019. MitoZ: a toolkit for animal mitochondrial genome assembly, annotation and visualization. Nucleic Acids Res 47, e63–e63. https://doi.org/10.1093/nar/gkz173

Messing, J., Crea, R., Seeburg, P.H., 1981. A system for shotgun DNA sequencing. Nucleic Acids Res 9, 309–321. https://doi.org/10.1093/NAR/9.2.309

Meyer, A., Schloissnig, S., Franchini, P., Du, K., Woltering, J.M., Irisarri, I., Wong, W.Y., Nowoshilow, S., Kneitz, S., Kawaguchi, A., Fabrizius, A., Xiong, P., Dechaud, C., Spaink, H.P., Volff, J.N., Simakov, O., Burmester, T., Tanaka, E.M., Schartl, M., 2021. Giant lungfish genome elucidates the conquest of land by vertebrates. Nature 2021 590:7845 590, 284–289. https://doi.org/10.1038/s41586-021-03198-8

Mikkelsen, N.T., Kocot, K.M., Halanych, K.M., 2018. Mitogenomics reveals phylogenetic relationships of caudofoveate aplacophoran molluscs. Mol Phylogenet Evol 127, 429–436. https://doi.org/10.1016/j.ympev.2018.04.031

Mikkelsen, N.T., Todt, C., Kocot, K.M., Halanych, K.M., Willassen, E., 2019. Molecular phylogeny of Caudofoveata (Mollusca) challenges traditional views. Mol Phylogenet Evol 132, 138–150. https://doi.org/10.1016/j.ympev.2018.10.037

Milani, L., Ghiselli, F., Guerra, D., Breton, S., Passamonti, M., 2013a. A Comparative Analysis of Mitochondrial ORFans: New Clues on Their Origin and Role in Species with Doubly Uniparental Inheritance of Mitochondria. Genome Biol Evol 5, 1408–1434. https://doi.org/10.1093/GBE/EVT101

Milani, L., Ghiselli, F., Iannello, M., Passamonti, M., 2014a. Evidence for somatic transcription of male-transmitted mitochondrial genome in the DUI species *Ruditapes philippinarum* (Bivalvia: Veneridae). Curr Genet 60, 163–173. https://doi.org/10.1007/S00294-014-0420-7/FIGURES/6

Milani, L., Ghiselli, F., Maurizii, M.G., Nuzhdin, S. v., Passamonti, M., 2014b. Paternally Transmitted Mitochondria Express a New Gene of Potential Viral Origin. Genome Biol Evol 6, 391–405. https://doi.org/10.1093/GBE/EVU021

Milani, L., Ghiselli, F., Nuzhdin, S. v., Passamonti, M., 2013b. Nuclear genes with sex bias in *Ruditapes philippinarum* (Bivalvia, veneridae): Mitochondrial inheritance and sex determination in DUI species. J Exp Zool B Mol Dev Evol 320, 442–454. https://doi.org/10.1002/JEZ.B.22520

Milani, L., Ghiselli, F., Passamonti, M., 2016. Mitochondrial selfish elements and the evolution of biological novelties. Curr Zool 62, 687–697. https://doi.org/10.1093/CZ/ZOW044

Milani, L., Ghiselli, F., Passamonti, M., 2012. Sex-Linked Mitochondrial Behavior During Early Embryo Development in *Ruditapes philippinarum* (Bivalvia Veneridae) a Species with the Doubly Uniparental Inheritance (DUI) of Mitochondria. J Exp Zool B Mol Dev Evol 318, 182–189. https://doi.org/10.1002/JEZ.B.22004

Milbury, C.A., Gaffney, P.M., 2005. Complete mitochondrial DNA sequence of the eastern oyster *Crassostrea virginica*. Marine Biotechnology 7, 697–712. https://doi.org/10.1007/S10126-005-0004-0/TABLES/6

Milbury, C.A., Lee, J.C., Cannone, J.J., Gaffney, P.M., Gutell, R.R., 2010. Fragmentation of the large subunit ribosomal RNA gene in oyster mitochondrial genomes. BMC Genomics 11, 1–17. https://doi.org/10.1186/1471-2164-11-485/FIGURES/7

Miller, J.R., Zhou, P., Mudge, J., Gurtowski, J., Lee, H., Ramaraj, T., Walenz, B.P., Liu, J., Stupar, R.M., Denny, R., Song, L., Singh, N., Maron, L.G., McCouch, S.R., McCombie, W.R., Schatz, M.C., Tiffin, P., Young, N.D., Silverstein, K.A.T., 2017. Hybrid assembly with long and short reads improves discovery of gene family expansions. BMC Genomics 18, 541. https://doi.org/10.1186/s12864-017-3927-8

Min, G.-S., Park, J.-K., 2009. Eurotatorian paraphyly: Revisiting phylogenetic relationships based on the complete mitochondrial genome sequence of *Rotaria rotatoria* (Bdelloidea: Rotifera: Syndermata). BMC Genomics 10, 533. https://doi.org/10.1186/1471-2164-10-533

Mita, S., Rizzuto, R., Moraes, C.T., Shanske, S., Arnaudo, E., Fabrizi, G.M., Koga, Y., Dimauro, S., Schon, E.A., 1990. Recombination via flanking direct repeats is a major cause of large-scale deletions of human mitochondrial DNA. Nucleic Acids Res 18, 561–567. https://doi.org/10.1093/NAR/18.3.561

Mitchell, A., Guerra, D., Stewart, D., Breton, S., 2016. In silico analyses of mitochondrial ORFans in freshwater mussels (Bivalvia: Unionoida) provide a framework for future studies of their origin and function. BMC Genomics 17, 1–22. https://doi.org/10.1186/S12864-016-2986-6/FIGURES/7

Mitra, R.D., Church, G.M., 1999. In situ localized amplification and contact replication of many individual DNA molecules. Nucleic Acids Res 27, e34–e39. https://doi.org/10.1093/NAR/27.24.E34

Mitra, R.D., Shendure, J., Olejnik, J., Krzymanska-Olejnik, E., Church, G.M., 2003. Fluorescent in situ sequencing on polymerase colonies. Anal Biochem 320, 55–65. https://doi.org/10.1016/S0003-2697(03)00291-4

Modesto, V., Ilarri, M., Souza, A.T., Lopes-Lima, M., Douda, K., Clavero, M., Sousa, R., 2018. Fish and mussels: Importance of fish for freshwater mussel conservation. Fish and Fisheries 19, 244–259. https://doi.org/10.1111/FAF.12252

Modica, M.V., Lombardo, F., Franchini, P., Oliverio, M., 2015. The venomous cocktail of the vampire snail *Colubraria reticulata* (Mollusca, Gastropoda). BMC Genomics 16, 441. https://doi.org/10.1186/s12864-015-1648-4

Moorkens, E., Cordeiro, J., Seddon, M., von Proschwitz, T., Woolnough, D., 2018. *Margaritifera margaritifera* (errata version published in 2018). The IUCN Red List of Threatened Species 2018 e.T12799A128686456. https://doi.org/http://dx.doi.org/10.2305/IUCN.UK.2017-3.RLTS.T12799A508865.en

Moorkens, E., Cordeiro, J., Seddon, M.B., von Proschwitz, T. Woolnough, D., 2017. *Margaritifera margaritifera* (Freshwater Pearl Mussel). The IUCN Red List of Threatened Species. URL https://www.iucnredlist.org/species/12799/128686456 (accessed 5.20.22).

Moorkens, E.A., 2018. Short-term breeding: releasing post-parasitic juvenile *Margaritifera* into ideal small-scale receptor sites: a new technique for the augmentation of declining populations. Hydrobiologia 810, 145–155. https://doi.org/10.1007/s10750-017-3138-y

Morin, P.A., Archer, F.I., Foote, A.D., Vilstrup, J., Allen, E.E., Wade, P., Durban, J., Parsons, K., Pitman, R., Li, L., Bouffard, P., Nielsen, S.C.A., Rasmussen, M., Willerslev, E., Gilbert, M.T.P., Harkins, T., 2010. Complete mitochondrial genome phylogeographic analysis of killer whales (*Orcinus orca*) indicates multiple species. Genome Res 20, 908–916. https://doi.org/10.1101/GR.102954.109

Mudge, J.M., Harrow, J., 2016. The state of play in higher eukaryote gene annotation. Nat Rev Genet. https://doi.org/10.1038/nrg.2016.119

Mun, S., Kim, Y.-J., Markkandan, K., Shin, W., Oh, S., Woo, J., Yoo, J., An, H., Han, K., 2017. The Whole-Genome and Transcriptome of the Manila Clam (*Ruditapes philippinarum*). Genome Biol Evol 9, 1487–1498. https://doi.org/10.1093/gbe/evx096

Murgarella, M., Puiu, D., Novoa, B., Figueras, A., Posada, D., Canchaya, C., 2016. Correction: A first insight into the genome of the filter-feeder mussel *Mytilus galloprovincialis*. PLoS One 11, 1–22. https://doi.org/10.1371/journal.pone.0160081

Myers, E.W., Sutton, G.G., Delcher, A.L., Dew, I.M., Fasulo, D.P., Flanigan, M.J., Kravitz, S.A., Mobarry, C.M., Reinert, K.H.J., Remington, K.A., Anson, E.L., Bolanos, R.A., Chou, H.-H., Jordan, C.M., Halpern, A.L., Lonardi, S., Beasley, E.M., Brandon, R.C., Chen, L., Dunn, P.J., Lai, Z., Liang, Y., Nusskern, D.R., Zhan, M., Zhang, Q., Zheng, X., Rubin, G.M., Adams, M.D., Venter, J.C., 2000. A Whole-Genome Assembly of *Drosophila*. Science (1979) 287, 2196–2204. https://doi.org/10.1126/science.287.5461.2196

Naimo, T.J., Damschen, E.D., Rada, R.G., Monroe, E.M., 1998. Nonlethal Evaluation of the Physiological Health of Unionid Mussels: Methods for Biopsy and Glycogen Analysis. J North Am Benthol Soc 17, 121–128. https://doi.org/10.2307/1468056

Nam, B.H., Kwak, W., Kim, Y.O., Kim, D.G., Kong, H.J., Kim, W.J., Kang, J.H., Park, J.Y., An, C.M., Moon, J.Y., Park, C.J., Yu, J.W., Yoon, J., Seo, M., Kim, K., Kim, D.K., Lee, S.B., Sung, S., Lee, C., Shin, Y., Jung, M., Kang, B.C., Shin, G.H., Ka, S., Caetano-Anolles, K., Cho, S., Kim, H., 2017. Genome sequence of pacific abalone (*Haliotis discus hannai*): the first draft genome in family Haliotidae. Gigascience 6, 1–8. https://doi.org/10.1093/gigascience/gix014

Neiman, M., Taylor, D.R., 2009. The causes of mutation accumulation in mitochondrial genomes. Proceedings of the Royal Society B: Biological Sciences 276, 1201–1209. https://doi.org/10.1098/RSPB.2008.1758

Nguyen, L.T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32, 268–274. https://doi.org/10.1093/molbev/msu300

Nguyen, T.T.T., Hayes, B.J., Ingram, B.A., 2014. Genetic parameters and response to selection in blue mussel (*Mytilus galloprovincialis*) using a SNP-based pedigree. Aquaculture 420–421, 295–301. https://doi.org/10.1016/J.AQUACULTURE.2013.11.021

Nicholls, D.G., Ferguson, S.J., 2002. Bioenergetics 3. Gulf Professional Publishing.

Nolan, J.R., Bergthorsson, U., Adema, C.M., 2014. *Physella acuta*: atypical mitochondrial gene order among panpulmonates (Gastropoda). Journal of Molluscan Studies 80, 388–399. https://doi.org/10.1093/MOLLUS/EYU025

Noll, D., Leon, F., Brandt, D., Pistorius, P., le Bohec, C., Bonadonna, F., Trathan, P.N., Barbosa, A., Rey, A.R., Dantas, G.P.M., Bowie, R.C.K., Poulin, E., Vianna, J.A., 2022. Positive selection over the mitochondrial genome and its role in the diversification of gentoo penguins in response to adaptation in isolation. Scientific Reports 2022 12:1 12, 1–13. https://doi.org/10.1038/s41598-022-07562-0

Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, A. v., Mikheenko, A., Vollger, M.R., Altemose, N., Uralsky, L., Gershman, A., Aganezov, S., Hoyt, S.J., Diekhans, M., Logsdon, G.A., Alonge, M., Antonarakis, S.E., Borchers, M., Bouffard, G.G., Brooks, S.Y., Caldas, G. v., Chen, N.-C., Cheng, H., Chin, C.-S., Chow, W., Lima, L.G. de, Dishuck, P.C., Durbin, R., Dvorkina, T., Fiddes, I.T., Formenti, G., Fulton, R.S., Fungtammasan, A., Garrison, E., Grady, P.G.S., Graves-Lindsay, T.A., Hall, I.M., Hansen, N.F., Hartley, G.A., Haukness, M., Howe, K., Hunkapiller, M.W., Jain, C., Jain, M., Jarvis, E.D., Kerpedjiev, P., Kirsche, M., Kolmogorov, M., Korlach, J., Kremitzki, M., Li, H., Maduro, V. v., Marschall, T., McCartney, A.M., McDaniel, J., Miller, D.E., Mullikin, J.C., Myers, E.W., Olson, N.D., Paten, B., Peluso, P., Pevzner, P.A., Porubsky, D., Potapova, T., Rogaev, E.I., Rosenfeld, J.A., Salzberg, S.L., Schneider, V.A., Sedlazeck, F.J., Shafin, K., Shew, C.J., Shumate, A., Sims, Y., Smit, A.F.A., Soto, D.C., Sović, I., Storer, J.M., Streets, A., Sullivan, B.A., Thibaud-Nissen, F., Torrance, J., Wagner, J., Walenz, B.P., Wenger, A., Wood, J.M.D., Xiao, C., Yan, S.M., Young, A.C., Zarate, S., Surti, U., McCoy, R.C., Dennis, M.Y., Alexandrov, I.A., Gerton, J.L., O'Neill, R.J., Timp, W., Zook, J.M., Schatz, M.C., Eichler, E.E., Miga, K.H., Phillippy, A.M., 2022. The complete sequence of a human genome. Science (1979) 376, 44–53. https://doi.org/10.1126/SCIENCE.ABJ6987

Nyrén, P., 1987. Enzymatic method for continuous monitoring of DNA polymerase activity. Anal Biochem 167, 235–238. https://doi.org/10.1016/0003-2697(87)90158-8

Nyrén, P., Lundin, A., 1985. Enzymatic method for continuous monitoring of inorganic pyrophosphate synthesis. Anal Biochem 151, 504–509. https://doi.org/10.1016/0003-2697(85)90211-8

Obata, M., Kamiya, C., Kawamura, K., Komaru, A., 2006. Sperm mitochondrial DNA transmission to both male and female offspring in the blue mussel *Mytilus galloprovincialis*. Dev Growth Differ 48, 253–261. https://doi.org/10.1111/J.1440-169X.2006.00863.X

Ojala, D., Montoya, J., Attardi, G., 1981. tRNA punctuation model of RNA processing in human mitochondria. Nature 1981 290:5806 290, 470–474. https://doi.org/10.1038/290470a0

Oliver, K.R., Greene, W.K., 2009. Transposable elements: powerful facilitators of evolution. BioEssays 31, 703–714. https://doi.org/10.1002/BIES.200800219

Oppen, M.J.H. van, Coleman, M.A., 2022. Advancing the protection of marine life through genomics. PLoS Biol 20, e3001801. https://doi.org/10.1371/JOURNAL.PBIO.3001801

Osca, D., Irisarri, I., Todt, C., Grande, C., Zardoya, R., 2014. The complete mitochondrial genome of *Scutopus ventrolineatus* (Mollusca: Chaetodermomorpha) supports the Aculifera hypothesis. BMC Evol Biol 14, 197. https://doi.org/10.1186/s12862-014-0197-9

Padmanabhan, R., Padmanabhan, Raji, Wu, R., 1972. Nucleotide sequence analysis of DNA: IX. Use of oligonucleotides of defined sequence as primers in DNA sequence analysis. Biochem Biophys Res Commun 48, 1295–1302. https://doi.org/10.1016/0006-291X(72)90852-2

Paez, S., Kraus, R.H.S., Shapiro, B., Gilbert, M.T.P., Jarvis, E.D., Group, V.G.P.C., Al-Ajli, F.O., Ceballos, G., Crawford, A.J., Fedrigo, O., Johnson, R.N., Johnson, W.E., Marques-Bonet, T., Morin, P.A., Mueller, R.C., Ryder, O.A., Teeling, E.C., Venkatesh, B., 2022. Reference genomes for conservation. Science (1979) 377, 364–366. https://doi.org/10.1126/SCIENCE.ABM8127

Paps, J., 2018. What Makes an Animal? The Molecular Quest for the Origin of the Animal Kingdom. Integr Comp Biol 58, 654–665. https://doi.org/10.1093/ICB/ICY036

Paps, J., Xu, F., Zhang, G., Holland, P.W.H., 2015. Reinforcing the egg-timer: Recruitment of novel Lophotrochozoa homeobox genes to early and late development in the Pacific oyster. Genome Biol Evol 7, 677–688. https://doi.org/10.1093/gbe/evv018

Park, J.-K., Kim, K.-H., Kang, S., Kim, W., Eom, K.S., Littlewood, D., 2007. A common origin of complex life cycles in parasitic flatworms: evidence from the complete mitochondrial genome of *Microcotyle sebastis* (Monogenea: Platyhelminthes). BMC Evol Biol 7, 11. https://doi.org/10.1186/1471-2148-7-11

Parmalee, P.W., Bogan, A.E., 1998. Freshwater mussels of Tennessee. University of Tennessee Press.

Passamonti, M., Calderone, M., Delpero, M., Plazzi, F., 2020. Clues of in vivo nuclear gene regulation by mitochondrial short non-coding RNAs. Scientific Reports 2020 10:1 10, 1–12. https://doi.org/10.1038/s41598-020-65084-z

Passamonti, M., Ghiselli, F., 2009. Doubly Uniparental Inheritance: Two Mitochondrial Genomes, One Precious Model for Organelle DNA Inheritance and Evolution. https://home.liebertpub.com/dna 28, 79–89. https://doi.org/10.1089/DNA.2008.0807

Passamonti, M., Ricci, A., Milani, L., Ghiselli, F., 2011. Mitochondrial genomes and Doubly Uniparental Inheritance: New insights from *Musculista senhousia* sex-linked mitochondrial DNAs (Bivalvia Mytilidae). BMC Genomics 12, 1–19. https://doi.org/10.1186/1471-2164-12-442/TABLES/8

Patnaik, B.B., Wang, T.H., Kang, S.W., Hwang, H.J., Park, S.Y., Park, E.B., Chung, J.M., Song, D.K., Kim, C., Kim, S., Lee, J.S., Han, Y.S., Park, H.S., Lee, Y.S., 2016. Sequencing, De Novo Assembly, and Annotation of the Transcriptome of the Endangered Freshwater Pearl Bivalve, *Cristaria plicata*, Provides Novel Insights into Functional Genes and Marker Discovery. PLoS One 11, e0148622. https://doi.org/10.1371/JOURNAL.PONE.0148622

Pauls, S.U., Alp, M., Bálint, M., Bernabò, P., Čiampor, F., Čiamporová-Zaťovičová, Z., Finn, D.S., Kohout, J., Leese, F., Lencioni, V., Paz-Vinas, I., Monaghan, M.T., 2014. Integrating molecular tools into freshwater ecology: developments and opportunities. Freshw Biol 59, 1559–1576. https://doi.org/10.1111/fwb.12381

Peñarrubia, L., Araguas, R.-M., Pla, C., Sanz, N., Viñas, J., Vidal, O., 2015a. Identification of 246 microsatellites in the Asiatic clam (*Corbicula fluminea*). Conserv Genet Resour 7, 393–395. https://doi.org/10.1007/s12686-014-0378-2

Peñarrubia, L., Sanz, N., Pla, C., Vidal, O., Viñas, J., 2015b. Using massive parallel sequencing for the development, validation, and application of population genetics markers in the invasive bivalve zebra mussel (*Dreissena polymorpha*). PLoS One 10, e0120732. https://doi.org/10.1371/journal.pone.0120732

Penn, O., Privman, E., Ashkenazy, H., Landan, G., Graur, D., Pupko, T., 2010. GUIDANCE: A web server for assessing alignment confidence scores. Nucleic Acids Res 38, W23–W28. https://doi.org/10.1093/nar/gkq443

Pérez-Parallé, M.L., Pazos, A.J., Mesías-Gansbiller, C., Sánchez, J.L., 2016. Hox, Parahox, Ehgbox, and NK Genes in Bivalve Molluscs: Evolutionary Implications. J Shellfish Res 35, 179–190. https://doi.org/10.2983/035.035.0119

Pett, W., Ryan, J.F., Pang, K., Mullikin, J.C., Martindale, M.Q., Baxevanis, A.D., Lavrov, D. v., 2011. Extreme mitochondrial evolution in the ctenophore *Mnemiopsis leidyi*: Insight from mtDNA and the nuclear genome. Mitochondrial DNA 22, 130–142. https://doi.org/10.3109/19401736.2011.624611/SUPPL_FILE/IMDN_A_624611_SM0001.PDF

Pfeiffer, J.M., Breinholt, J.W., Page, L.M., 2019. Unioverse: A phylogenomic resource for reconstructing the evolution of freshwater mussels (Bivalvia, Unionoida). Mol Phylogenet Evol 137, 114–126. https://doi.org/10.1016/J.YMPEV.2019.02.016

Pfeiffer, J.M., Graf, D.L., Cummings, K.S., Page, L.M., 2021. Taxonomic revision of a radiation of South-east Asian freshwater mussels (Unionidae : Gonideinae : Contradentini+Rectidentini). https://doi.org/10.1071/IS20044 35, 394–470. https://doi.org/10.1071/IS20044

Pfenninger, M., Lerp, H., Tobler, M., Passow, C., Kelley, J.L., Funke, E., Greshake, B., Erkoc, U.K., Berberich, T., Plath, M., 2014. Parallel evolution of cox genes in H2S-tolerant fish as key adaptation to a toxic environment. Nature Communications 2014 5:1 5, 1–7. https://doi.org/10.1038/ncomms4873

Pollard, S.L., Holland, P.W.H., 2000. Evidence for 14 homeobox gene clusters in human genome ancestry. Current Biology 10, 1059–1062. https://doi.org/10.1016/S0960-9822(00)00676-X

Ponder, W.F., Lindberg, D.R., Ponder, J.M., 2019. Biology and Evolution of the Mollusca, Volume 1,2. CRC Press.

Powell, D., Subramanian, S., Suwansa-ard, S., Zhao, M., O’Connor, W., Raftos, D., Elizur, A., 2018. The genome of the oyster *Saccostrea* offers insight into the environmental resilience of bivalves. DNA Research 25, 655–665. https://doi.org/10.1093/dnares/dsy032

Pozzi, A., Dowling, D.K., Sloan, D., 2019. The Genomic Origins of Small Mitochondrial RNAs: Are They Transcribed by the Mitochondrial DNA or by Mitochondrial Pseudogenes within the Nucleus (NUMTs)? Genome Biol Evol 11, 1883–1896. https://doi.org/10.1093/GBE/EVZ132

Pozzi, A., Plazzi, F., Milani, L., Ghiselli, F., Passamonti, M., 2017. SmithRNAs: Could Mitochondria “Bend” Nuclear Regulation? Mol Biol Evol 34, 1960–1973. https://doi.org/10.1093/molbev/msx140

Prada, C.F., Boore, J.L., 2019. Gene annotation errors are common in the mammalian mitochondrial genomes database. BMC Genomics 20, 1–8. https://doi.org/10.1186/S12864-019-5447-1/FIGURES/4

Prober, J.M., Trainor, G.L., Dam, R.J., Hobbs, F.W., Robertson, C.W., Zagursky, R.J., Cocuzza, A.J., Jensen, M.A., Baumeister, K., 1987. A System for Rapid DNA Sequencing with Fluorescent Chain-Terminating Dideoxynucleotides. Science (1979) 238, 336–341. https://doi.org/10.1126/SCIENCE.2443975

Pruitt, K.D., Tatusova, T., Maglott, D.R., 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res 35, D61–D65. https://doi.org/10.1093/nar/gkl842

Punta, M., Coggill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., Heger, A., Holm, L., Sonnhammer, E.L.L., Eddy, S.R., Bateman, A., Finn, R.D., 2012. The Pfam protein families database. Nucleic Acids Res 40, D290–D301. https://doi.org/10.1093/NAR/GKR1065

Punzi, E., Milani, L., Ghiselli, F., Passamonti, M., 2018. Lose it or keep it: (how bivalves can provide) insights into mitochondrial inheritance mechanisms. J Exp Zool B Mol Dev Evol 330, 41–51. https://doi.org/10.1002/JEZ.B.22788

Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., Lopez, R., 2005. InterProScan: Protein domains identifier. Nucleic Acids Res 33, W116–W120. https://doi.org/10.1093/nar/gki442

Quince, C., Walker, A.W., Simpson, J.T., Loman, N.J., Segata, N., 2017. Shotgun metagenomics, from sampling to analysis. Nature Biotechnology 2017 35:9 35, 833–844. https://doi.org/10.1038/nbt.3935

Raghavan, N., Knight, M., 2006. The snail (Biomphalaria glabrata) genome project. Trends Parasitol 22, 148–151. https://doi.org/10.1016/J.PT.2006.02.008

Raimond, R., Marcadé, I., Bouchon, D., Rigaud, T., Bossy, J.P., Souty-Grosset, C., 1999. Organization of the Large Mitochondrial Genome in the Isopod *Armadillidium vulgare*. Genetics 151, 203–210. https://doi.org/10.1093/GENETICS/151.1.203

Rak, M., Tzagoloff, A., 2009. F1-dependent translation of mitochondrially encoded Atp6p and Atp8p subunits of yeast ATP synthase. Proc Natl Acad Sci U S A 106, 18509–18514.

https://doi.org/10.1073/PNAS.0910351106/ASSET/3C152803-7151-40D0-97CE-5167555D8FCE/ASSETS/GRAPHIC/ZPQ9990900520007.JPEG

Ranallo-Benavidez, T.R., Jaron, K.S., Schatz, M.C., 2020. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nat Commun 11, 1–10. https://doi.org/10.1038/s41467-020-14998-3

Rawlings, T.A., Collins, T.M., Bieler, R., 2001. A Major Mitochondrial Gene Rearrangement Among Closely Related Species. Mol Biol Evol 18, 1604–1609. https://doi.org/10.1093/OXFORDJOURNALS.MOLBEV.A003949

Rawlings, T.A., Collinst, T.M., Bieler, R., 2003. Changing identities: tRNA duplication and remolding within animal mitochondrial genomes. Proc Natl Acad Sci U S A 100, 15700–15705. https://doi.org/10.1073/PNAS.2535036100/SUPPL_FILE/5036SUPPTABLE1.XLS

Rawlings, T.A., MacInnis, M.J., Bieler, R., Boore, J.L., Collins, T.M., 2010. Sessile snails, dynamic genomes: Gene rearrangements within the mitochondrial genome of a family of caenogastropod molluscs. BMC Genomics 11, 1–24. https://doi.org/10.1186/1471-2164-11-440/FIGURES/6

Ray, W., Tu, C. pei D., Padmanabhan, R., 1973. Nucleotide sequence analysis of DNA XII. The chemical synthesis and sequence analysis of a dodecadeoxynucleotide which binds to the endolysin gene of bacteriophage lambda. Biochem Biophys Res Commun 55, 1092–1099. https://doi.org/10.1016/S0006-291X(73)80007-5

Regan, T., Stevens, L., Peñaloza, C., Houston, R.D., Robledo, D., Bean, T.P., 2021. Ancestral Physical Stress and Later Immune Gene Family Expansions Shaped Bivalve Mollusc Evolution. Genome Biol Evol 13. https://doi.org/10.1093/gbe/evab177

Renaut, S., Guerra, D., Hoeh, W.R., Stewart, D.T., Bogan, A.E., Ghiselli, F., Milani, L., Passamonti, M., Breton, S., 2018. Genome Survey of the Freshwater Mussel *Venustaconcha ellipsiformis* (Bivalvia: Unionida) Using a Hybrid De Novo Assembly Approach. Genome Biol Evol 10, 1637–1646. https://doi.org/10.1093/gbe/evy117

Rhie, A., McCarthy, S.A., Fedrigo, O., Damas, J., Formenti, G., Koren, S., Uliano-Silva, M., Chow, W., Fungtammasan, A., Kim, J., Lee, C., Ko, B.J., Chaisson, M., Gedman, G.L., Cantin, L.J., Thibaud-Nissen, F., Haggerty, L., Bista, I., Smith, M., Haase, B., Mountcastle, J., Winkler, S., Paez, S., Howard, J., Vernes, S.C., Lama, T.M., Grutzner, F., Warren, W.C., Balakrishnan, C.N., Burt, D., George, J.M., Biegler, M.T., Iorns, D., Digby, A., Eason, D., Robertson, B., Edwards, T., Wilkinson, M., Turner, G., Meyer, A., Kautt, A.F., Franchini, P., Detrich, H.W., Svardal, H., Wagner, M., Naylor, G.J.P., Pippel, M., Malinsky, M., Mooney, M., Simbirsky, M., Hannigan, B.T., Pesout, T., Houck, M., Misuraca, A., Kingan, S.B., Hall, R., Kronenberg, Z., Sović, I., Dunn, C., Ning, Z., Hastie, A., Lee, J., Selvaraj, S., Green, R.E., Putnam, N.H., Gut, I., Ghurye, J., Garrison, E., Sims, Y., Collins, J., Pelan, S., Torrance, J., Tracey, A., Wood, J., Dagnew, R.E., Guan, D., London, S.E., Clayton, D.F., Mello, C. v., Friedrich, S.R., Lovell, P. v., Osipova, E., Al-Ajli, F.O., Secomandi, S., Kim, H., Theofanopoulou, C., Hiller, M., Zhou, Y., Harris, R.S., Makova, K.D., Medvedev, P., Hoffman, J., Masterson, P., Clark, K., Martin, F., Howe, Kevin, Flicek, P., Walenz, B.P., Kwak, W., Clawson, H., Diekhans, M., Nassar, L., Paten, B., Kraus, R.H.S., Crawford, A.J., Gilbert, M.T.P., Zhang, G., Venkatesh, B., Murphy, R.W., Koepfli, K.-P., Shapiro, B., Johnson, W.E., di Palma, F., Marques-Bonet, T., Teeling, E.C., Warnow, T., Graves, J.M., Ryder, O.A., Haussler, D., O’Brien, S.J., Korlach, J., Lewin, H.A., Howe, Kerstin, Myers, E.W., Durbin, R., Phillippy, A.M., Jarvis, E.D., 2021. Towards complete and error-free genome assemblies of all vertebrate species. Nature 592, 737–746. https://doi.org/10.1038/s41586-021-03451-0

Richards, P.M., Liu, M.M., Lowe, N., Davey, J.W., Blaxter, M.L., Davison, A., 2013. RAD-Seq derived markers flank the shell colour and banding loci of the *Cepaea nemoralis* supergene. Mol Ecol 22, 3077–3089. https://doi.org/10.1111/mec.12262

Richards, S., 2015. It’s more than stamp collecting: how genome sequencing can unify biological research. Trends in Genetics 31, 411–421. https://doi.org/https://doi.org/10.1016/j.tig.2015.04.007

Rigaa, A., Monnerot, M., Sellos, D., 1995. Molecular cloning and complete nucleotide sequence of the repeated unit and flanking gene of the scallop *Pecten maximus* mitochondrial DNA: Putative replication origin features. Journal of Molecular Evolution 1995 41:2 41, 189–195. https://doi.org/10.1007/BF00170672

Roberts, N.G., Kocot, K.M., 2021. Developing a contaminant-aware pipeline to resolve lophotrochozoan relationships in the genomics age. INTEGRATIVE AND COMPARATIVE BIOLOGY. pp. E1237–E1238.

Robertson, L.S., Galbraith, H.S., Iwanowicz, D., Blakeslee, C.J., Cornman, R.S., 2017. RNA sequencing analysis of transcriptional change in the freshwater mussel *Elliptio complanata* after environmentally relevant sodium chloride exposure. Environ Toxicol Chem 36, 2352–2366. https://doi.org/10.1002/ETC.3774

Robicheau, B.M., Breton, S., Stewart, D.T., 2017. Sequence motifs associated with paternal transmission of mitochondrial DNA in the horse mussel, *Modiolus modiolus* (Bivalvia: Mytilidae). Gene 605, 32–42. https://doi.org/10.1016/J.GENE.2016.12.025

Rogers, R.L., Grizzard, S.L., Titus-McQuillan, J.E., Bockrath, K., Patel, S., Wares, J.P., Garner, J.T., Moore, C.C., 2021. Gene family amplification facilitates adaptation in freshwater unionid bivalve *Megalonaias nervosa*. Mol Ecol 30, 1155–1173. https://doi.org/10.1111/mec.15786

Romero, P.E., Weigand, A.M., Pfenninger, M., 2016. Positive selection on panpulmonate mitogenomes provide new clues on adaptations to terrestrial life. BMC Evol Biol 16, 1–13. https://doi.org/10.1186/S12862-016-0735-8/TABLES/5

Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlén, M., Nyrén, P., 1996. Real-Time DNA Sequencing Using Detection of Pyrophosphate Release. Anal Biochem 242, 84–89. https://doi.org/10.1006/abio.1996.0432

Ronaghi, M., Uhlén, M., Nyrén, P., 1998. A Sequencing Method Based on Real-Time Pyrophosphate. Science (1979) 281, 363–365. https://doi.org/10.1126/science.281.5375.363

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. Syst Biol 61, 539–542. https://doi.org/10.1093/sysbio/sys029

Rosenfeld, J.A., Foox, J., DeSalle, R., 2016. Insect genome content phylogeny and functional annotation of core insect genomes. Mol Phylogenet Evol 97, 224–232. https://doi.org/10.1016/J.YMPEV.2015.10.014

Roznere, I., Sinn, B.T., Watters, G.T., 2018. The *Amblema plicata* Transcriptome as a Resource to Assess Environmental Impacts on Freshwater Mussels. Freshwater Mollusk Biology and Conservation 21, 57–64. https://doi.org/10.31931/fmbc.v21i2.2018.57–64

Ryan, J.F., 2013. Baa.pl: A tool to evaluate de novo genome assemblies with RNA transcripts.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., Erlich, H.A., 1988. Primer-Directed Enzymatic Amplification of DNA with a Thermostable DNA Polymerase. Science (1979) 239, 487–491. https://doi.org/10.1126/science.2448875

Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A., Arnheim, N., 1985. Enzymatic Amplification of β-Globin Genomic Sequences and Restriction Site Analysis for Diagnosis of Sickle Cell Anemia. Science (1979) 230, 1350–1354. https://doi.org/10.1126/science.2999980

Sambrook, Joseph., Russell, D.W., Maniatis, T., 1989. Molecular cloning : a laboratory manual. Cold Spring Harbor Laboratory Press, New York, NY.

Sanderson, M.J., 2008. Phylogenetic signal in the eukaryotic tree of life. Science (1979) 321, 121–123. https://doi.org/10.1126/SCIENCE.1154449/SUPPL_FILE/SANDERSON.SOM.PDF

Sanger, F., 1988. SEQUENCES, SEQUENCES, AND SEQUENCES. Annu Rev Biochem 57, 1–29. https://doi.org/10.1146/annurev.bi.57.070188.000245

Sanger, F., Air, G.M., Barrell, B.G., Brown, N.L., Coulson, A.R., Fiddes, J.C., Hutchison, C.A., Slocombe, P.M., Smith, M., 1977a. Nucleotide sequence of bacteriophage φX174 DNA. Nature 1977 265:5596 265, 687–695. https://doi.org/10.1038/265687a0

Sanger, F., Brownlee, G.G., Barrell, B.G., 1965. A two-dimensional fractionation procedure for radioactive nucleotides. J Mol Biol 13, 373-IN4. https://doi.org/10.1016/S0022-2836(65)80104-8

Sanger, F., Coulson, A.R., 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J Mol Biol 94, 441–448. https://doi.org/10.1016/0022-2836(75)90213-2

Sanger, F., Coulson, A.R., Hong, G.F., Hill, D.F., Petersen, G.B., 1982. Nucleotide sequence of bacteriophage λ DNA. J Mol Biol 162, 729–773. https://doi.org/10.1016/0022-2836(82)90546-0

Sanger, F., Nicklen, S., Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences 74, 5463–5467. https://doi.org/10.1073/PNAS.74.12.5463

Sano, I., Saito, T., Ito, S., Ye, B., Uechi, T., Seo, T., Do, V.T., Kimura, K., Hirano, T., Yamazaki, D., Shirai, A., Kondo, T., Miura, O., Miyazaki, J.I., Chiba, S., 2022. Resolving species-level diversity of *Beringiana* and *Sinanodonta* mussels (Bivalvia: Unionidae) in the Japanese archipelago using genome-wide data. Mol Phylogenet Evol 175, 107563. https://doi.org/10.1016/J.YMPEV.2022.107563

Sasuga, J., Yokobori, S.I., Kaifu, M., Ueda, T., Nishikawa, K., Watanabe, K., 1999. Gene Contents and Organization of a Mitochondrial DNA Segment of the Squid *Loligo bleekeri*. Journal of Molecular Evolution 1999 48:6 48, 692–702. https://doi.org/10.1007/PL00006513

Sato, K., Sato, M., 2017. Multiple ways to prevent transmission of paternal mitochondrial DNA for maternal inheritance in animals. The Journal of Biochemistry 162, 247–253. https://doi.org/10.1093/JB/MVX052

Savolainen, O., Lascoux, M., Merilä, J., 2013. Ecological genomics of local adaptation. Nat Rev Genet 14, 807–820. https://doi.org/10.1038/nrg3522

Schell, T., Feldmeyer, B., Schmidt, H., Greshake, B., Tills, O., Truebano, M., Rundle, S.D., Paule, J., Ebersberger, I., Pfenninger, M., 2017. An Annotated Draft Genome for *Radix auricularia* (Gastropoda, Mollusca). Genome Biol Evol 9, 585–592. https://doi.org/10.1093/gbe/evx032

Schlüter, J., Rätsch, C., 1999. Perlen und Perlmutt. Ellert und Richter, Hamburg.

Schrödl, M., Stöger, I., 2014. A review on deep molluscan phylogeny: old markers, integrative approaches, persistent problems. J Nat Hist 48, 2773–2804. https://doi.org/10.1080/00222933.2014.963184

Schultzhaus, J.N., Taitt, C.R., Orihuela, B., Smerchansky, M., Schultzhaus, Z.S., Rittschof, D., Wahl, K.J., Spillmann, C.M., 2019. Comparison of seven methods for DNA extraction from prosomata of the acorn barnacle, *Amphibalanus amphitrite*. Anal Biochem 586, 113441. https://doi.org/10.1016/J.AB.2019.113441

Schwentner, M., Combosch, D.J., Pakes Nelson, J., Giribet, G., 2017. A Phylogenomic Solution to the Origin of Insects by Resolving Crustacean-Hexapod Relationships. Current Biology 27, 1818-1824.e5. https://doi.org/10.1016/J.CUB.2017.05.040

Sebastian, W., Sukumaran, S., Zacharia, P.U., Muraleedharan, K.R., Dinesh Kumar, P.K., Gopalakrishnan, A., 2020. Signals of selection in the mitogenome provide insights into adaptation mechanisms in heterogeneous habitats in a widely distributed pelagic fish. Scientific Reports 2020 10:1 10, 1–14. https://doi.org/10.1038/s41598-020-65905-1

Sedlazeck, F.J., Lee, H., Darby, C.A., Schatz, M.C., 2018. Piercing the dark matter: Bioinformatics of long-range sequencing and mapping. Nat Rev Genet. https://doi.org/10.1038/s41576-018-0003-4

Sevigny, J.L., Kirouac, L.E., Thomas, W.K., Ramsdell, J.S., Lawlor, K.E., Sharifi, O., Grewal, S., Baysdorfer, C., Curr, K., Naimie, A.A., Okamoto, K., Murray, J.A., Newcomb, J.M., 2015. The Mitochondrial Genomes of the Nudibranch Mollusks, *Melibe leonina* and *Tritonia diomedea*, and Their Impact on Gastropod Phylogeny. PLoS One 10, e0127519. https://doi.org/10.1371/JOURNAL.PONE.0127519

Shadel, G.S., Clayton, D.A., 1997. MITOCHONDRIAL DNA MAINTENANCE IN VERTEBRATES. Annu Rev Biochem 66, 409–435. https://doi.org/10.1146/annurev.biochem.66.1.409

Sharbrough, J., Cruise, J.L., Beetch, M., Enright, N.M., Neiman, M., Pecon-Slattery, J., 2017. Genetic Variation for Mitochondrial Function in the New Zealand Freshwater Snail Potamopyrgus antipodarum. Journal of Heredity 108, 759–768. https://doi.org/10.1093/JHERED/ESX041

Shen, X., Meng, X.P., Chu, K.H., Zhao, N.N., Tian, M., Liang, M., Hao, J., 2014. Comparative mitogenomic analysis reveals cryptic species: A case study in Mactridae (Mollusca: Bivalvia). Comp Biochem Physiol Part D Genomics Proteomics 12, 1–9. https://doi.org/10.1016/J.CBD.2014.08.002

Shen, X.X., Zhou, X., Kominek, J., Kurtzman, C.P., Hittinger, C.T., Rokas, A., 2016. Reconstructing the backbone of the Saccharomycotina yeast phylogeny using genome-scale data. G3: Genes, Genomes, Genetics 6, 3927–3939. https://doi.org/10.1534/G3.116.034744/-/DC1

Shendure, J., Balasubramanian, S., Church, G.M., Gilbert, W., Rogers, J., Schloss, J.A., Waterston, R.H., 2017. DNA sequencing at 40: past, present and future. Nature 550, 345–353. https://doi.org/10.1038/nature24286

Shendure, J., Porreca, G.J., Reppas, N.B., Lin, X., McCutcheon, J.P., Rosenbaum, A.M., Wang, M.D., Zhang, K., Mitra, R.D., Church, G.M., 2005. Molecular biology: Accurate multiplex polony sequencing of an evolved bacterial genome. Science (1979) 309, 1728–1732. https://doi.org/10.1126/SCIENCE.1117389/SUPPL_FILE/SHENDURE.SOM.PDF

Shi, J., Hong, Y., Sheng, J., Peng, K., Wang, J., 2015. De novo transcriptome sequencing to identify the sex-determination genes in *Hyriopsis schlegelii*. Biosci Biotechnol Biochem 79, 1257–1265. https://doi.org/10.1080/09168451.2015.1025690

Sigwart, J.D., Sumner-Rooney, L.H., 2015. Mollusca: Caudofoveata, Monoplacophora, Polyplacophora, Scaphopoda, And Solenogastres, in: Structure and Evolution of

Invertebrate Nervous Systems. Oxford University Press, pp. 172–189. https://doi.org/10.1093/acprof:oso/9780199682201.003.0018

Simakov, O., Marletaz, F., Cho, S.-J., Edsinger-Gonzales, E., Havlak, P., Hellsten, U., Kuo, D.-H., Larsson, T., Lv, J., Arendt, D., Savage, R., Osoegawa, K., de Jong, P., Grimwood, J., Chapman, J.A., Shapiro, H., Aerts, A., Otillar, R.P., Terry, A.Y., Boore, J.L., Grigoriev, I. v., Lindberg, D.R., Seaver, E.C., Weisblat, D.A., Putnam, N.H., Rokhsar, D.S., 2012. Insights into bilaterian evolution from three spiralian genomes. Nature 493, 526–531. https://doi.org/10.1038/nature11696

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E. v., Zdobnov, E.M., 2015. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Simison, W.B., Lindberg, D.R., Boore, J.L., 2006. Rolling circle amplification of metazoan mitochondrial genomes. Mol Phylogenet Evol 39, 562–567. https://doi.org/10.1016/J.YMPEV.2005.11.006

Simon, M., Faye, G., 1984. Organization and processing of the mitochondrial oxi3/oli2 multigenic transcript in yeast. Molecular and General Genetics MGG 1984 196:2 196, 266–274. https://doi.org/10.1007/BF00328059

Skibinski, D.O.F., Gallagher, C., Beynon, C.M., 1994. Mitochondrial DNA inheritance. Nature 368, 817–818. https://doi.org/10.1038/368817b0

Smit, A., Hubley, R., 2015a. RepeatModeler.

Smit, A., Hubley, R., 2015b. RepeatMasker.

Smith, C.H., 2021. A High-Quality Reference Genome for a Parasitic Bivalve with Doubly Uniparental Inheritance (Bivalvia: Unionida). Genome Biol Evol 13. https://doi.org/10.1093/gbe/evab029

Smith, C.H., Pfeiffer, J.M., Johnson, N.A., 2020. Comparative phylogenomics reveal complex evolution of life history strategies in a clade of bivalves with parasitic larvae (Bivalvia: Unionoida: Ambleminae). Cladistics 36, 505–520. https://doi.org/10.1111/cla.12423

Smith, C.H., Pinto, B.J., Kirkpatrick, M., Hillis, D.M., Pfeiffer, J.M., Havird, J.C., 2022. A tale of two paths: The evolution of mitochondrial recombination in bivalves with doubly uniparental inheritance. bioRxiv 2022.10.22.513339. https://doi.org/10.1101/2022.10.22.513339

Smith, D.R., Snyder, M., 2007. Complete mitochondrial DNA sequence of the scallop *Placopecten magellanicus*: Evidence of transposition leading to an uncharacteristically large mitochondrial genome. J Mol Evol 65, 380–391. https://doi.org/10.1007/S00239-007-9016-X/FIGURES/7

Smith, L.M., Fung, S., Hunkapiller, M.W., Hunkapiller, T.J., Hood, L.E., 1985. The synthesis of oligonucleotides containing an aliphatic amino group at the 5′ terminus: synthesis of fluorescent DNA primers for use in DNA sequence analysis. Nucleic Acids Res 13, 2399–2412. https://doi.org/10.1093/NAR/13.7.2399

Smith, M.L., Hahn, M.W., 2021. New Approaches for Inferring Phylogenies in the Presence of Paralogs. Trends in Genetics 37, 174–187. https://doi.org/10.1016/J.TIG.2020.08.012

Smith, S.A., Wilson, N.G., Goetz, F.E., Feehery, C., Andrade, S.C.S., Rouse, G.W., Giribet, G., Dunn, C.W., 2011. Resolving the evolutionary relationships of molluscs with phylogenomic tools. Nature 480, 364–367. https://doi.org/10.1038/nature10526

Smith-Unna, R., Boursnell, C., Patro, R., Hibberd, J.M., Kelly, S., 2016. TransRate: reference free quality assessment of de novo transcriptome assemblies. Genome Res 26, gr.196469.115. https://doi.org/10.1101/GR.196469.115

Snyder, M., Fraser, A.R., Laroche, J., Gartner-Kepkay, K.E., Zouros, E., Kafatos, F.C., 1987. Atypical mitochondrial DNA from the deep-sea scallop *Placopecten magellanicus*. Proceedings of the National Academy of Sciences 84, 7595–7599. https://doi.org/10.1073/PNAS.84.21.7595

Sokolov, E.P., 2000. An improved method for DNA isolation from mucopolysaccharide-rich molluscan tissues. Journal of Molluscan Studies 66, 573–575. https://doi.org/10.1093/mollus/66.4.573

Song, L., Florea, L., 2015. Rcorrector: Efficient and accurate error correction for Illumina RNA-seq reads. Gigascience 4, 48. https://doi.org/10.1186/s13742-015-0089-y

Sousa, R., Antunes, C., Guilhermino, L., 2008. Ecology of the invasive Asian clam *Corbicula fluminea* (Müller, 1774) in aquatic ecosystems: an overview. Annales de Limnologie - International Journal of Limnology 44, 85–94. https://doi.org/10.1051/limn:2008017

Spooner, D.E., Frost, P.C., Hillebrand, H., Arts, M.T., Puckrin, O., Xenopoulos, M.A., 2013. Nutrient loading associated with agriculture land use dampens the importance of consumer-mediated niche construction. Ecol Lett 16, 1115–1125. https://doi.org/10.1111/ele.12146

Staden, R., 1979. A strategy of DNA sequencing employing computer programs. Nucleic Acids Res 6, 2601–2610. https://doi.org/10.1093/NAR/6.7.2601

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Stanton, D.J., Daehler, L.L., Moritz, C.C., Brown, W.M., 1994. Sequences with the potential to form stem-and-loop structures are associated with coding-region duplications in animal mitochondrial DNA. Genetics 137, 233–241. https://doi.org/10.1093/GENETICS/137.1.233

Stephan, T., Burgess, S.M., Cheng, H., Danko, C.G., Gill, C.A., Jarvis, E.D., Koepfli, K.P., Koltes, J.E., Lyons, E., Ronald, P., Ryder, O.A., Schriml, L.M., Soltis, P., VandeWoude, S., Zhou, H., Ostrander, E.A., Karlsson, E.K., 2022. Darwinian genomics and diversity in the tree of life. Proc Natl Acad Sci U S A 119. https://doi.org/10.1073/pnas.2115644119

Sternecker, K., Geist, J., Beggel, S., Dietz-Laursonn, K., de La Fuente, M., Frank, H.G., Furia, J.P., Milz, S., Schmitz, C., 2018. Exposure of zebra mussels to extracorporeal shock waves demonstrates formation of new mineralized tissue inside and outside the focus zone. Biol Open 7. https://doi.org/10.1242/bio.033258

Stewart, Donald T, Breton, Sophie, Blier, Pierre U, Hoeh, Walter R, Stewart, D T, Breton, S, Hoeh, W R, Blier, P U, 2009. Masculinization Events and Doubly Uniparental Inheritance of Mitochondrial DNA: A Model for Understanding the Evolutionary Dynamics of Gender-Associated mtDNA in Mussels. Evol Biol 163–173. https://doi.org/10.1007/978-3-642-00952-5_9

Stewart, D.T., Kenchington, E.R., Singh, R.K., Zonros, E., 1996. Degree of Selective Constraint as an Explanation of the Different Rates of Evolution of Gender-Specific Mitochondrial DNA Lineages in the *Mussel Mytilus*. Genetics 143, 1349–1357. https://doi.org/10.1093/GENETICS/143.3.1349

Stöger, I., Kocot, K.M., Poustka, A.J., Wilson, N.G., Ivanov, D., Halanych, K.M., Schrödl, M., 2016. Monoplacophoran mitochondrial genomes: Convergent gene arrangements and little phylogenetic signal. BMC Evol Biol 16, 274. https://doi.org/10.1186/s12862-016-0829-3

Stöger, I., Schrödl, M., 2013. Mitogenomics does not resolve deep molluscan relationships (yet?). Mol Phylogenet Evol 69, 376–392. https://doi.org/10.1016/J.YMPEV.2012.11.017

Stone, J., Barndt, S., Gangloff, M., 2011. Spatial Distribution and Habitat Use of the Western Pearlshell Mussel (*Margaritifera falcata*) in a Western Washington Stream. http://dx.doi.org/10.1080/02705060.2004.9664907 19, 341–352. https://doi.org/10.1080/02705060.2004.9664907

Strack, E., 2015. European Freshwater Pearls: Part 1-Russia. The Journal of Gemmology 34, 580–592. https://doi.org/10.15506/jog.2015.34.7.580

Strayer, D.L., 2009. Twenty years of zebra mussels: lessons from the mollusk that made headlines. Front Ecol Environ 7, 135–141. https://doi.org/10.1890/080020

Strayer, D.L., 2008. Freshwater mussel ecology: a multifactor approach to distribution and abundance. Univ of California Press.

Strayer, D.L., Dudgeon, D., 2010. Freshwater biodiversity conservation: recent progress and future challenges. J North Am Benthol Soc 29, 344–358. https://doi.org/10.1899/08-171.1

Strong, E.E., Gargominy, O., Ponder, W.F., Bouchet, P., 2007. Global diversity of gastropods (Gastropoda; Mollusca) in freshwater. Freshwater Animal Diversity Assessment 149–166. https://doi.org/10.1007/978-1-4020-8259-7_17

Suleria, H.A.R., Masci, P.P., Gobe, G.C., Osborne, S.A., 2017. Therapeutic potential of abalone and status of bioactive molecules: A comprehensive review. http://dx.doi.org/10.1080/10408398.2015.1031726 57, 1742–1748. https://doi.org/10.1080/10408398.2015.1031726

Sun, J., Chen, C., Miyamoto, N., Li, R., Sigwart, J.D., Xu, T., Sun, Y., Wong, W.C., Ip, J.C.H., Zhang, W., Lan, Y., Bissessur, D., Watsuji, T., Watanabe, H.K., Takaki, Y., Ikeo, K., Fujii, N., Yoshitake, K., Qiu, J.-W., Takai, K., Qian, P.-Y., 2020. The Scaly-foot Snail genome and implications for the origins of biomineralised armour. Nature Communications 2020 11:1 11, 1–12. https://doi.org/10.1038/s41467-020-15522-3

Sun, J., Mu, H., Ip, J.C.H., Li, R., Xu, T., Accorsi, A., Alvarado, A.S., Ross, E., Lan, Y., Sun, Y., Castro-Vazquez, A., Vega, I.A., Heras, H., Ituarte, S., van Bocxlaer, B., Hayes, K.A., Cowie, R.H., Zhao, Z., Zhang, Y., Qian, P.Y., Qiu, J.W., 2019. Signatures of divergence, invasiveness, and terrestrialization revealed by four apple snail genomes. Mol Biol Evol 36, 1507–1520. https://doi.org/10.1093/molbev/msz084

Sun, J., Zhang, Yu, Xu, T., Zhang, Yang, Mu, H., Zhang, Yanjie, Lan, Y., Fields, C.J., Hui, J.H.L., Zhang, W., Li, R., Nong, W., Cheung, F.K.M., Qiu, J.-W., Qian, P.-Y., 2017. Adaptation to deep-sea chemosynthetic environments as revealed by mussel genomes. Nat Ecol Evol 1, 0121. https://doi.org/10.1038/s41559-017-0121

Sun, S., Kong, L., Yu, H., Li, Q., 2015. The complete mitochondrial DNA of Tegillarca granosa and comparative mitogenomic analyses of three Arcidae species. Gene 557, 61–70. https://doi.org/10.1016/J.GENE.2014.12.011

Sun, S., Kong, L., Yu, H., Li, Q., 2014. The complete mitochondrial genome of *Scapharca kagoshimensis* (Bivalvia: Arcidae). http://dx.doi.org/10.3109/19401736.2013.865174 26, 957–958. https://doi.org/10.3109/19401736.2013.865174

Sun, S., Li, Q., Kong, L., Yu, H., 2018. Multiple reversals of strand asymmetry in molluscs mitochondrial genomes, and consequences for phylogenetic inferences. Mol Phylogenet Evol 118, 222–231. https://doi.org/10.1016/J.YMPEV.2017.10.009

Swart, E.M., Davison, A., Ellers, J., Filangieri, R.R., Jackson, D.J., Mariën, J., van der Ouderaa, I.B.C., Roelofs, D., Koene, J.M., 2019. Temporal expression profile of an accessory-gland protein that is transferred via the seminal fluid of the simultaneous hermaphrodite *Lymnaea stagnalis*. Journal of Molluscan Studies 85, 177–183. https://doi.org/10.1093/mollus/eyz005

Swerdlow, H., Gesteland, R., 1990. Capillary gel electrophoresis for rapid, high resolution DNA sequencing. Nucleic Acids Res 18, 1415–1419. https://doi.org/10.1093/NAR/18.6.1415

Takeuchi, M., Okada, A., Kakino, W., 2015. Phylogenetic relationships of two freshwater pearl mussels, *Margaritifera laevis* (Haas, 1910) and *Margaritifera togakushiensis* Kondo & Kobayashi, 2005 (Bivalvia: Margaritiferidae), in the Japanese archipelago. http://dx.doi.org/10.1080/13235818.2015.1053165 35, 218–226. https://doi.org/10.1080/13235818.2015.1053165

Takeuchi, T., 2017. Molluscan Genomics: Implications for Biology and Aquaculture. Curr Mol Biol Rep 3, 297–305. https://doi.org/10.1007/s40610-017-0077-3

Takeuchi, T., Kawashima, T., Koyanagi, R., Gyoja, F., Tanaka, M., Ikuta, T., Shoguchi, E., Fujiwara, M., Shinzato, C., Hisata, K., Fujie, M., Usami, T., Nagai, K., Maeyama, K., Okamoto, K., Aoki, H., Ishikawa, T., Masaoka, T., Fujiwara, A., Endo, K., Endo, H., Nagasawa, H., Kinoshita, S., Asakawa, S., Watabe, S., Satoh, N., 2012. Draft Genome of the Pearl Oyster *Pinctada fucata*: A Platform for Understanding Bivalve Biology. DNA Research 19, 117–130. https://doi.org/10.1093/dnares/dss005

Takeuchi, T., Koyanagi, R., Gyoja, F., Kanda, M., Hisata, K., Fujie, M., Goto, H., Yamasaki, S., Nagai, K., Morino, Y., Miyamoto, H., Endo, K., Endo, H., Nagasawa, H., Kinoshita, S., Asakawa, S., Watabe, S., Satoh, N., Kawashima, T., 2016. Bivalve-specific gene expansion in the pearl oyster genome: implications of adaptation to a sessile lifestyle. Zoological Lett 2, 3. https://doi.org/10.1186/s40851-016-0039-2

Takeuchi, T., Suzuki, Y., Watabe, S., Nagai, K., Masaoka, T., Fujie, M., Kawamitsu, M., Satoh, N., Myers, E.W., 2022. A high-quality, haplotype-phased genome reconstruction reveals unexpected haplotype diversity in a pearl oyster. DNA Research 29, 1–13. https://doi.org/10.1093/DNARES/DSAC035

Talmage, S.C., Gobler, C.J., 2010. Effects of past, present, and future ocean carbon dioxide concentrations on the growth and survival of larval shellfish. Proc Natl Acad Sci U S A 107, 17246–51. https://doi.org/10.1073/pnas.0913804107

Tang, Y., Zheng, X., Ma, Y., Cheng, R., Li, Q., 2018. The complete mitochondrial genome of *Amphioctopus marginatus* (Cephalopoda: Octopodidae) and the exploration for the optimal DNA barcoding in Octopodidae. Conserv Genet Resour 10, 115–118. https://doi.org/10.1007/S12686-017-0777-2/FIGURES/2

Taylor, R.S., Manseau, M., Horn, R.L., Keobouasone, S., Golding, G.B., Wilson, P.J., 2020. The role of introgression and ecotypic parallelism in delineating intraspecific conservation units. Mol Ecol 29, 2793–2809. https://doi.org/10.1111/MEC.15522

Teacher, A.G., André, C., Merilä, J., Wheat, C.W., 2012. Whole mitochondrial genome scan for population structure and selection in the Atlantic herring. BMC Evol Biol 12, 1–14. https://doi.org/10.1186/1471-2148-12-248/FIGURES/5

Telford, M.J., Budd, G.E., 2003. The place of phylogeny and cladistics in Evo-Devo research. International Journal of Developmental Biology 47, 479–490. https://doi.org/10.1387/IJDB.14756323

Telford, M.J., Budd, G.E., Philippe, H., 2015. Phylogenomic Insights into Animal Evolution. Current Biology 25, R876–R887. https://doi.org/10.1016/J.CUB.2015.07.060

Terrett, J.A., Miles, S., Thomas, R.H., 1996. Complete DNA sequence of the mitochondrial genome of *Cepaea nemoralis* (Gastropoda: Pulmonata). Journal of Molecular Evolution 1996 42:2 42, 160–168. https://doi.org/10.1007/BF02198842

The Arabidopsis Genome Initiative, 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408, 796–815. https://doi.org/10.1038/35048692

Timmis, J.N., Ayliff, M.A., Huang, C.Y., Martin, W., 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nature Reviews Genetics 2004 5:2 5, 123–135. https://doi.org/10.1038/nrg1271

Toews, D.P.L., Mandic, M., Richards, J.G., Irwin, D.E., 2014. MIGRATION, MITOCHONDRIA, AND THE YELLOW-RUMPED WARBLER. Evolution (N Y) 68, 241–255. https://doi.org/10.1111/EVO.12260

Tomita, K., Yokobori, S. ichi, Oshima, T., Ueda, T., Watanabe, K., 2002. The Cephalopod *Loligo bleekeri* Mitochondrial Genome: Multiplied Noncoding Regions and Transposition of tRNA Genes. Journal of Molecular Evolution 2002 54:4 54, 486–500. https://doi.org/10.1007/S00239-001-0039-4

Uliano-Silva, M., Americo, J.A., Costa, I., Schomaker-Bastos, A., de Freitas Rebelo, M., Prosdocimi, F., 2016. The complete mitochondrial genome of the golden mussel *Limnoperna fortunei* and comparative mitogenomics of Mytilidae. Gene 577, 202–208. https://doi.org/10.1016/J.GENE.2015.11.043

Uliano-Silva, M., Dondero, F., Dan Otto, T., Costa, I., Lima, N.C.B., Americo, J.A., Mazzoni, C.J., Prosdocimi, F., Rebelo, M. de F., 2018. A hybrid-hierarchical genome assembly strategy to sequence the invasive golden mussel, *Limnoperna fortunei*. Gigascience 7, 1–10. https://doi.org/10.1093/GIGASCIENCE/GIX128

Uribe, J.E., González, V.L., Irisarri, I., Kano, Y., Herbert, D.G., Strong, E.E., Harasewych, M.G., 2022. A Phylogenomic Backbone for Gastropod Molluscs. Syst Biol 71, 1271–1280. https://doi.org/10.1093/SYSBIO/SYAC045

Uribe, J.E., Irisarri, I., Templado, J., Zardoya, R., 2019. New patellogastropod mitogenomes help counteracting long-branch attraction in the deep phylogeny of gastropod mollusks. Mol Phylogenet Evol 133, 12–23. https://doi.org/10.1016/J.YMPEV.2018.12.019

Uribe, J.E., Zardoya, R., 2017. Revisiting the phylogeny of Cephalopoda using complete mitochondrial genomes. Journal of Molluscan Studies 83, 133–144. https://doi.org/10.1093/MOLLUS/EYW052

van der Schatte Olivier, A., Jones, L., Vay, L. le, Christie, M., Wilson, J., Malham, S.K., 2020. A global review of the ecosystem services provided by bivalve aquaculture. Rev Aquac 12, 3–25. https://doi.org/10.1111/RAQ.12301

Varney, R.M., Speiser, D.I., McDougall, C., Degnan, B.M., Kocot, K.M., 2021. The Iron-Responsive Genome of the Chiton *Acanthopleura granulata*. Genome Biol Evol 13, 2020.05.19.102897. https://doi.org/10.1093/gbe/evaa263

Vaughn, C.C., 2017. Ecosystem services provided by freshwater mussels. Hydrobiologia 2017 810:1 810, 15–27. https://doi.org/10.1007/S10750-017-3139-X

Vaughn, C.C., Hakenkamp, C.C., 2001. The functional role of burrowing bivalves in freshwater ecosystems. Freshw Biol 46, 1431–1446. https://doi.org/10.1046/j.1365-2427.2001.00771.x

Vaughn, C.C., Hoellein, T.J., 2018. Bivalve Impacts in Freshwater and Marine Ecosystems. Annu Rev Ecol Evol Syst 49. https://doi.org/10.1146/annurev-ecolsys-110617-062703

Vaughn, C.C., Nichols, S.J., Spooner, D.E., 2015. Community and foodweb ecology of freshwater mussels. https://doi.org/10.1899/07-058.1 27, 409–423. https://doi.org/10.1899/07-058.1

Vieira, L.D., Silva-Junior, O.B., Novaes, E., Collevatti, R.G., 2022. Comparative population genomics in *Tabebuia alliance* shows evidence of adaptation in Neotropical tree species. Heredity 2022 128:3 128, 141–153. https://doi.org/10.1038/s41437-021-00491-0

Vikhrev, I. v., Makhrov, A.A., Artamonova, V.S., Ermolenko, A. v., Gofarov, M.Y., Kabakov, M.B., Kondakov, A. v., Chukhchin, D.G., Lyubas, A.A., Bolotov, I.N., 2019a. Fish hosts, glochidia features and life cycle of the endemic freshwater pearl mussel *Margaritifera dahurica* from the Amur Basin. Scientific Reports 2019 9:1 9, 1–13. https://doi.org/10.1038/s41598-019-44752-9

Vikhrev, I. v., Makhrov, A.A., Artamonova, V.S., Ermolenko, A. v., Gofarov, M.Y., Kabakov, M.B., Kondakov, A. v., Chukhchin, D.G., Lyubas, A.A., Bolotov, I.N., 2019b. Fish hosts, glochidia features and life cycle of the endemic freshwater pearl mussel *Margaritifera* *dahurica* from the Amur Basin. Scientific Reports 2019 9:1 9, 1–13. https://doi.org/10.1038/s41598-019-44752-9

Vinther, J., Sperling, E.A., Briggs, D.E.G., Peterson, K.J., 2012. A molecular palaeobiological hypothesis for the origin of aplacophoran molluscs and their derivation from chiton-like ancestors. Proceedings of the Royal Society B: Biological Sciences 279, 1259–1268. https://doi.org/10.1098/rspb.2011.1773

Voolstra, C.R., GIGA Community of Scientists (COS), Wörheide, G., Lopez, J. v., 2017. Corrigendum to: Advancing genomics through the Global Invertebrate Genomics Alliance (GIGA). Invertebr Syst 31, 231. https://doi.org/10.1071/IS16059_CO

Walker, J.M., Curole, J.P., Wade, D.E., Chapman, E.G., Bogan, A.E., Watters, G.T., Hoeh, W.R., 2006. Taxonomic distribution and phylogenetic utility of gender-associated mitochondrial genomes in the Unionoida (Bivalvia). MALACOLOGIA-PHILADELPHIA- 48, 265.

Walters, E.T., Moroz, L.L., 2009. Molluscan memory of injury: Evolutionary insights into Chronic pain and neurological disorders. Brain Behav Evol. https://doi.org/10.1159/000258667

Wang, M., Kong, L., 2019. pblat: A multithread blat algorithm speeding up aligning sequences to genomes. BMC Bioinformatics 20, 28. https://doi.org/10.1186/s12859-019-2597-8

Wang, R., Li, C., Stoeckel, J., Moyer, G., Liu, Z., Peatman, E., 2015. Rapid development of molecular resources for a freshwater mussel, *Villosa lienosa* (Bivalvia:Unionidae), using an RNA-seq-based approach. https://doi.org/10.1899/11-149.1 31, 695–708. https://doi.org/10.1899/11-149.1

Wang, S., Zhang, J., Jiao, W., Li, J., Xun, X., Sun, Y., Guo, X., Huan, P., Dong, B., Zhang, L., Hu, X., Sun, X., Wang, J., Zhao, C., Wang, Y., Wang, D., Huang, X., Wang, R., Lv, J., Li, Yuli, Zhang, Z., Liu, B., Lu, W., Hui, Y., Liang, J., Zhou, Z., Hou, R., Li, Xue, Liu, Y., Li, H., Ning, X., Lin, Y., Zhao, L., Xing, Q., Dou, J., Li, Yangping, Mao, J., Guo, H., Dou, H., Li, T., Mu, C., Jiang, W., Fu, Q., Fu, X., Miao, Y., Liu, J., Yu, Q., Li, Ruojiao, Liao, H., Li, Xuan, Kong, Y., Jiang, Z., Chourrout, D., Li, Ruiqiang, Bao, Z., 2017. Scallop genome provides insights into evolution of bilaterian karyotype and development. Nat Ecol Evol 1, 0120. https://doi.org/10.1038/s41559-017-0120

Wang, X., Liu, Z., Wu, W., 2017. Transcriptome analysis of the freshwater pearl mussel (*Cristaria plicata*) mantle unravels genes involved in the formation of shell and pearl.

Molecular Genetics and Genomics 292, 343–352. https://doi.org/10.1007/S00438-016-1278-9/FIGURES/4

Wanninger, A., Wollesen, T., 2019. The evolution of molluscs. Biological Reviews 94, 102–115. https://doi.org/10.1111/brv.12439

Watson, J.D., Crick, F.H.C., 1953. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. Nature 1953 171:4356 171, 737–738. https://doi.org/10.1038/171737a0

Wendel, J.F., Greilhuber, J., Dolezel, J., Leitch, I.J., 2012. Plant genome diversity volume 1: Plant genomes, their residents, and their evolutionary dynamics. Springer.

Wenger, A.M., Peluso, P., Rowell, W.J., Chang, P.C., Hall, R.J., Concepcion, G.T., Ebler, J., Fungtammasan, A., Kolesnikov, A., Olson, N.D., Töpfer, A., Alonge, M., Mahmoud, M., Qian, Y., Chin, C.S., Phillippy, A.M., Schatz, M.C., Myers, G., DePristo, M.A., Ruan, J., Marschall, T., Sedlazeck, F.J., Zook, J.M., Li, H., Koren, S., Carroll, A., Rank, D.R., Hunkapiller, M.W., 2019. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. Nature Biotechnology 2019 37:10 37, 1155–1162. https://doi.org/10.1038/s41587-019-0217-9

Whelan, N., Whelan, N. v, Johnson, P.D., Garner, J.T., Garrison, N.L., Strong, E.E., 2022. Prodigious polyphyly in Pleuroceridae (Gastropoda: Cerithioidea). Bulletin of the Society of Systematic Biologists 1. https://doi.org/10.18061/BSSB.V1I2.8419

Whelan, N. v., Geneva, A.J., Graf, D.L., 2011. Molecular phylogenetic analysis of tropical freshwater mussels (Mollusca: Bivalvia: Unionoida) resolves the position of *Coelatura* and supports a monophyletic Unionidae. Mol Phylogenet Evol 61, 504–514. https://doi.org/10.1016/J.YMPEV.2011.07.016

White, T.R., Conrad, M.M., Tseng, R., Balayan, S., Golding, R., de Frias Martins, A., Dayrat, B.A., 2011. Ten new complete mitochondrial genomes of pulmonates (Mollusca: Gastropoda) and their impact on phylogenetic relationships. BMC Evol Biol 11, 1–15. https://doi.org/10.1186/1471-2148-11-295/FIGURES/4

Williams, J.D., Bogan, A.E., Butler, R.S., Cummings, K.S., Garner, J.T., Harris, J.L., Johnson, N.A., Watters, G.T., 2017. A Revised List of the Freshwater Mussels (Mollusca: Bivalvia: Unionida) of the United States and Canada. https://doi.org/10.31931/fmbc.v20i2.2017.33-58 20, 33–58. https://doi.org/10.31931/FMBC.V20I2.2017.33-58

Williams, S.T., Foster, P.G., Hughes, C., Harper, E.M., Taylor, J.D., Littlewood, D.T.J., Dyal, P., Hopkins, K.P., Briscoe, A.G., 2017. Curious bivalves: Systematic utility and unusual properties of anomalodesmatan mitochondrial genomes. Mol Phylogenet Evol 110, 60–72. https://doi.org/10.1016/J.YMPEV.2017.03.004

Winson F, P., David R, L., 2008. Phylogeny and Evolution of the Mollusca. University of California Press. https://doi.org/10.1525/california/9780520250925.001.0001

Woese, C.R., Fox, G.E., 1977. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. Proc Natl Acad Sci U S A 74, 5088–5090. https://doi.org/10.1073/PNAS.74.11.5088/SUPPL_FILE/SCIENCEOFMICROBESPODCAST.MP3

Woody, C.A., Holland-Bartels, L., 2011. Reproductive Characteristics of a Population of the Washboard Mussel *Megalonaias nervosa* (Rafinesque 1820) in the Upper Mississippi River. http://dx.doi.org/10.1080/02705060.1993.9664724 8, 57–66. https://doi.org/10.1080/02705060.1993.9664724

Wu, R., Kaiser, A.D., 1968. Structure and base sequence in the cohesive ends of bacteriophage lambda DNA. J Mol Biol 35, 523–537. https://doi.org/10.1016/S0022-2836(68)80012-9

Wu, R., Taylor, E., 1971. Nucleotide sequence analysis of DNA: II. Complete nucleotide sequence of the cohesive ends of bacteriophage λ DNA. J Mol Biol 57, 491–511. https://doi.org/10.1016/0022-2836(71)90105-7

Wu, R.W., Liu, X.J., Wang, S., Roe, K.J., Ouyang, S., Wu, X.P., 2019. Analysis of mitochondrial genomes resolves the phylogenetic position of Chinese freshwater mussels (Bivalvia, Unionidae). ZooKeys 812: 23-46 812, 23–46. https://doi.org/10.3897/ZOOKEYS.812.29908

Wu, X., Li, X., Li, L., Xu, X., Xia, J., Yu, Z., 2012a. New features of Asian *Crassostrea oyster* mitochondrial genomes: A novel alloacceptor tRNA gene recruitment and two novel ORFs. Gene 507, 112–118. https://doi.org/10.1016/J.GENE.2012.07.032

Wu, X., Li, X., Li, L., Yu, Z., 2012b. A unique tRNA gene family and a novel, highly expressed ORF in the mitochondrial genome of the silver-lip pearl oyster, *Pinctada maxima* (Bivalvia: Pteriidae). Gene 510, 22–31. https://doi.org/10.1016/J.GENE.2012.08.037

Wu, X., Li, X., Yu, Z., 2015. The mitochondrial genome of the scallop *Mimachlamys senatoria* (Bivalvia, Pectinidae). http://dx.doi.org/10.3109/19401736.2013.823181 26, 242–244. https://doi.org/10.3109/19401736.2013.823181

Wu, X., Xu, X., Yu, Z., Kong, X., 2009. Comparative mitogenomic analyses of three scallops (Bivalvia: Pectinidae) reveal high level variation of genomic organization and a diversity of transfer RNA gene sets. BMC Res Notes 2, 1–7. https://doi.org/10.1186/1756-0500-2-69/FIGURES/3

Wyman, S.K., Jansen, R.K., Boore, J.L., 2004. Automatic annotation of organellar genomes with DOGMA. Bioinformatics 20, 3252–3255. https://doi.org/10.1093/BIOINFORMATICS/BTH352

Xie, G.L., Köhler, F., Huang, X.C., Wu, R.W., Zhou, C.H., Ouyang, S., Wu, X.P., 2019. A novel gene arrangement among the Stylommatophora by the complete mitochondrial genome of the terrestrial slug *Meghimatium bilineatum* (Gastropoda, Arionoidea). Mol Phylogenet Evol 135, 177–184. https://doi.org/10.1016/J.YMPEV.2019.03.002

Xin, Y., Ren, J., Liu, X., 2011. Mitogenome of the small abalone *Haliotis diversicolor* Reeve and phylogenetic analysis within Gastropoda. Mar Genomics 4, 253–262. https://doi.org/10.1016/J.MARGEN.2011.06.005

Xu, X., Wu, X., Yu, Z., 2010. The mitogenome of *Paphia euglypta* (Bivalvia: Veneridae) and comparative mitogenomic analyses of three venerids. Genome 53, 1041–1052. https://doi.org/10.1139/G10-096/SUPPL_FILE/G10-096ESUPPL.DOCX

Yamazaki, N., Ueshima, R., Terrett, J.A., Yokobori, S.I., Kaifu, M., Segawa, R., Kobayashi, T., Numachi, K.I., Ueda, T., Nishikawa, K., Watanabe, K., Thomas, R.H., 1997. Evolution of Pulmonate Gastropod Mitochondrial Genomes: Comparisons of Gene Organizations of Euhadra, Cepaea and Albinaria and Implications of Unusual tRNA Secondary Structures. Genetics 145, 749–758. https://doi.org/10.1093/GENETICS/145.3.749

Yan, X., Nie, H., Huo, Z., Ding, J., Li, Z., Yan, L., Jiang, L., Mu, Z., Wang, H., Meng, X., Chen, P., Zhou, M., Rbbani, Md.G., Liu, G., Li, D., 2019. Clam Genome Sequence Clarifies the Molecular Basis of Its Benthic Adaptation and Extraordinary Shell Color Diversity. iScience 19, 1225–1237. https://doi.org/10.1016/j.isci.2019.08.049

Yandell, M., Ence, D., 2012. A beginner’s guide to eukaryotic genome annotation. Nat Rev Genet 13, 329–342. https://doi.org/10.1038/nrg3174

Yang, Q., Guo, K., Zhou, X., Tang, X., Yu, X., Yao, W., Wu, Z., 2021. Histopathology, antioxidant responses, transcriptome and gene expression analysis in triangle sail mussel *Hyriopsis cumingii* after bacterial infection. Dev Comp Immunol 124, 104175. https://doi.org/10.1016/j.dci.2021.104175

Yang, S., Mi, Z., Tao, G., Liu, X., Wei, M., Wang, H., 2014. The complete mitochondrial genome sequence of *Margaritiana dahurica* Middendorff. http://dx.doi.org/10.3109/19401736.2013.845755 26, 716–717. https://doi.org/10.3109/19401736.2013.845755

Yang, Z., Zhang, L., Hu, J., Wang, J., Bao, Z., Wang, S., 2020. The evo-devo of molluscs: Insights from a genomic perspective. Evol Dev 22, 409–424. https://doi.org/10.1111/EDE.12336

Yarra, T., Blaxter, M., Clark, M.S., 2021. A Bivalve Biomineralization Toolbox. Mol Biol Evol 38, 4043–4055. https://doi.org/10.1093/molbev/msab153

Ye, Y.Y., Wu, C.W., Li, J.J., 2015. Genetic Population Structure of *Macridiscus multifarius* (Mollusca: Bivalvia) on the Basis of Mitochondrial Markers: Strong Population Structure in a Species with a Short Planktonic Larval Stage. PLoS One 10, e0146260. https://doi.org/10.1371/JOURNAL.PONE.0146260

Yokobori, S., Iseto, T., Asakawa, S., Sasaki, T., Shimizu, N., Yamagishi, A., Oshima, T., Hirose, E., 2008. Complete nucleotide sequences of mitochondrial genomes of two solitary entoprocts, *Loxocorone allax* and *Loxosomella aloxiata*: Implications for lophotrochozoan phylogeny. Mol Phylogenet Evol 47, 612–628. https://doi.org/10.1016/J.YMPEV.2008.02.013

Yokobori, S.I., Fukuda, N., Nakamura, M., Aoyama, T., Oshima, T., 2004. Long-Term Conservation of Six Duplicated Structural Genes in Cephalopod Mitochondrial Genomes. Mol Biol Evol 21, 2034–2046. https://doi.org/10.1093/MOLBEV/MSH227

Yoshida, M.A., Ogura, A., Ikeo, K., Shigeno, S., Moritaki, T., Winters, G.C., Kohn, A.B., Moroz, L.L., 2015. Molecular Evidence for Convergence and Parallelism in Evolution of Complex Brains of Cephalopod Molluscs: Insights from Visual Systems. Integr Comp Biol 55, 1070–1083. https://doi.org/10.1093/ICB/ICV049

Young, A.D., Lemmon, A.R., Skevington, J.H., Mengual, X., Ståhls, G., Reemer, M., Jordaens, K., Kelso, S., Lemmon, E.M., Hauser, M., de Meyer, M., Misof, B., Wiegmann, B.M., 2016. Anchored enrichment dataset for true flies (order Diptera) reveals insights into the phylogeny of flower flies (family Syrphidae). BMC Evol Biol 16, 1–13. https://doi.org/10.1186/S12862-016-0714-0/FIGURES/1

Yuan, M.L., Zhang, L.J., Zhang, Q.L., Zhang, L., Li, M., Wang, X.T., Feng, R.Q., Tang, P.A., 2020. Mitogenome evolution in ladybirds: Potential association with dietary adaptation. Ecol Evol 10, 1042–1053. https://doi.org/10.1002/ECE3.5971

Yuan, Y., Li, Q., Kong, L., Yu, H., 2012. The complete mitochondrial genome of the grand jackknife clam, *Solen grandis* (Bivalvia: Solenidae): A novel gene order and unusual non-coding region. Mol Biol Rep 39, 1287–1292. https://doi.org/10.1007/S11033-011-0861-8/FIGURES/3

Yusa, Y., Breton, S., Hoeh, W.R., 2013. Population Genetics of Sex Determination in *Mytilus* Mussels: Reanalyses and a Model. Journal of Heredity 104, 380–385. https://doi.org/10.1093/JHERED/EST014

Zallen, D.T., 2003. Despite Franklin's work, Wilkins earned his Nobel. Nature 2003 425:6953 425, 15–15. https://doi.org/10.1038/425015b

Zampini, É., Lepage, É., Tremblay-Belzile, S., Truche, S., Brisson, N., 2015. Organelle DNA rearrangement mapping reveals U-turn-like inversions as a major source of genomic instability in *Arabidopsis* and humans. Genome Res 25, 645–654. https://doi.org/10.1101/GR.188573.114

Zanatta, D.T., Stoeckle, B.C., Inoue, K., Paquet, A., Martel, A.L., Kuehn, R., Geist, J., 2018. High genetic diversity and low differentiation in north american *Margaritifera margaritifera* (Bivalvia: Unionida: Margaritiferidae). Biological Journal of the Linnean Society 123, 850–863. https://doi.org/10.1093/biolinnean/bly010

Zapata, F., Wilson, N.G., Howison, M., Andrade, S.C.S., Jörger, K.M., Schrödl, M., Goetz, F.E., Giribet, G., Dunn, C.W., 2014. Phylogenomic analyses of deep gastropod relationships reject Orthogastropoda. Proceedings of the Royal Society B: Biological Sciences 281, 20141739. https://doi.org/10.1098/rspb.2014.1739

Zardoya, R., Pérez-Martos, A., Bautista, J.M., Montoya, J., 1995. Analysis of the transcription products of the rainbow trout (*Oncorynchus mykiss*) liver mitochondrial genome: detection of novel mitochondrial transcripts. Current Genetics 1995 28:1 28, 67–70. https://doi.org/10.1007/BF00311883

Zarrella, I., Herten, K., MAES, G., Tai, S., Yang, M., Seuntjens, E., Ritschard, E., Zach, M., Styfhals, R., Sanges, R., Simakov, O., Ponte, G., Fiorito, G., 2019. The survey and reference assisted assembly of the *Octopus vulgaris* genome. Sci Data 6, 13. https://doi.org/10.1038/s41597-019-0017-6

Zbawicka, M., Wenne, R., Burzyński, A., 2014. Mitogenomics of recombinant mitochondrial genomes of Baltic Sea *Mytilus* mussels. Molecular Genetics and Genomics 289, 1275–1287. https://doi.org/10.1007/S00438-014-0888-3/TABLES/5

Zeng, X., Hourset, A., Tzagoloff, A., 2007. The Saccharomyces cerevisiae ATP22 Gene Codes for the Mitochondrial ATPase Subunit 6-Specific Translation Factor. Genetics 175, 55–63. https://doi.org/10.1534/GENETICS.106.065821

Zhang, C., Rabiee, M., Sayyari, E., Mirarab, S., 2018. ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics 19, 153. https://doi.org/10.1186/s12859-018-2129-y

Zhang, Guofan, Fang, X., Guo, X., Li, L., Luo, R., Xu, F., Yang, P., Zhang, L., Wang, X., Qi, H., Xiong, Z., Que, H., Xie, Y., Holland, P.W.H., Paps, J., Zhu, Y., Wu, F., Chen, Y., Wang, Jiafeng, Peng, C., Meng, J., Yang, L., Liu, J., Wen, B., Zhang, N., Huang, Z., Zhu, Q., Feng, Y., Mount, A., Hedgecock, D., Xu, Z., Liu, Y., Domazet-Lošo, T., Du, Y., Sun, X., Zhang, Shoudu, Liu, B., Cheng, P., Jiang, X., Li, J., Fan, D., Wang, W., Fu, W., Wang, T., Wang, B., Zhang, J., Peng, Z., Li, Yingxiang, Li, Na, Wang, Jinpeng, Chen, M., He, Y., Tan, F., Song, X., Zheng, Q., Huang, R., Yang, Hailong, Du, X., Chen, L., Yang, M., Gaffney, P.M., Wang, S., Luo, L., She, Z., Ming, Y., Huang, W., Zhang, Shu, Huang, B., Zhang, Y., Qu, T., Ni, P., Miao, G., Wang, Junyi, Wang, Q., Steinberg, C.E.W., Wang, H., Li, Ning, Qian, L., Zhang, Guojie, Li, Yingrui, Yang, Huanming, Liu, X., Wang, Jian, Yin, Y., Wang, Jun, 2012. The oyster genome reveals stress adaptation and complexity of shell formation. Nature 490, 49–54. https://doi.org/10.1038/nature11413

Zhang, J., Lindsey, A.R.I., Peters, R.S., Heraty, J.M., Hopper, K.R., Werren, J.H., Martinson, E.O., Woolley, J.B., Yoder, M.J., Krogmann, L., 2020. Conflicting signal in transcriptomic markers leads to a poorly resolved backbone phylogeny of chalcidoid wasps. Syst Entomol 45, 783–802. https://doi.org/10.1111/SYEN.12427

Zhang, W., Chen, J., Yang, Y., Tang, Y., Shang, J., Shen, B., 2011. A Practical Comparison of De Novo Genome Assembly Software Tools for Next-Generation Sequencing Technologies. PLoS One 6, e17915. https://doi.org/10.1371/journal.pone.0017915

Zhao, T., Xue, J., Kao, S., Li, Z., Zwaenepoel, A., Schranz, M.E., Peer, Y. van de, 2020. Novel phylogeny of angiosperms inferred from whole-genome microsynteny analysis. bioRxiv 2020.01.15.908376. https://doi.org/10.1101/2020.01.15.908376

Zheng, R., Li, J., Niu, D., 2010. The complete DNA sequence of the mitochondrial genome of *Sinonovacula constricta* (Bivalvia: Solecurtidae). Acta Oceanologica Sinica 2010 29:2 29, 88–92. https://doi.org/10.1007/S13131-010-0026-Y

Zhu, H.C., Shen, H.D., Zheng, P., Zhang, Y., 2012. Complete mitochondrial genome of the jackknife clam *Solen grandis* (Veneroida, Solenidae). http://dx.doi.org/10.3109/19401736.2011.653803 23, 115–117. https://doi.org/10.3109/19401736.2011.653803

Zouros, E., 2013. Biparental Inheritance Through Uniparental Transmission: The Doubly Uniparental Inheritance (DUI) of Mitochondrial DNA. Evol Biol 40, 1–31. https://doi.org/10.1007/s11692-012-9195-2

Zouros, E., 2012. Biparental Inheritance Through Uniparental Transmission: The Doubly Uniparental Inheritance (DUI) of Mitochondrial DNA. Evolutionary Biology 2012 40:1 40, 1–31. https://doi.org/10.1007/S11692-012-9195-2

Zouros, E., 2000. The exceptional mitochondrial DNA system of the mussel family Mytilidae. Genes Genet Syst 75, 313–318. https://doi.org/10.1266/GGS.75.313

Zouros, E., Ball, A.O., Saavedra, C., Freeman, K.R., 1994. Mitochondrial DNA inheritance. Nature 368, 818. https://doi.org/10.1038/368818a0