

**A glimpse into mollusk genomics from the *de novo*  
genome of *Arion vulgaris* Moquin-Tandon, 1855  
(Gastropoda, Stylommatophora)**

Dissertation

zur Erlangung des Doktorgrades der Naturwissenschaften (Dr. rer. nat.)

der Fakultät für Biologie

der Ludwig-Maximilians-Universität München



vorgelegt von

Zeyuan Chen

München, 24. Oktober 2022

Erstgutachter: Prof. Dr. Michael Schrödl

Zweitgutachter: Prof. Dr. Gerhard Haszprunar

Datum der Einreichung: 24. Oktober 2022

Datum der mündlichen Prüfung: 27. Februar 2023

This study was funded by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 764840





*für meine Eltern Yayong, Xiaoru und meine Schwester Zefang*



## TABLE OF CONTENTS

<b>1. ABSTRACT</b>	<b>8</b>
<b>2. INTRODUCTION</b>	<b>10</b>
<b>2.1 The treasure of evolution—mollusks</b>	<b>10</b>
<b>2.2 Complexity and inconsistency of molluscan phylogenetics</b>	<b>11</b>
<b>2.3 Towards molluscan genomics</b>	<b>12</b>
2.3.1 Development of genome sequencing strategy	12
2.3.2 Underrepresented genetic resources of mollusks	13
2.3.3 Unbalanced genetic resources among mollusks	15
2.3.4 Challenges and genomic features of mollusks	16
2.3.5 Concerns of currently published mollusk genomic studies	17
<b>2.4 Aims of this thesis</b>	<b>18</b>
<b>3. RESULTS</b>	<b>20</b>
<b>3.1 Publication I</b>	<b>20</b>
Özgül Doğan, Michael Schrödl, <b>Zeyuan Chen</b> (2019): The complete mitogenome of <i>Arion vulgaris</i> Moquin-Tandon, 1855 (Gastropoda: Stylommatophora): mitochondrial genome architecture, evolution and phylogenetic considerations within Stylommatophora. <i>PeerJ</i> 8:e8603	
Supplemental Tables and Figures for Publication I	51
<b>3.2 Publication II</b>	<b>72</b>
<b>Zeyuan Chen</b> , Michael Schrödl (2022): How many single-copy orthologous genes from whole genomes reveal deep gastropod relationships? <i>PeerJ</i> 10:e13285	
Supplemental Tables and Figures for Publication II	87
<b>3.3 Publication III</b>	<b>96</b>
<b>Zeyuan Chen</b> , Özgül Doğan, Nadège Guiguelmoni, Anne Guichard, Michael Schrödl (2022): Pulmonate slug evolution is reflected in the <i>de novo</i> genome of <i>Arion vulgaris</i> Moquin-Tandon, 1855. <i>Scientific Reports</i> 12, 14226	
Supplemental Tables and Figures for Publication III	113
<b>4. DISCUSSION</b>	<b>169</b>
<b>4.1 From <i>A. vulgaris</i> to mollusk mitogenomes</b>	<b>169</b>
<b>4.2 From gastropod to molluscan phylogeny</b>	<b>170</b>
<b>4.3 Deep dive into the <i>A. vulgaris</i> genome</b>	<b>171</b>
4.3.1 Calmodulin-like proteins are reduced/lost in <i>A. vulgaris</i> genome	171

4.3.2 Mucin related genes are expanded in land snails and slugs	174
<b>5. CONCLUSION AND OUTLOOK</b>	<b>176</b>
<b>6. REFERENCES</b>	<b>177</b>
<b>7. ADDITIONAL FILE: Statistics for Mollusca genomes</b>	<b>186</b>
<b>8. ACKNOWLEDGEMENTS</b>	<b>199</b>
<b>9. APPENDIX</b>	<b>200</b>
<b>9.1 Declaration of contributions to each publication</b>	<b>200</b>
<b>9.2 Curriculum Vitae</b>	<b>201</b>
<b>9.3 Eidesstattliche Versicherung und Erklärung</b>	<b>203</b>

## 1. Abstract

Mollusks are among nature's greatest innovators, representing some of the most ancient and evolutionary successful animals, with a tremendous diversity in morphology, behavior, lifestyle and habitats. Since the beginning of human civilization, this taxon has received a great deal of attention, for example as ornaments, currency, and diet, and is nowadays closely linked to the economy, health and ecology of human society. For more than two centuries, malacologists have laid a solid foundation for our understanding on the evolution of the phylum Mollusca through behavioral and environmental observations, as well as morphological and anatomical studies. The beginning of the era of genetic and genomic reconstruction of natural history has allowed gaining a complementary perspective on molluscan diversity and evolution.

A genome contains the complete genetic information of an organism or a cell, passed from one generation to the next. A reference genome, i.e., a representative nucleic acid sequence database of a target species, reveals the structure, organization and functional features of genes, and therefore serves as a foundation for molecular and genomic studies. Recent improvements in sequencing technologies have provided the ability to produce high-quality genome assemblies in a shorter period of time and with affordable costs. However, relative to the huge number of extant species, molluscan genomic resources lag far behind, and are highly uneven across classes and taxa. For example, up to now (August 2022) no sequenced and assembled genomes are available for four out of the eight mollusk classes (Scaphopoda, Solenogastres, Caudofoveata, Monoplacophora). Gastropods account for 80% of extant species in the phylum Mollusca. While the total number of reported gastropod genomes (49) accounts for 45% of published mollusk genomes, this only represents 0.05% of all known gastropod species.

During the course of my dissertation, we generated whole-genome and transcriptome data for *Arion vulgaris* Moquin-Tandon, 1855 (Mollusca: Gastropoda: Stylommatophora), a notorious agricultural pest in Europe and thus an ecologically and economically important species. We assembled a chromosome-level genome of *A. vulgaris* with considerable completeness and contiguous, annotated genes and repetitive elements, as well as its mitochondrial genome. Using the mitogenome, we explored the mitochondrial evolutionary patterns as well as phylogenetic relationships of stylommatophoran land snails and slugs. Using whole-genome data, we identified single-copy orthologous genes among major gastropod groups, and recovered Psilogastropoda *s.l.*, that is Patellogastropoda as sister to Vetigastropoda and Neomphalina, for deep gastropod phylogeny. As the first reported genome of a land slug, we have also made extensive comparisons between the genome of *Arion vulgaris* with land snails and other aquatic and marine species of the clade Heterobranchia. We identified genes that are specific/expanded/positively selected in *A. vulgaris*, which show functional relatedness to its strong adaptive capacities. Furthermore, we showed that the whole-genome duplication event that occurred approximately 93–109 Mya shared by Stylommatophora species might have promoted stylommatophoran land



invasion, speciation, and adaptive radiation. Small-scale gene duplication, i.e., a formidable recent expansion of transposable elements of *A. vulgaris*, might have additionally driven its genetic innovation and quick adaptation, promoting invasion of new areas and success in changing habitats and environments.

## 2. Introduction

### 2.1 The treasure of evolution—mollusks

Dating back to 580 million years ago, during the Precambrian, the common ancestor of living mollusks has entered the stage of history (Kocot et al. 2020; Slater & Bohlin 2022; Vinther 2015). The first mollusk was probably a small animal with a single non-mineralized “shell”/cuticula and a rasping tongue (radula with scraping-type feeding) (Haszprunar 1992; Wanninger & Wollesen 2019, but see Kocot et al. 2020 on *Kimberella*-type large stem group mollusks). However, this unremarkable creature was able to compete with other successful Cambrian metazoans like arthropods and brachiopods; its descendants survived several mass extinctions events and underwent significant adaptive radiation (Barnosky et al. 2011; Crick et al. 2020). Today, Mollusca is the second largest phylum in the animal kingdom, with around 130,000 living and 70,000 fossil species described, and an estimated actual extant diversity around 200,000 (Haszprunar et al. 2008).

Mollusks thus have evolved tremendous diversity of morphology, behavior and lifestyle over a long-time span (Hochner & Glanzman 2016). The extant mollusks are recognized as divided into eight classes: worm-like and shell-less marine mollusks Solenogastres and Caudofoveata, Monoplacophora (*sensu* Tryblidia) with a single, cup-like true shell, Polyplacophora with eight serial shell plates, Scaphopoda with elephant tusk-like shells, Cephalopoda (nautilus, sepias, squids and octopuses) with external or internal, reduced or lost shells and well developed senses and large brains, Bivalvia (e.g., mussels and clams) with bisected shells connected dorsally by a hinge, and Gastropoda (such as snails, whelks, slugs, and limpets) with an asymmetrical body showing the effects of torsion (Haszprunar & Wanninger 2012). The body size of mollusks ranges from 0.4 mm (omalogyrid gastropods) to over 7 m (colossal squids), lifespan ranges from few months to 500 years, e.g., bivalve *Arctica islandica* (Linnaeus, 1767) (Butler et al. 2013). The phylum includes some of the slowest or permanently cemented groups, such as most Bivalvia, feeding through sedentary filtering (Kappes & Haase 2011), and some of the swiftest carnivorous predators, e.g., octopuses, with the largest nervous systems of any invertebrates, complex behaviors such as instantaneous camouflage, and arms studded with dexterous suckers (Villanueva et al. 2017). The diversity of molluscan life forms has become variously adapted for a great range of habitats, from the tropics to polar seas, from the coasts to the deepest part of the ocean exceeding 7000 m. Mollusks originated in the ocean and most species still are marine, only bivalves and gastropods have invaded freshwater habitats, and Gastropoda is the only group which conquered terrestrial habitats, several times independently (Aktipis et al. 2008).

Mollusks are highly related to human life. For centuries, mollusks have been used commercially and artistically, for example as currency or ornaments. Also, mollusks comprise a significant share of seafood (mussels, squids, clams, octopuses, snails), which are a rich source of protein with essential amino acids, micro/macro minerals

and vitamins (Chakraborty & Joy 2020). In recent years, some of the mollusks' metabolic components have also been used medically or pharmaceutically, such as nutraceuticals, potential anti-inflammatory ingredients, or as an analgesic (Chakraborty & Joy 2020; Pereira 2016). Bivalves (oysters, clams, mussels, etc.) also play an important role in the ecosystem, such as water purification, carbon sequestration, nutrient remediation and so on (Schatte Olivier et al. 2018). However, there are also negative impacts. Many freshwater gastropod species act as intermediate hosts for various human-pathogenic trematodes, such as schistosomiasis (Fried & Huffman 2017), and some bivalve and gastropod mollusks are invasive species and agricultural pests, causing huge ecological and economical damage (Herbert et al. 2016; Sin 2003; Zemanova et al. 2017).

## **2.2 Complexity and inconsistency of molluscan phylogenetics**

As mollusks are an extremely ancient, diverse, successful and important phylum, they have been employed as excellent models in studies of developmental biology, neurobiology, physiology, evolution and population genetics for over 100 years (Davison & Neiman 2021). The understanding of their biology and evolution provides clues to answering fundamental questions on genotypic versus phenotypic adaptation, origins of evolutionary novelties, and macroevolutionary processes, which have been studied extensively by morphologists, taxonomists and paleontologists. However, despite the fact that so many researchers have studied mollusks, their old age and aberrant morphology have made it difficult to resolve their relationships: firstly, within the system of (Lopho)Trochozoa (Kocot 2016) but also among the eight classes of mollusks (Haszprunar & Wanninger 2012; Stöger et al 2013, and in many cases on order and family level (Haszprunar 1988; Korshunova et al. 2017).

With the development of novel DNA sequencing technology, molecular data are being applied to gradually understand the evolution and phylogeny of mollusks: from the use of single or several molecular markers, like e.g., mitochondrial cytochrome c oxidase subunit I (COI) gene, 16S rRNA, and nuclear 18S rRNA to the application of mitochondrial genomes, large-scale transcriptome data, ultra-conserved elements (Pierce 2019), and whole-genome data (Gomes-dos-Santos et al. 2019). Although there are still large conflicts and controversies in different studies (regarding the choice of included taxa and the type of DNA markers), and sometimes challenging traditional views of molluscan evolutionary histories (Formaggioni et al. 2022; Haszprunar & Wanninger 2012; Schrödl & Stöger 2014; Sigwart & Lindberg 2015), phylogenetic inferences based on molecular level have greatly contributed to the understanding of mollusk relationships. Challenging traditional concepts of molluscan evolution, recent large-scale phylogenomic studies converge towards the Aculifera hypothesis with polyplacophorans sister to the aplacophoran Solenogastres and Caudofoveata (Kocot et al. 2011; Smith et al. 2022) and, e.g., recover Monoplacophora as sister to all other conchiferans (Gastropoda+Bivalvia+Scaphopoda+Cephalopoda) with strong support (Kocot et al. 2020), as had been proposed based on morphology (Haszprunar 2000). Transcriptomic data from large taxon sampling by Cunha & Giribet (2019) and Zapata

et al. (2014) resulted in congruent topologies for deep gastropod relationships, while a study by Moles & Giribet (2021) using ultra-conserved elements from major groups of marine heterobranchs recovered the monophyly of all orders tested as well as the larger clades Nudipleura, Panpulmonata, and Euopisthobranchia which had been proposed based on a couple of molecular markers during the last 15 years (e.g., Jörger et al 2010).

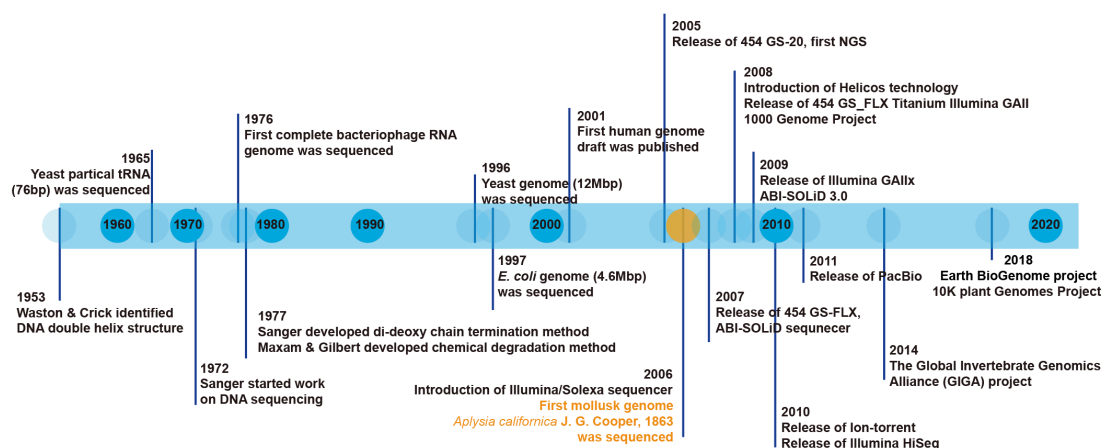
Many reasons may have caused the complexity and inconsistency of mollusk phylogeny. For example, the absence of fossils of taxa without shells (mollusks fossils are usually well preserved because of their hard shell, but those without a shell, such as slugs and octopus, are rarely found as fossils) (Parkhaev 2008; Parkhaev 2017); poor taxon sampling for crucial clades; the lack of knowledge on potential transitory characters that are restricted to given ontogenetic stages led to incomplete data matrices for the morphology-based phylogenetic analyses (Wanninger & Wollesen 2019) which were further complicated by long evolutionary history, multiple character losses and uncertain homologies (Schrödl & Stöger 2014). Rapid evolution, adaptive radiation and convergent evolution provide difficulty in revealing relationships among species (Morton 1983; Serb & Eernisse 2008; Stöger et al. 2013; Van Bocxlaer et al. 2020; Vinther et al. 2012) and tree reconstructions are hampered by incomplete lineage sorting and long internodes with sudden bursts of different species. In the early days of molecular phylogenetics, not only the quantity of markers was poor, but also the quality of sequences, alignments, model assumptions and phylogenetic analyses may have been inadequate (e.g., Stöger & Schrödl 2013). In addition, molecular markers themselves have limitations. For example, the commonly used mitochondrial markers, might be influenced by introgression and sex-biased reproductive dispersal (Kern et al. 2020). Furthermore, the biases of accelerated substitution rates and compositional heterogeneity of mitochondrial sequence also affects phylogenetic analyses (Masta et al. 2009). Contrary to difficulties in the resolution of deeper, older evolutionary relationships, mitochondrial genes and some genomic data already have been useful in resolving more recent, intrafamilial phylogenies (Ghiselli et al. 2021a). Evolving sequencing techniques, analytical power of computers and bioinformatics paved the way to access the full information of whole genomes.

## **2.3 Towards molluscan genomics**

### **2.3.1 Development of genome sequencing strategy**

Since the first DNA molecule (12 bases long complementary extremities of phage  $\lambda$  cos-site) was sequenced in 1968 (Wu & Kaiser 1968), we have witnessed a dramatical development of sequencing technology in the past 50 years (**Fig 1**). From Sanger sequencing (1 kbp, 1% error rate) (Sanger et al. 1977), to next generation sequencing (NGS, 110 bp-350 bp, 0.087-0.613 error rate) (Stoler & Nekrutenko 2021), to Third-generation sequencing represented by Oxford Nanopore Technologies (Nanopore, 100 kb-1 M, 1% error rate for new R10 flow cell) (Sereika et al. 2022) and Pacific Biosciences (Single Molecule RealTime sequencing, 20 kb, 1% error rate for High Fidelity (HiFi) reads) (Wenger et al. 2019). The greatly improved throughput,

sequencing speed, accuracy, and the dramatically decreased sequencing price would permit sequencing a vast number of species (Giani et al. 2020; Shendure et al. 2017). At the same time, advances in big data processing and assembly algorithms and techniques allowed to correctly reconstruct the original chromosome sequences from sequencing reads (Guiglielmoni et al. 2022). For larger genomes, the Bionano optical mapping (Schwartz et al. 1993) and Chromosome Conformation Capture (3C) based technique (Hi-C) allows to orient and order the contigs, thereby significantly increasing the continuity of assembly and even raising the scaffold to the chromosome level (Lieberman-Aiden et al. 2009).



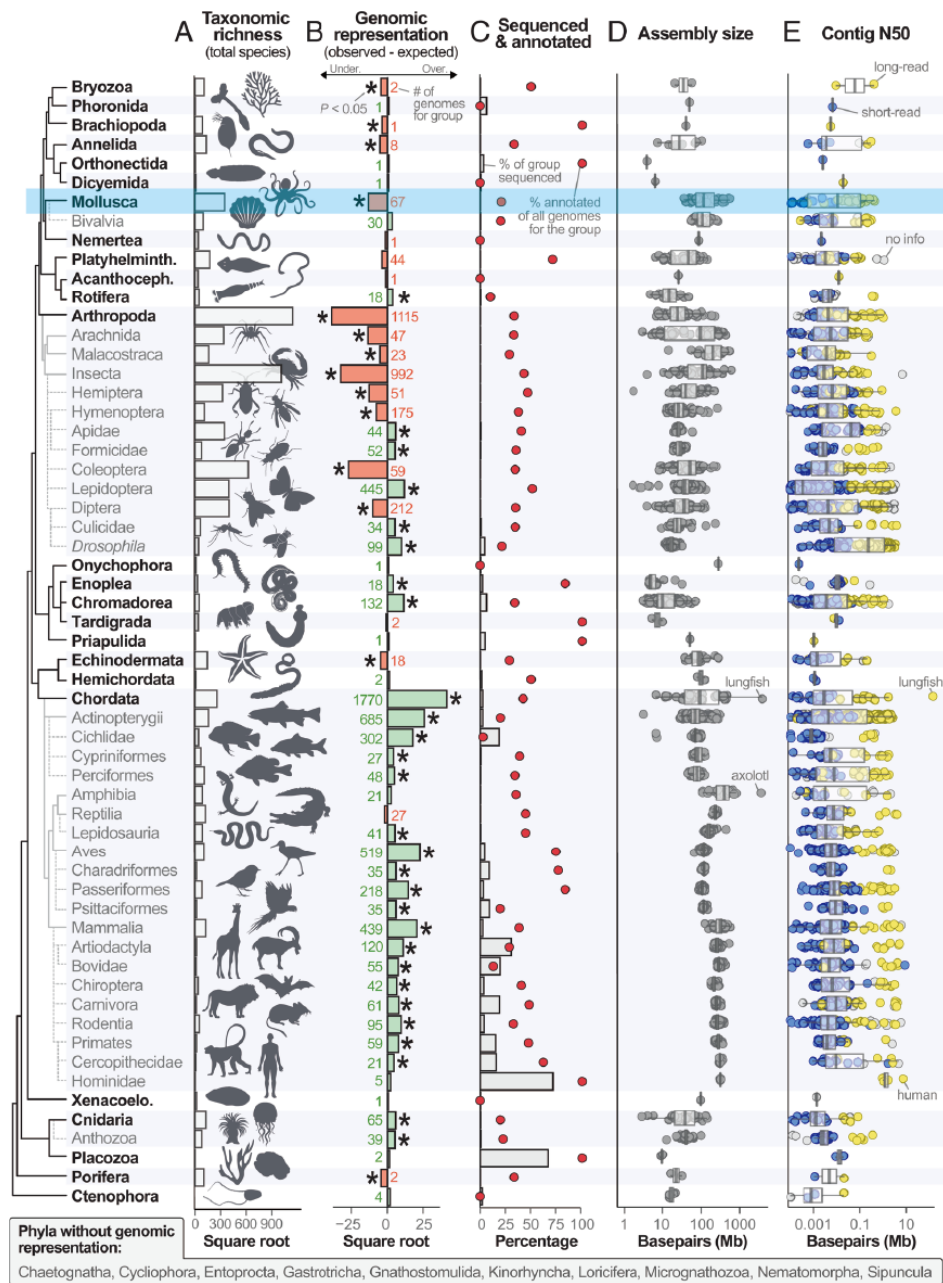
**Fig 1. Timeline of introduction of DNA sequencing technologies, platforms and Milestones in genome assembly** (See Giani et al. 2019; supplemented). The year of the first reported mollusk genome was highlighted in orange.

### 2.3.2 Underrepresented genetic resources of mollusks

The advent of high-throughput sequencing techniques has extended mitogenome characterization to non-model, understudied mollusks, with the number of sequenced mitogenomes of mollusks increasing rapidly in the last few years. As of 20 August 2022, there were 1,296 molluscan mitochondrial genomes deposited in GenBank (the same species was counted only once, and only species with assembly lengths greater than 10,000 bp were counted). However, the available database of mitogenomes is still largely underrepresented and taxon-biased. For example, the proportion of sequenced mitogenomes to the number of species per lineage for Mammals was 21%, Osteichthyes was 9%, Aves was 7% (as before April 2020) (Zardoya 2020), but as of 20 August 2022, for Mollusca was only 1.2%, whereas the smallest was in gastropods (0.7%) and the largest was in bivalves (4.5%).

The field of animal genomics is thriving with the advancement of sequencing technologies, with around 4.07 genomes deposited in GenBank per day in a recent estimate (Hotaling et al. 2021). However, efforts have exceptionally focused on vertebrates: Out of 10,509 metazoan assemblies available in the GenBank database

(accessed on August 20, 2022), 54.8% (5,755) belong to Vertebrata. Although vertebrates represent only 3.5% of the diversity of described metazoan species (IUCN 2021). Conversely, mollusks were underrepresented with 134 assemblies (1.3% of the dataset) for a group that comprises 5.4% of animal species (IUCN 2021). Therefore, the overall resources for comparative molluscan genomics are still lacking (Fig 2).



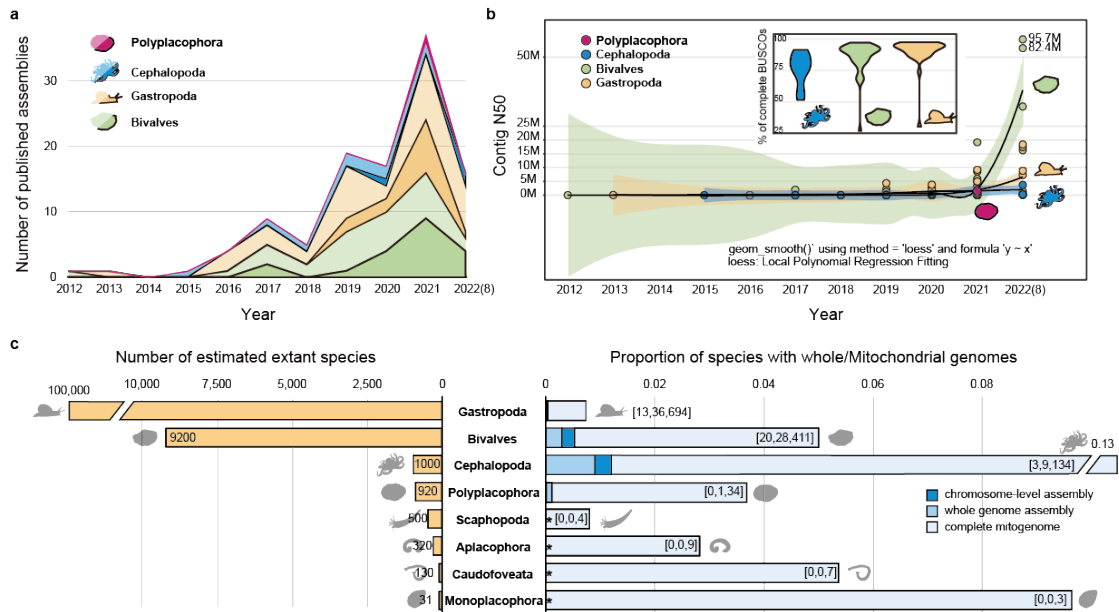
**Fig 2. Variation in taxonomic richness and genome availability, quality, and assembly size across kingdom Animalia in GenBank (as of 28 June 2021)**  
 Figure was modified from Hotaling et al (2021). Mollusca are highlighted in blue.  
 Permission has been obtained from the corresponding author.

### 2.3.3 Unbalanced genetic resources among mollusks

Despite the maturation of sequencing technologies since a decade ago, the genomic resources available for mollusks have been slow to arrive, and only in the last few years has the number increased rapidly (**Fig 3a**). As of August 20, 2022, there are a total of 110 species whose genomes are available in GenBank (for species with multiple assembled versions, we only count once), of which 32.7% (36) have reached the chromosome level (**Fig 3d**). At the same time, the continuity of the assemblies (using contig N50 (Alhakami et al. 2017) as an indicator) has been greatly improved, with two newly reported bivalve genomes *Ostrea edulis* Linnaeus, 1758 contig N50 reaching 95.7M and *Mercenaria mercenaria* (Linnaeus, 1758) reaching 82.4M (**Fig 3b**) (Gundappa et al. 2022; Varney et al. 2021b). Except for cephalopods and several outliers, genome integrity (using % of complete BUSCOs as an indicator) also exceeded 80% (Simao et al. 2015).

However, genomic resources still are highly uneven among the various classes within the phylum Mollusca (**Fig 3c**). Gastropoda has the largest number of genome assemblies (49), yet it has a large deficit relative to its species number. Bivalves (48) and Cephalopoda (11) occupy a higher percentage of assemblies relative to the number of extant species, however, they only account for 0.005% and 0.012% of the total number of currently known species respectively. Currently (as of September 2022) Polyplacophora has only one available genome assembly *Acanthopleura granulata* (Gmelin, 1791) (Varney et al. 2021b), and four classes (Scaphopoda, Solenogastres, Caudofoveata, and Monoplacophora) have no publicly available sequenced and assembled genomes.

Similarly, the resources of the mitochondrial genome are also unbalanced in mollusks. The species numbers with an assembled mitogenome size >10 kbp were counted from NCBI, and Gastropoda were largely underrepresented, as well as Scaphopoda in comparison with other orders (0.7%, 0.8% of extant species respectively, **Fig 3c**). Cephalopoda and Monoplacophora have the highest number of mitochondrial genomes as a proportion of species number (13%, 10% of extant species respectively, **Fig 3c**).



**Fig 3. Variation in taxonomic richness and genome/mitogenome availability, continuity, and integrity among mollusks**

(Data from GenBank, as of 20 August, 2022; own analysis)

Note: Light and dark colors in Fig 3a represent non-chromosome level and chromosome level assemblies, respectively. In Fig 3b, contigN50 of all assemblies (non-chromosomal/chromosomal level assemblies) of different classes were counted.

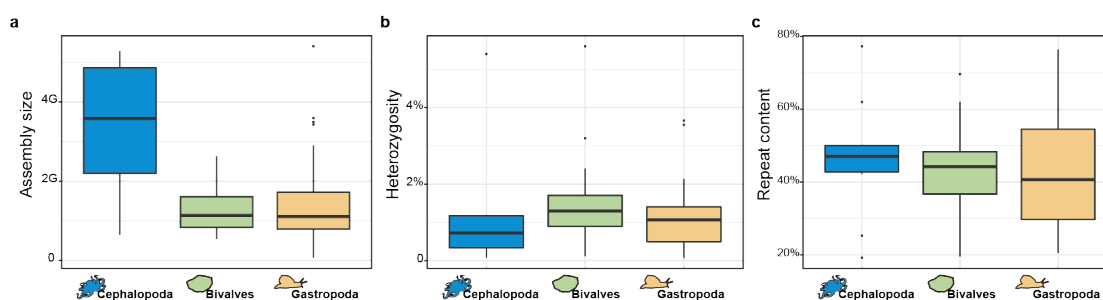
### 2.3.4 Challenges and genomic features of mollusks

Many factors may have impeded and slowed down the advancement of mollusk genomics. First, many mollusks, especially pelagic and deep-sea mollusks, are difficult to sample and are not easily preserved in optimal conditions for genomic studies. Samples that have not been adequately preserved or are taken from animals with very small body size, make it difficult to obtain enough High Molecular Weight DNA suitable for whole genome sequencing. In addition, sticky mucopolysaccharides and inhibitory polyphenolic proteins make DNA extraction exceptionally difficult (Adema 2021), therefore it was trapped at the first step. Moreover, mollusks in general have large and complex genomes (Guiglielmoni et al. 2022). A high-quality assembly with large genome size usually requires sufficient sequencing data (hence requirements for even higher quality and quantity of DNA), exponentially increased financial investment and computational power. Coupled with the fact that the genome is usually highly heterozygous and highly repetitive, this greatly increases the difficulty of assembling (Leffler et al. 2012).

Genomic characteristics vary widely among and within classes. Specifically, except for the only Polyplacophora genome, with an assembled genome size of 607M, 0.65% heterozygosity and 23.56% repeat content (Varney et al. 2021b), Cephalopoda have the largest assembly size with an average genome size of 3.35 Gb and the number of



haploid chromosomes range from 30-48 (**Fig 4a, Addition file**). Gastropoda have an average assembly size of 1.37 Gb and the number of 9-35 chromosomes and bivalves have an average assembly size of 1.24 Gb, with 10-19 chromosomes (**Fig 4a, Addition file**). They all have very high heterozygosity, with the vast majority of Gastropoda and Bivalvia being greater than 1%, and some species are extremely heterogeneous, e.g., *Watasenia scintillans* (Berry, 1911) (Cephalopoda, estimated heterozygosity of 4.9-5.9%), *Mya arenaria* Linnaeus, 1758 (Bivalves, estimated heterozygosity of 4.9-5.9%), *Elysia chlorotica* Gould, 1870 (Gastropoda, estimated heterozygosity of 3.66%) (**Fig 4b, Addition file**). The repeat content varies between species in all classes, from 19.2% (*W. scintillans*, Cephalopoda) to 77.3% (*Sepia pharaonis* Ehrenberg, 1831, Cephalopoda) (**Fig 4c, Addition file**).



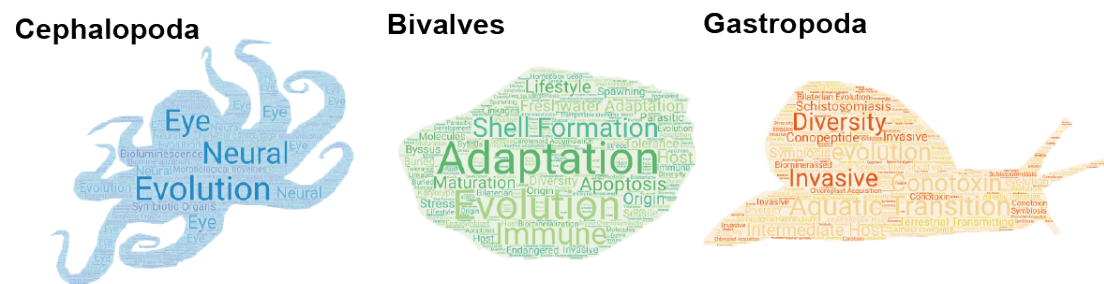
**Fig 4. Assembly size, heterozygosity, repeat content among available molluscan genomes in different classes**  
(As of 20 August, 2022; own analysis)

### 2.3.5 Concerns of currently published mollusk genomic studies

Whole genome assemblies provide a vast resource for expanding our knowledge regarding the understanding of species biodiversity, evolution, and adaptation. By means of genomic resources, studies have focused on various aspects in different classes of mollusks. For Polyplacophora, the genome was explored to explain its iron-response and biomineralization. As the first reported Polyplacophora genome, it was also used to infer molluscan phylogeny and recovered Polyplacophora as the sister taxon of all other mollusks (Varney et al. 2021b). For Cephalopoda, comparative genomics were used to investigate the development of brain and nervous system (Albertin et al. 2022; Albertin et al. 2015), symbiotic organs (Belcaid et al. 2019), eye evolution (Zhang et al. 2021), Bioluminescence (Yoshida et al. 2020) and so on (**Fig 5**). For bivalves, efforts were greatly focused on economically important species that contribute significantly to aquaculture or on invasive species and pests that have a huge destructive effect (Gundappa et al. 2022; Li et al. 2020; Mun et al. 2017; Penaloza et al. 2021; Uliano-Silva et al. 2018; Yang et al. 2021). In addition to that, there were also studies focusing on host-endosymbiont genome integration (Ip et al. 2021) and biomineralization (Du et al. 2017) (**Fig 5**). For Gastropoda, genomic studies were more diverse, including of aquaculturally important species, e.g., abalone (Botwright et al.

2019; Masonbrink et al. 2019; Orland et al. 2022), pests and invasive species (Ebbs et al. 2018; Sun et al. 2019), or exploring the origin of conotoxin diversity (Barghi et al. 2016; Pardos-Blas et al. 2021; Peng et al. 2021), host-parasite interactions (Nong et al. 2022; Young et al. 2022), gammaproteobacterial endosymbionts (Lan et al. 2021), photosynthesis (Maeda et al. 2021) and territory adaption (Liu et al. 2020). From the level of molecular evolution, scientists have also focused on the expansion of transposons in the molluscan genome, DNA methylation, whole-genome duplication, evolution and recombination of hox genes and so on (Chen et al. 2022; Thomas-Bulle et al. 2018).

As one of the main clades of gastropods, Heterobranchia includes marine, aquatic and terrestrial species, with very diverse forms and lifestyles (Haszprunar 1985; Wägele et al. 2008). Many interesting questions, such as why and how several Heterobranchia clades accomplished the aquatic-land transition and adapted to such a diverse environment, have not yet been explained on a molecular level. Most of the published genomic studies of Heterobranchia nowadays stay in the aspect of data reporting, and for some taxa, it is still a big challenge to getting good sequencing data and assembling into high quality genomes (**Additional file**).



**Fig 5. Concerns of currently published mollusk genomic studies**  
(As of 20 August, 2022; own analysis)

## 2.4 Aims of this thesis

In this thesis, we started to address this huge gap of knowledge, sequencing, assembling, and annotating the whole genome of the “Spanish” slug *Arion vulgaris* Moquin-Tandon, 1855. This is a notorious pest land slug in Europe. It causes serious damage in agriculture, transmits plant pathogens, outcompetes native species, and thus has great negative impact on the economy, ecology, health, and social system. Recently, *A. vulgaris* has been widely noted for its invasiveness, tolerance and adaptability to the urban environment. As a member of stylommatophoran pulmonates (land slugs and snails), which have completed the water-to-land transition from an aquatic ancestor and now flourish on land, we were very curious about the evolutionary mechanism and driving force behind it. We expected that the *A. vulgaris* genome, as the first published land slug genome, provides a valuable perspective to answer these questions on genetic

level. Meanwhile, we used the new mitogenomic and whole genomic data in an undersampled taxon to reconstruct the so far controversial phylogeny of Stylommatophora and also Gastropoda. Moreover, with full genomic data available, we wanted to test the influence of markers and their numbers selected for phylogenetic analyses on the inferred trees. In comparison with the recently published genomes of land snails, the genome of the shell-less *A. vulgaris* will serve as a good molecular resource for studying mollusk shell formation, loss, and biomineralization. Furthermore, using up-to-date hybrid sequencing techniques and bioinformatics, we expected to provide this new genome in such a high quality that it can serve as a reliable resource and reference genome for other gastropods and mollusks that are far less well studied than expected, in order to facilitate future comparative genomics studies on mollusk evolution and diversification.

In the first step, we assembled and annotated the complete mitochondrial genome of *A. vulgaris* (**Publication I**). We analyzed its mitochondrial genome composition (tRNA and rRNA genes, protein coding genes, non-coding and overlapping regions) and characteristics (codon usage), and compared the selective pressures and gene rearrangements in stylommatophoran mitogenomes. Using the mitogenomic data from other 34 published stylommatophorans and six outgroup species, we also inferred the stylommatophoran phylogenetic tree and estimated their divergence time.

Second, using single-copy orthologous genes (SOGs) identified from *A. vulgaris* and other 13 gastropods species with whole genomes available which cover five main gastropod subclasses, we explored the influence of 1) numbers of SOGs, and 2) the represented taxa on the inference of deep Gastropoda relationships, and consequently obtained robust gastropod phylogenetic hypotheses (**Publication II**).

Finally, we assembled and annotated a chromosome-level genome of *A. vulgaris* with up to date sequencing techniques and bioinformatics. Using this reference genome, we discovered genomic signatures for the water-land transition of stylommatophoran ancestors, the specific terrestrial adaptation of shell-less *A. vulgaris* and for its potential invasiveness and we discussed possible evolutionary driving forces i.e., whole-genome duplication, small scale gene duplication, positive selection, and the expansion of transposable elements (**Publication III**). We hope that this work will also promote future population genetic and phylogeographic studies on *A. vulgaris*, to further reveal its origin, population expansion, or invasiveness.

### 3. Results

#### 3.1 Publication I

Özgül Doğan, Michael Schrödl, **Zeyuan Chen** (2019): The complete mitogenome of *Arion vulgaris* Moquin-Tandon, 1855 (Gastropoda: Stylommatophora): mitochondrial genome architecture, evolution and phylogenetic considerations within Stylommatophora. PeerJ 8:e8603

A pdf of the article is available at:

<http://doi.org/10.7717/peerj.8603>

This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, reproduction and adaptation in any medium and for any purpose provided that it is properly attributed. For attribution, the original author(s), title, publication source (PeerJ) and either DOI or URL of the article must be cited.

# The complete mitogenome of *Arion vulgaris* Moquin-Tandon, 1855 (Gastropoda: Stylommatophora): mitochondrial genome architecture, evolution and phylogenetic considerations within Stylommatophora

Özgül Doğan<sup>1,2</sup>, Michael Schrödl<sup>2,3,4</sup> and Zeyuan Chen<sup>2,3</sup>

<sup>1</sup> Department of Molecular Biology and Genetics, Faculty of Science, Sivas Cumhuriyet University, Sivas, Turkey

<sup>2</sup> SNSB-Bavarian State Collection of Zoology, Munich, Germany

<sup>3</sup> Department Biology II, Ludwig-Maximilians-Universität, Munich, Germany

<sup>4</sup> GeoBio-Center LMU, Munich, Germany

## ABSTRACT

Stylommatophora is one of the most speciose orders of Gastropoda, including terrestrial snails and slugs, some of which are economically important as human food, agricultural pests, vectors of parasites or due to invasiveness. Despite their great diversity and relevance, the internal phylogeny of Stylommatophora has been debated. To date, only 34 stylommatophoran mitogenomes were sequenced. Here, the complete mitogenome of an invasive pest slug, *Arion vulgaris* Moquin-Tandon, 1855 (Stylommatophora: Arionoidea), was sequenced using next generation sequencing, analysed and compared with other stylommatophorans. The mitogenome of *A. vulgaris* measures 14,547 bp and contains 13 protein-coding, two rRNA, 22 tRNA genes, and one control region, with an A + T content of 70.20%. All protein coding genes (PCGs) are initiated with ATN codons except for *COX1*, *ND5* and *ATP8* and all are ended with TAR or T-stop codons. All tRNAs were folded into a clover-leaf secondary structure except for *trnC* and *trnS1* (AGN). Phylogenetic analyses confirmed the position of *A. vulgaris* within the superfamily Arionoidea, recovered a sister group relationship between Arionoidea and Orthalicoidea, and supported monophyly of all currently recognized superfamilies within Stylommatophora except for the superfamily Helicoidea. Initial diversification time of the Stylommatophora was estimated as 138.55 million years ago corresponding to Early Cretaceous. The divergence time of *A. vulgaris* and *Arion rufus* (Linnaeus, 1758) was estimated as 15.24 million years ago corresponding to one of Earth's most recent, global warming events, the Mid-Miocene Climatic Optimum. Furthermore, selection analyses were performed to investigate the role of different selective forces shaping stylommatophoran mitogenomes. Although purifying selection is the predominant selective force shaping stylommatophoran mitogenomes, six genes (*ATP8*, *COX1*, *COX3*, *ND3*, *ND4* and *ND6*) detected by the branch-specific aBSREL approach and three genes (*ATP8*, *CYTB* and *ND4L*) detected by codon-based BEB, FUBAR and MEME approaches were exposed to diversifying selection. The positively selected substitutions at the mitochondrial PCGs of stylommatophoran

Submitted 23 October 2019

Accepted 19 January 2020

Published 21 February 2020

Corresponding author

Özgül Doğan,  
odogan@cumhuriyet.edu.tr

Academic editor

Walter de Azevedo Jr.

Additional Information and  
Declarations can be found on  
page 21

DOI 10.7717/peerj.8603

© Copyright  
2020 Doğan et al.

Distributed under  
Creative Commons CC-BY 4.0

**OPEN ACCESS**

**How to cite this article** Doğan Ö, Schrödl M, Chen Z. 2020. The complete mitogenome of *Arion vulgaris* Moquin-Tandon, 1855 (Gastropoda: Stylommatophora): mitochondrial genome architecture, evolution and phylogenetic considerations within Stylommatophora. *PeerJ* 8:e8603 <http://doi.org/10.7717/peerj.8603>

species seems to be adaptive to environmental conditions and affecting mitochondrial ATP production or protection from reactive oxygen species effects. Comparative analysis of stylommatophoran mitogenome rearrangements using MLGO revealed conservatism in Stylommatophora; exceptions refer to potential apomorphies for several clades including rearranged orders of *trnW-trnY* and of *trnE-trnQ-rnrS-trnM-trnL2-ATP8-trnN-ATP6-trnR* clusters for the genus *Arion*. Generally, tRNA genes tend to be rearranged and tandem duplication random loss, transitions and inversions are the most basic mechanisms shaping stylommatophoran mitogenomes.

**Subjects** Bioinformatics, Evolutionary Studies, Genomics, Zoology

**Keywords** Mollusca, Pulmonate phylogeny, Garden slug, Gene rearrangement, Next generation sequencing, Positive selection

## INTRODUCTION

Gastropoda is the most speciose class of Mollusca, including snails and slugs with very diverse feeding habits and a wide range of habitats (Barker, 2009). The about 63,000 gastropod species represent 476 families (Bouchet et al., 2017) and radiated in marine, freshwater and terrestrial ecosystems with detritivorous, herbivorous, carnivorous, predatory or parasitic life styles (Ponder & Lindberg, 1997). Most of the terrestrial gastropods are stylommatophoran pulmonates, with approximately 30,000 species distributed from polar to tropical regions (Mordan & Wade, 2008). Stylommatophorans are economically important as human food and because of their status of being major agricultural pests and/or vectors of parasites and invasiveness (Barker, 2009). The origin of Stylommatophora is within panpulmonate heterobranchs (Jörger et al., 2010) and the monophyly of the order is undisputed. Internal phylogenetic relationships of stylommatophorans were poorly resolved based on morphology but then investigated molecularly in different sampling sets of taxa with various methods and basically relatively short sequences. Tillier, Masselmot & Tillirt (1996) used the D2 region of 28S rRNA to explore the phylogenetic relationships of pulmonates including a few stylommatophoran species; however, they reported that these short sequences would not have sufficient resolving power for investigating the relationships owing to the probable rapid radiation of pulmonate species. Wade, Mordan & Clarke (2001) and Wade, Mordan & Naggs (2006) presented more comprehensive molecular phylogenies based on relatively longer sequence information of the rRNA gene-cluster using 104 species (Wade, Mordan & Clarke, 2001) and 160 species (Wade, Mordan & Naggs, 2006) from Stylommatophora. Although these phylogenetic reconstructions accurately supported the monophyly of achatinoid and non-achatinoide clades, some clades of families that traditionally have been assumed to be monophyletic and some of the morphological groups based on excretory system; in particular, monophyly of some families and morphological groups were not supported.

The emergence and divergence time of Stylommatophora is also doubtful due to the fragmentary fossil records. The earliest land snails identified as stylommatophoran species are from upper Carboniferous and Permian but their classification has still

been controversial (Solem & Yochelson, 1979; Hausdorf, 2000). Bandel (1991) and Roth et al. (1996) suggested the oldest known fossil records from late Jurassic and Early Cretaceous (*Cheruscicola*) and Early Cretaceous (Pupilloidea). Tillier, Masselmot & Tillirt (1996) inferred that the Stylommatophora emerged in the transition between late Cretaceous and Paleocene (65–55 Ma) congruent with fossil records, based on the molecular data. However, all of the previous molecular dating analyses on Stylommatophora have been performed either with limited numbers of taxa or molecular markers (Tillier, Masselmot & Tillirt, 1996; Jörger et al., 2010; Dinapoli & Klussmann-Kolb, 2010; Zapata et al., 2014), therefore there is a need for further investigations in a more comprehensive sampling using more markers for better understanding of the phylogeny and timing of evolution of Stylommatophora.

In recent years, there has been a rapid increase in the number of sequenced mitochondrial genomes (mitogenomes) in parallel to revolution on high throughput DNA sequencing technology and data mining, providing a powerful tool for phylogenetic analysis (Moritz, Dowling & Brown, 1987; Boore, 1999; Bernt et al., 2013a). Animal mitogenomes are double-stranded circular molecules which are ~16 kb in length and contain 13 protein coding genes (PCGs) forming the respiratory chain complexes: Complex I or NADH: ubiquinone oxidoreductase contains seven subunits of NADH dehydrogenase (*ND1–6* and *ND4L*), complex III or ubiquinol: cytochrome *c* oxidoreductase consists of cytochrome *b* (*CYTB*), complex IV or cytochrome *c* oxidase comprises three subunits of cytochrome *c* oxidase (*COX1–COX3*) and complex V or ATP synthase includes two subunits of the ATPase (*ATP6* and *ATP8*). The mitochondrial PCGs have generally been supposed to be evolving under neutral or nearly neutral selection (Ballard & Kreitman, 1995). Although it has been suggested that these genes are likely to be under strong purifying selection considering their functional importance, the selective pressures might vary even among closely related species and be influenced by environmental conditions (Meiklejohn, Montooth & Rand, 2007). The mitogenomes also encode the small and large subunit rRNAs (*rrnL* and *rrnS*) and twenty-two tRNA genes for the translation process of PCGs. In general, they harbour a single large non-coding region containing control elements necessary for replication and transcription (Boore, 1999). Mitogenomes have become widely used tools in recent phylogeny, phylogeography and molecular dating analyses in various taxa, because of their (1) relatively small size, (2) the high copy number, (3) maternal inheritance type and (4) relatively rapid rate of evolutionary change (Moritz, Dowling & Brown, 1987; Gray, 1989). The sequence information of mitogenomes has also been used in reconstructing phylogenies of several taxonomic groups within/including Gastropoda (White et al., 2011; Stöger & Schrödl, 2013; Sevigny et al., 2015; Uribe et al., 2016a; Uribe et al., 2016b; Romero, Weigand & Pfenninger, 2016; Yang et al., 2019). Although there have been some criticisms about the usage of mitogenomes in construction of gastropod phylogeny because of long branch attraction, substitution saturation and strand-specific skew bias (Stöger & Schrödl, 2013), within the recently diversified lineages of gastropods, the use of mitogenomes resulted in highly resolved phylogenies (Williams, Foster & Littlewood, 2014; Osca, Templado & Zardoya, 2014). Besides the use of the mitogenome in sequence-based phylogenies, mitogenome rearrangements can also provide phylogenetic signals

(Grande, Templado & Zardoya, 2008; Stöger & Schrödl, 2013; Xie et al., 2019b). Although the mitogenome is widely used in phylogeny of many gastropod groups (Arquez, Colgan & Castro, 2014; Osca, Templado & Zardoya, 2014; Sevigny et al., 2015; Uribe, Zardoya & Puillandre, 2018), there are limited numbers of reported stylommatophoran mitogenomes and phylogenetic studies in Stylommatophora in terms of usage of mitogenome sequence and rearrangement (Romero, Weigand & Pfenninger, 2016; Xie et al., 2019a; Yang et al., 2019). To date, complete or nearly complete mitogenomes have been reported for only 34 stylommatophoran species (NCBI, September, 2019).

In this study, we sequenced and annotated the complete mitogenome of *Arion vulgaris* Moquin-Tandon, 1855 (Gastropoda: Stylommatophora), which is considered as a serious invasive pest both in agriculture and private gardens. We compared it with the mitogenome of its congener *Arion rufus* (Linnaeus, 1758), and with all other previously reported stylommatophoran mitogenomes. We also reconstructed a phylogeny from stylommatophoran mitogenomes to estimate the phylogenetic position of *A. vulgaris* and to test the informativeness of mitogenome data in the reconstruction of Stylommatophora phylogeny. In addition, we obtained a dated phylogeny using this mitogenome dataset and fossil calibrations to estimate divergence times within Stylommatophora. Furthermore, selection analyses were performed to investigate the role of different selective forces shaping stylommatophoran mitogenomes. Finally, we compared the mitogenome organisations of stylommatophoran species using a comparative and phylogeny based method and tried to uncover the evolutionary pathways of mitogenome rearrangements.

## MATERIALS AND METHODS

### Specimen collection and DNA extraction

The specimen of *A. vulgaris* was collected from the garden of the Zoologische Staatssammlung München (ZSM), Germany. Total genomic DNA was extracted from mantle tissue using CTAB method (Doyle & Doyle, 1987).

### Mitogenome sequencing, annotation and analyses

The whole-genome sequencing was conducted with 150 bp pair-end reads on the Illumina HiSeq4000 Platform (Illumina, San Diego, CA) using 350 bp insert size libraries. Raw reads were processed by removing low quality reads, adapter sequences and possible contaminated reads using Fastp v0.20.0 (Chen et al., 2018) and Lighter v1.0.7 (Song, Florea & Langmead, 2014). In total, about 7.5G high quality base pairs of sequence data were obtained and the mitogenome was assembled using the MitoZ software (Meng et al., 2019), followed by manual curation using Geneious R9 (Kearse et al., 2012).

The annotation of tRNA genes of the *A. vulgaris* mitogenome was performed using MITOS (<http://mitos.bioinf.uni-leipzig.de/index.py>) (Bernt et al., 2013b) and ARWEN web servers (Laslett & Canbäck, 2008) based on their secondary structures and anticodon sequences. The locations and boundaries of PCGs and rRNA genes were identified manually by comparing with the *A. rufus* (KT626607) homologous gene sequences. The visualization of the secondary structure of tRNA genes was performed using VARNA v3-93 (Darty, Denise & Ponty, 2009) and RNAviz 2.0.3 (De Rijk, Wuyts & De Wachter, 2003). Intergenic



spacers and overlapping regions between genes were estimated manually. The largest non-coding region was defined as control region and the Mfold server (Zuker, 2003) was used to predict the secondary structure of this region. The “palindrome” tool within the European Molecular Biology Open Software Suite (EMBOSS) (Rice, Longden & Bleasby, 2000) was used for searching the palindromic sequences in the control region. Finally, the complete mitogenome of *A. vulgaris* was deposited in GenBank under accession number MN607980. The mitogenome of *A. vulgaris* is visualized using OrganellarGenomeDRAW (OGDRAW) (Greiner, Lehwark & Bock, 2019).

The nucleotide compositions, average nucleotide and amino acid sequence divergences and the relative synonymous codon usages (RSCU) of PCGs were computed using MEGA v7.0 (Kumar, Stecher & Tamura, 2016). The strand asymmetries were calculated according to the following formulas: AT-skew =  $[A - T]/[A + T]$  and GC-skew =  $[G - C]/[G + C]$  (Perna & Kocher, 1995).

## Phylogenetic and comparative analyses

### Alignment and model selection

Phylogenetic and comparative analyses were performed using the mitogenome dataset of 35 stylommatophoran species representing 18 families, and using one species from Systellommatophora, one species from Hygrophila, and one species from Ellobioidea as outgroups (Table 1). Each tRNA and rRNA gene was aligned individually using MAFFT (Kato & Standley, 2013) algorithm in Geneious R9 (Kearse et al., 2012). The alignment of nucleotide sequences of each PCG was performed using MAFFT algorithm and the “translation align” option implemented in Geneious R9. The final alignment files were then concatenated using SequenceMatrix v.1.7.8 (Vaidya, Lohman & Meier, 2011). The optimal partitioning scheme and substitution models were inferred by PartitionFinder v1.1.1 (Lanfear et al., 2012) using the Bayesian Information Criterion (BIC) and the “greedy” algorithm with the option of “unlinked” branch lengths. The best-fit partitioning scheme and nucleotide substitution models were used in phylogenetic analyses (Table S1).

### Assessing the substitution saturation level

The substitution saturation levels in different genes and codon positions were estimated comparing the uncorrected p-distances and the distances calculated by applying the GTR + G + I evolutionary model selected based on the BIC using jModelTest v2.1.7 (Durriba et al., 2012). All genetic distances were computed with PAUP v4.0 b10 (Swofford, 2002).

### Phylogenetic reconstruction

Two different datasets were created for phylogenetic analyses to test the influence of saturated genes and codon positions: (1) 13 PCGs including all codon positions plus the 22 tRNAs and two rRNAs (P123RNA) and (2) PCGs excluding the five saturated genes and third codon positions, plus 22 tRNAs and two rRNAs (8P12RNA, Table S2). Maximum likelihood (ML) trees were constructed with RAxML v8.0.9 (Stamatakis, 2014) implemented in Geneious R9 applying the best-fit evolutionary model for each partition under 1,000 bootstrap replicates. For Bayesian Inference (BI) analyses, MrBayes v3.2.2 (Ronquist et al., 2012) was employed with two independent runs of 10 million generations

**Table 1** List of stylommatophoran mitogenomes used in phylogenetic and comparative analyses.

	Species	Family	Accession number	References
	<i>Arion vulgaris</i>	Arionidae	MN607980	This study
	<i>Arion rufus</i>	Arionidae	KT626607	Romero, Weigand & Pfenninger (2016)
	<i>Achatinella fulgens</i>	Achatinellidae	MG925058	Price et al. (2018)
	<i>Achatinella mustelina</i>	Achatinellidae	NC030190	Price et al. (2016a)
	<i>Achatinella sowerbyana</i>	Achatinellidae	KX356680	Price et al. (2016b)
	<i>Partulina redfieldi</i>	Achatinellidae	MG925057	Price et al. (2018)
	<i>Achatina fulica</i>	Achatinidae	KM114610	He et al. (2016)
	<i>Deroceras reticulatum</i>	Agriolimacidae	NC035495	Ahn et al. (2017)
	<i>Aegista aubryana</i>	Bradybaenidae	NC029419	Yang et al. (2016)
	<i>Aegista diversifamilia</i>	Bradybaenidae	NC027584	Huang, Lin & Wu (2015)
	<i>Dolicheulota formosensis</i>	Bradybaenidae	NC027493	Huang, Lin & Wu (2015)
	<i>Mastiguelota kiangsinensis</i>	Bradybaenidae	NC024935	Deng et al. (2016)
	<i>Camaena cicatricosa</i>	Camaenidae	NC025511	Wang et al. (2014)
	<i>Camaena poyuensis</i>	Camaenidae	KT001074	Unpublished
	<i>Cerion incanum</i>	Cerionidae	NC025645	González et al. (2016)
	<i>Cerion tridentatum costellata</i>	Cerionidae	KY249249	Unpublished
STYLOMMATOPHORA	<i>Cerion uva</i>	Cerionidae	KY124261	Harasewych et al. (2017)
	<i>Albinaria caerulea</i>	Clausiliidae	NC001761	Hatzoglou, Rodakis & Lecanidou (1995)
	<i>Gastrocopta cristata</i>	Gastrocoptidae	NC026043	Unpublished
	<i>Cernuella virgata</i>	Geomitridae	NC030723	Lin et al. (2016)
	<i>Helicella itala</i>	Geomitridae	KT696546	Romero, Weigand & Pfenninger (2016)
	<i>Cepaea nemoralis</i>	Helicidae	NC001816	Yamazaki et al. (1997)
	<i>Cylindrus obtusus</i>	Helicidae	NC017872	Groenenberg et al. (2012)
	<i>Cornu aspersum</i>	Helicidae	NC021747	Gaitán-Espitia, Nespolo & Opazo (2013)
	<i>Helix pomatia</i>	Helicidae	NC041247	Korábek, Petrusek & Rovatsos (2019)
	<i>Orcula dolium</i>	Orculidae	NC034782	Groenenberg et al. (2017)
	<i>Naesiotus nux</i>	Orthalicidae	NC028553	Hunter et al. (2016)
	<i>Meghimatium bilineatum</i>	Philomycidae	NC035429	Xie et al. (2019a) and Xie et al. (2019b)
	<i>Philomycus bilineatus</i>	Philomycidae	MG722906	Yang et al. (2019)
	<i>Polygyra cereolus</i>	Polygyridae	NC032036	Unpublished
	<i>Praticolella mexicana</i>	Polygyridae	KX240084	Minton et al. (2016)
	<i>Pupilla muscorum</i>	Pupillidae	NC026044	Unpublished
	<i>Succinea putris</i>	Succineidae	NC016190	White et al. (2011)
	<i>Microceramus pontificus</i>	Urocoptidae	NC036381	Unpublished
	<i>Vertigo pusilla</i>	Vertiginidae	NC026045	Unpublished
Ellobioidea	<i>Carychium tridentatum</i>	Ellobiidae	KT696545	Romero, Weigand & Pfenninger (2016)
Hygrophila	<i>Galba pervia</i>	Lymnaeidae	NC018536	Liu et al. (2012)
Systellommatophora	<i>Platevindex mortoni</i>	Onchidiidae	GU475132	Sun et al. (2016)

with four Markov chains (three cold, one heated), sampling every 1,000 generations and a burn-in of 25% trees. The stationarity of the chains was assessed using the program Tracer v1.7 (Rambaut et al., 2018). The consensus phylogenetic trees were visualized using FigTree v1.4.0 (Rambaut, 2012).

### **Divergence time estimation**

MCMCTree program implemented in the Phylogenetic Analysis by Maximum Likelihood (PAML) package v4.9 (Yang, 2007) was used for Bayesian estimation of divergence times of each species. Substitution rate per site was estimated by BASEML and was used to set the prior for the mean substitution rate in the Bayesian analysis. MCMC was run by  $50 \times 10,000$  iterations with the REV substitution model. The soft bounds of *Helix pomatia* + *Cornu aspersum* [divergence time between 34 million years ago (Ma) and 42 Ma], *Mastigeulota kiangsinensis* + (*Dolicheulota formosensis* + (*Aegista aubryana* + *Aegista diversifamilia*)) (divergence time between 25 Ma and 51 Ma), and *Camaena cicatricosa* + *Camaena poyuensis* (divergence time between 16 Ma and 39 Ma) were used as external calibrations (Razkin et al., 2015) and the estimated nodal age of *Tectipleura* [244 Ma (210–279 Ma)] was used for the calibration of the root (Kano et al., 2016).

### **Selection analyses**

The CODEML implemented in PAML was used to estimate the ratio of nonsynonymous/synonymous substitution rate ( $\omega = dN/dS$ ) and to explore the role of different selective constraints working on each PCG under the one-ratio model (Model A: model = 0, NSsites = 0, fix\_omega = 0, omega = 1). Gaps and ambiguous sites of sequence alignments were included in the analyses. For each PCG, likelihood ratio tests (LRTs) were used to compare the null neutral model (Model B: model = 2, NSsites = 2, fix\_omega = 1, omega = 1) against alternative models of branch-specific positive selection (Model C: model = 2, NSsites = 2, fix\_omega = 0, omega = 1.5). The Bayes Empirical Bayes (BEB) algorithm in CODEML was used to detect the positively selected sites. Furthermore, the adaptive branch-site random effects likelihood (aBSREL) (Smith et al., 2015) implemented in DATAMONKEY webserver (Weaver et al., 2018) was used to search the signatures of episodic positive diversifying selection testing each branch. In addition, mixed effects model of evolution (MEME) (Murrell et al., 2012) was used to detect episodic or diversifying selection at individual sites and a fast, unconstrained Bayesian approximation for inferring selection (FUBAR) (Murrell et al., 2013) was used for providing additional support to the detection of sites evolving under positive or negative selection. Each PCG was also evaluated in terms of properties and magnitude of amino acid changes using TreeSAAP v3.2 (Woolley et al., 2003), which uses 31 properties of amino acids and categorizes the degree of substitutions to eight categories (1–8).

### **Comparison of mitogenome organizations**

Mitogenome organizations and gene rearrangements of stylommatophoran species were analysed via the CREx web server (<http://pacosy.informatik.uni-leipzig.de/crex>) (Bernt et al., 2007). The gene orders of ancestral nodes were reconstructed using the Maximum Likelihood for Gene Order Analysis (MLGO, <http://geneorder.org/>) (Hu, Lin & Tang, 2014) with the input tree obtained by phylogenetic approaches, and the orders of the protein coding, rRNA and tRNA genes were compared with the inferred ancestral mitogenomes. A distance matrix was calculated based on number of common intervals, and the output diagram visually examined to identify shared and/or derived gene rearrangements as well as mechanisms of rearrangements.

## RESULTS AND DISCUSSION

### Mitogenome characteristics and nucleotide composition

The complete mitogenome sequence of *A. vulgaris* was obtained with a length of 14,547 bp (Table 2) and its size was within the range of the those of other reported stylommatophoran mitogenomes, varying between 13,797 bp in *Camaena poyuensis* and 16,879 bp in *Partulina redfieldi* (Price et al., 2018). It includes the entire set of 37 mitochondrial genes: 13 PCGs, 22 tRNAs and two rRNAs. Twenty-four genes were located on the J strand, while the remainings were encoded by the opposite N strand (Table 2, Fig. 1).

The nucleotide composition of *A. vulgaris* mitogenome was distinctly biased towards A and T, with a 70.20% A + T content, and comparable to other reported stylommatophoran mitogenomes, varying between 59.79% A + T in *Cepea nemoralis* (Yamazaki et al., 1997) and 80.07% A + T in *Achatinella mustelina* (Price et al., 2016a) (Table 3 and Table S3). A bias towards A and T nucleotides was also observed in PCGs of the *A. vulgaris* mitogenome with a 69.34% A + T content (Table 3). The A + T content of the 3rd codon position (79.64%) was higher than those of the 2nd (64.21%) and 1st codon positions (64.18%). Similar to other reported stylommatophoran mitogenomes (Table S3), the AT- and GC-skews were found slightly negative ( $-0.0756$ ) and positive ( $0.0431$ ) in the whole mitogenome of *A. vulgaris*, respectively. A pronounced T and G skew was also observed in all PCGs ( $-0.1508$ ,  $0.0472$ ), PCGs on the majority strand ( $-0.1447$ ,  $0.0596$ ), and tRNA genes ( $-0.0010$ ,  $0.1582$ ) (Table 3). The T- and G-skewed mitogenome of *A. vulgaris* might be explained by the spontaneous deamination of cytosine during replication and transcription processes (Reyes et al., 1998). The PCGs encoded on the minority strand displayed a T- and C-skewed pattern ( $-0.1783$  AT-skew,  $-0.0065$  GC-skew), contrary to the expected high rates of Ts and Gs on the minority strand for most of the metazoans (Hassanin, Léger & Deutsch, 2005).

### Protein coding genes and codon usage

In comparison, the lengths of the PCGs of *A. vulgaris* mitogenome were within the range of those of other stylommatophoran mitochondrial PCGs. The *ND4* gene was the most variable gene in length and has a variability of 53 codons among stylommatophorans (419 codons in *Microceramus pontificus* and 472 codons in *Orcula dolium*). The most conserved gene in length was *COX1* and it exhibits variability with only 16 codons between species of Stylommatophora (501 codons in *Achatinella mustelina* and 517 codons in *Achatina fulica*). Compared with the mitogenome of *A. rufus*, the lengths of PCGs of *A. vulgaris* were distinct except for *COX1*, *COX2*, *CYTB* and *ND1* genes. The *ND6* gene was the most variable gene in length and was longer in the *A. vulgaris* mitogenome by 11 codons. Based on the amino acid identities, the most conserved PCG was *COX1* (56.45%) whereas the least conserved was *ND6* (11.92%) among the stylommatophoran mitogenomes. The most conserved PCG was *COX1* (97.45%) whereas the least conserved was *ATP8* (68.18%) based on the amino acid identities between the two *Arion* mitogenomes (Table S4).

In the *A. vulgaris* mitogenome, most of the PCGs initiated with typical ATN start codon, except for *COX1*, *ND5* and *ATP8* genes which use TTG, ACA and GTG triplets as start codons, respectively (Table 2). The TTG and GTG start codons are also accepted

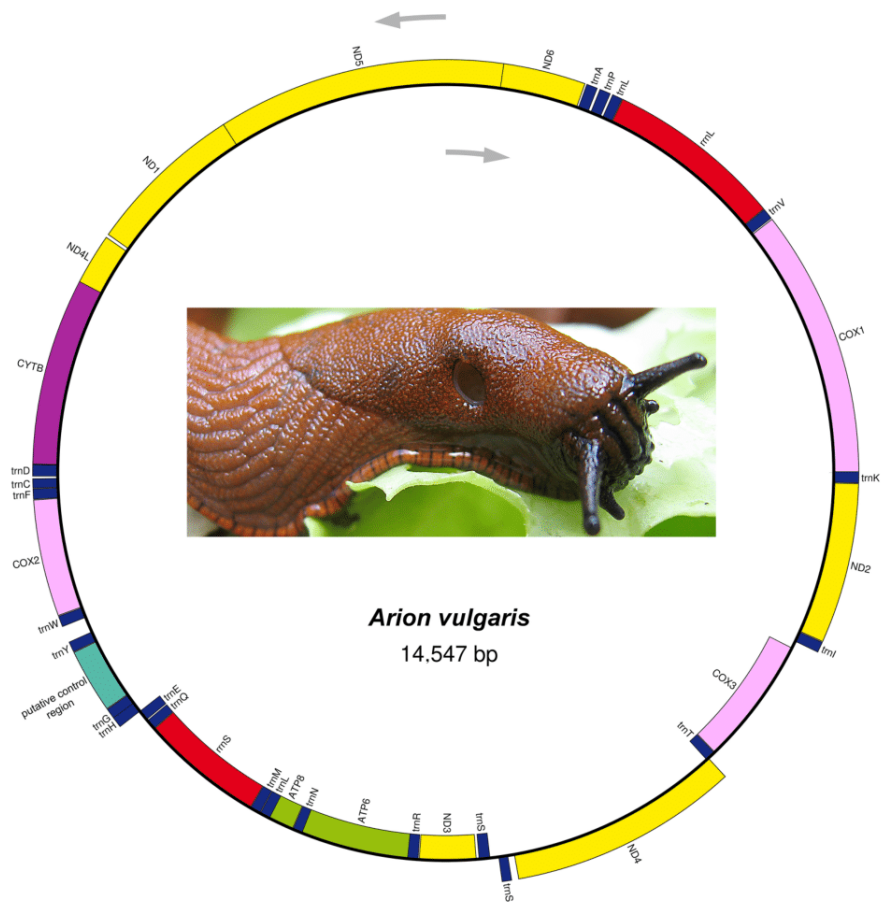
**Table 2** Mitogenome summary of *Arion vulgaris*.

Gene	Strand	From	To	Size	Start codon	Stop codon	Anticodon	IGN
<i>COX1</i>	J	1	1,530	1,530	TTG	TAG		4
<i>tRNA-Val</i>	J	1,535	1,600	66			UAC	0
<i>16S rRNA</i>	J	1,601	2,613	1,013				0
<i>tRNA-Leu</i>	J	2,614	2,675	62			UAG	11
<i>tRNA-Pro</i>	J	2,687	2,752	66			UGG	13
<i>tRNA-Ala</i>	J	2,766	2,831	66			UGC	7
<i>ND6</i>	J	2,839	3,312	474	ATG	TAG		-41
<i>ND5</i>	J	3,272	4,960	1,689	ACA	TAA		-10
<i>ND1</i>	J	4,951	5,853	903	ATG	TAG		15
<i>ND4L</i>	J	5,869	6,163	295	ATA	T-		-15
<i>CYTB</i>	J	6,149	7,228	1,080	ATG	TAA		-2
<i>tRNA-Asp</i>	J	7,227	7,296	70			GUC	10
<i>tRNA-Cys</i>	J	7,307	7,363	57			GCA	0
<i>tRNA-Phe</i>	J	7,364	7,425	62			GAA	0
<i>COX2</i>	J	7,426	8,094	669	ATG	TAG		1
<i>tRNA-Trp</i>	J	8,096	8,160	65			UCA	91
<i>tRNA-Tyr</i>	J	8,252	8,315	67			GUA	0
<i>Control region</i>	J	8,316	8,685	370				0
<i>tRNA-Gly</i>	J	8,686	8,763	78			UCC	-20
<i>tRNA-His</i>	J	8,744	8,809	66			GUG	-3
<i>tRNA-Glu</i>	N	8,807	8,873	67			UUC	5
<i>tRNA-Gln</i>	N	8,879	8,942	64			UUG	0
<i>12S rRNA</i>	N	8,943	9,689	747				0
<i>tRNA-Met</i>	N	9,690	9,754	65			CAU	0
<i>tRNA-Leu</i>	N	9,755	9,820	66			UAA	-32
<i>ATP8</i>	N	9,789	9,971	183	GTG	TAA		0
<i>tRNA-Asn</i>	N	9,972	10,033	62			GUU	-8
<i>ATP6</i>	N	10,026	10,688	663	ATA	TAA		-9
<i>tRNA-Arg</i>	N	10,680	10,746	67			UCG	3
<i>ND3</i>	N	10,750	11,094	345	ATG	TAA		13
<i>tRNA-Ser2</i>	N	11,108	11,176	69			UGA	49
<i>tRNA-Ser1</i>	J	11,226	11,283	58			GCU	36
<i>ND4</i>	J	11,320	12,633	1,314	ATA	TAG		-18
<i>tRNA-Thr</i>	N	12,616	12,681	66			UGU	0
<i>COX3</i>	N	12,682	13,462	781	ATG	T-		41
<i>tRNA-Ile</i>	J	13,504	13,567	64			GAU	1
<i>ND2</i>	J	13,569	144,86	918	ATG	TAA		0
<i>tRNA-Lys</i>	J	14,487	6	67			UUU	-6

**Notes.**

J, major; N, minor; IGN, intergenic nucleotides.

Minus indicates overlapping sequences between adjacent genes.



**Figure 1** Mitogenome organization of *Arion vulgaris*. Genes transcribed from the J and N strands are shown outside and inside of the circle, respectively. PCGs coding complex I, complex III, complex IV and complex V components are marked with yellow, purple, pink and green, respectively. rRNA genes are coloured with red and the putative control region is coloured with cyan, while tRNA genes are coloured with dark blue and labelled by the single letter amino acid code.

Full-size DOI: 10.7717/peerj.8603/fig-1

as canonical start codons for invertebrate mitogenomes (Yang *et al.*, 2019), however, ACA as start codon for *ND5* gene was reported for the first time for stylommatophoran mitogenomes. Most of the PCGs were inferred to use TAR as termination codon, except for *ND4L* and *COX3* which have an abbreviated T-termination codon and their products are probably completed via post-transcriptional polyadenylation (Anderson *et al.*, 1981; Ojala, Montoya & Attardi, 1981).

The most frequently used amino acids by the PCGs of the mitogenomes of *A. vulgaris* and *A. rufus* were leucine (16.71% and 15.91% respectively) and serine (10.33% and 10.18% respectively), similar to PCGs of the mitogenome of other stylommatophoran species (Leu 16,60%, Ser 10.21% on average). The codons rich in A and T, such as UUA-Leu, AUU-Ile, UUU-Phe, AUA-Met, UAU-Tyr, were the most frequently used codons in all stylommatophoran mitochondrial PCGs. The codons rich in terms of G and C content,

**Table 3** Nucleotide composition of the *Arion vulgaris* mitogenome.

Feature	T%	C%	A%	G%	A + T%	AT-skew	GC-skew
Whole mitogenome	37.75	14.26	32.45	15.54	70.20	-0.076	0.043
Protein coding genes	39.90	14.61	29.44	16.05	69.34	-0.151	0.047
First codon position	33.54	14.01	30.64	21.81	64.18	-0.045	0.218
Second codon position	45.65	19.42	18.55	16.38	64.21	-0.422	-0.085
Third codon position	40.50	10.39	39.14	9.97	79.64	-0.017	-0.021
Protein coding genes-J	39.78	14.34	29.72	16.16	69.50	-0.145	0.060
First codon position-J	32.94	13.93	31.23	21.90	64.17	-0.027	0.222
Second codon position-J	45.91	19.07	18.53	16.49	64.44	-0.425	-0.073
Third codon position-J	40.50	10.01	39.41	10.08	79.90	-0.014	0.003
Protein coding genes-N	40.42	15.80	28.19	15.60	68.60	-0.178	-0.006
First codon position-N	36.24	14.37	27.98	21.41	64.22	-0.129	0.197
Second codon position-N	44.50	20.95	18.65	15.90	63.15	-0.409	-0.137
Third codon position-N	40.52	12.08	37.92	9.48	78.44	-0.033	-0.121
tRNA genes	36.40	11.48	36.33	15.80	72.72	-0.001	0.158
rRNA genes	33.24	13.58	38.18	15.00	71.42	0.069	0.050
Control region	38.92	18.65	30.81	11.62	69.73	-0.116	-0.232

CGC-CGG-Arg, CAG-Gln, UGC-Cys, CUC-Leu and UCG-Ser were rarely used in both *Arion* mitogenomes (Table S5, Fig. 2). CGN-Arg, CCS-Pro, GCS, UCG and UGC codons are seldom used or never used also in the stylommatophoran mitogenomes and reflected a significant relationship between codon usage and nucleotide content (Table S5).

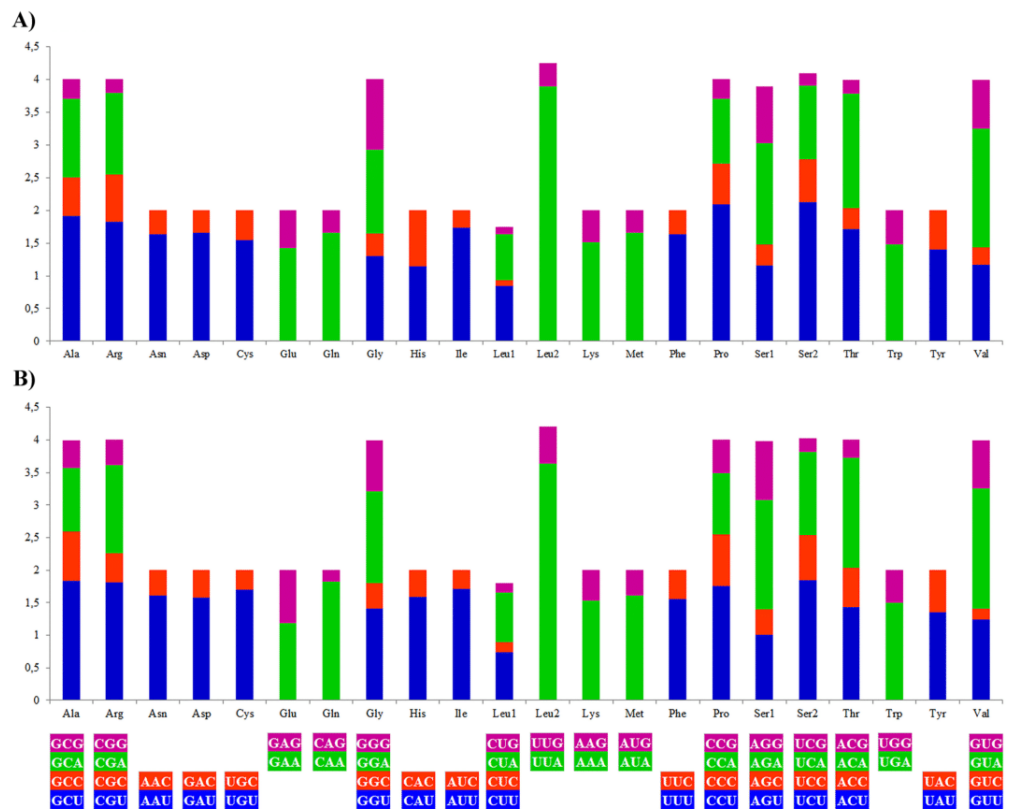
### tRNA and rRNA genes

All of the tRNA genes could be folded into a usual clover-leaf secondary structure, except for *trnS1* (AGN) and *trnC* which lacked dihydrouridine (DHU) and T $\Psi$ C arms, respectively and formed simple loops (Fig. S1). Their lengths ranged between 57 bp (*trnC*) and 78 bp (*trnG*), with an average 72.72% A + T content. 26 mismatched positions were observed in stem regions and all of the mismatches were G-U pairs (Fig. S1).

The exact boundaries of rRNA genes were determined as being bounded by the adjacent tRNA genes. The *rrnL* gene was located between *trnV* and *trnL1* genes, and the *rrnS* gene was located between *trnQ* and *trnM* genes. The length of the *rrnL* gene was 1,013 bp, with a 71.17% A + T content, while that of *rrnS* gene was 747 bp, with a 71.75% A + T content. These were comparable in ranges to homologous genes in other reported stylommatophoran species, ranging from 605 to 1215 bp in *rrnL* and from 564 to 857 bp in *rrnS*.

### Non-coding and overlapping regions

The total length of intergenic regions in the *A. vulgaris* mitogenome was 670 bp in 16 locations ranging between 1 and 370 bp (Table 2). In general, the largest non-coding region in the animal mitogenomes is considered to contain the signals for replication and transcription, and so called as the control region (Wolstenholme, 1992). The possible candidate for the control region in *A. vulgaris* mitogenome was the largest non-coding



**Figure 2** Relative synonymous codon usage (RSCU) of the (A) *A. vulgaris* and (B) *A. rufus* mitogenomes. Codon families are provided on the x axis. The stop codons are not given.

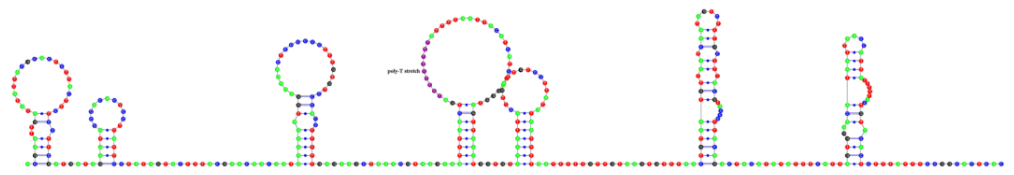
Full-size [DOI: 10.7717/peerj.8603/fig-2](https://doi.org/10.7717/peerj.8603/fig-2)

region located between *trnY* and *trnG* genes with 370 bp in length. This sequence did not give BLAST hits with other putative CRs of other molluscan mitogenomes, however a part of the sequence with 67 bp in length displayed 79.11% sequence similarity with the mitochondrial control region of an amphibian species (*Indotyphlus maharashtraensis*, [KF540157](https://doi.org/10.1016/j.mbs.2015.05.001)). Nucleotide composition of this region was slightly biased towards A + T with a 69.73% A + T content. The putative control region had a nine bp poly-T stretch and formed a stable secondary structure comprising seven stems and loops (Fig. 3). Furthermore, this sequence also contained a lot of palindromic sequences which are varying between 4 and 8 bp, but tandemly repeated sequences were not found.

The second largest non-coding region was found between *trnW* and *trnY* with a length of 91 bp (Table 2). The A + T composition of the sequence was higher than that of whole genome and putative control region with an 86.81% A + T. This non-coding region also contained a seven bp poly-A stretch and was folded into a secondary structure with two stem and loops. This secondary structure forming AT-rich sequence might function as the origin of the second strand (Wolstenholme, 1992).

Eleven overlapping regions with a total length of 164 bp were found throughout the mitogenome of *A. vulgaris*. The largest overlapping region was 41 bp in length and located





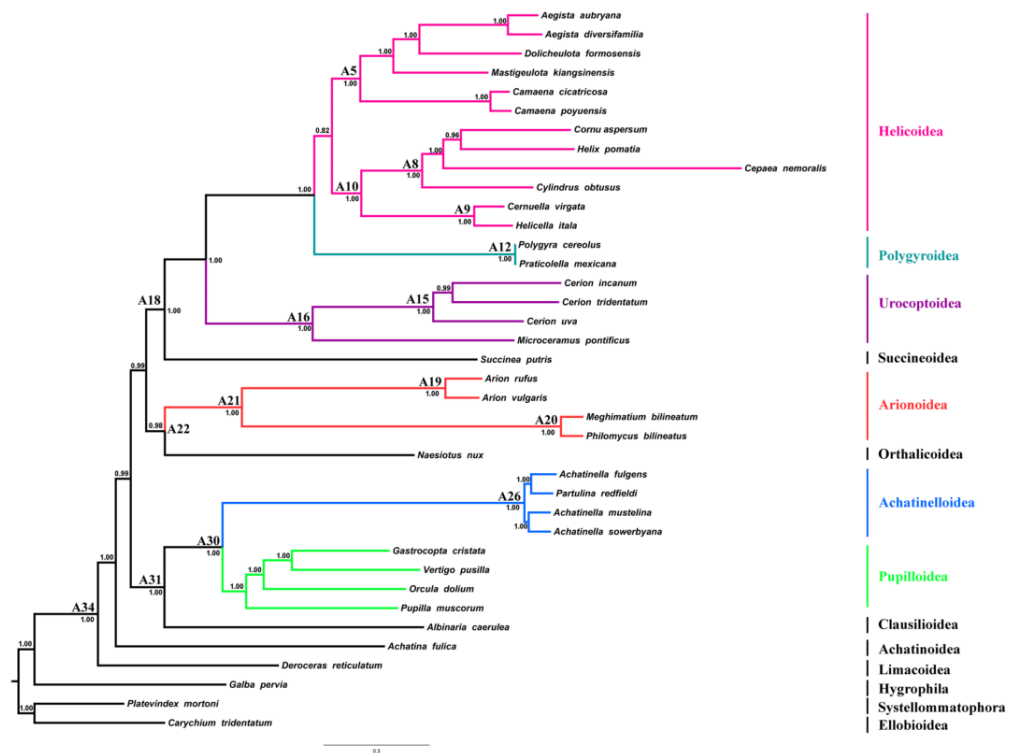
**Figure 3** Predicted secondary structure of putative control region of *A. vulgaris* mitogenome. Nucleotides are coloured as follows: Adenine is green, thymine is red, cytosine is blue and guanine is black. The poly-T stretch is labelled with purple.

Full-size DOI: 10.7717/peerj.8603/fig-3

between *ND6* and *ND5* genes, while the second largest was 32 bp and located between *trnL2* and *ATP8* (Table 2).

### Phylogeny and divergence times of stylommatophoran species

Regression analyses of pairwise distances revealed that the 1st and 2nd codon positions of *ATP8*, *ND2*, *ND3*, *ND4L* and *ND6* genes, as well as the 3rd codon positions of all PCGs were saturated (Table S2). Four phylogenetic reconstruction analyses were performed with combination of inference methods and different data matrices to test the influence of inference methods and saturation level of genes/codon positions on tree topology and nodal support. Three different tree topologies were obtained as the results of these analyses, and topologies were sensitive to both inference methods and exclusion of saturated genes/codon positions (Fig. 4 and Figs. S2–S4). Nodal support values were always higher in BI trees than ML trees of the corresponding dataset. The usage of all mitochondrial genes and codon positions (P123RNA dataset) under both approaches resulted with identical tree topology (Figs. S3 and S4), which were similar to the results of Yang *et al.* (2019) obtained using only amino acid sequences of mitochondrial PCGs. The results of these analyses supported the monophyly of all included superfamilies with high nodal supports except for the superfamily Helicoidea which recovered with low nodal support [Bayesian Posterior Probability (BPP) = 0.75, Bootstrap support (BS) = 58%] and recovered Arionoidea superfamily as sister group to Urocoptoidea + (Polygyroidea + Helicoidea) clade (BPP = 1.00, BS = 100%), and Succineoidea + Orthalicoidea clade was recovered as sister group to Arionoidea + (Urocoptoidea + (Polygyroidea + Helicoidea)) (BPP = 1.00, BS = 50%). The ML and BI analyses performed using the dataset constructed with the removal of the saturated PCGs and codon positions (8P12RNA) resulted in two different tree topologies (Fig. 4 and Fig. S2). The phylogenetic tree obtained from ML analysis did not support the monophyly of the superfamily of Helicoidea and the superfamily of Polygyroidea placed within the superfamily of Helicoidea (BS = 61%, Fig. S2). A highly resolved tree with higher nodal support values was obtained from the BI approach of the dataset 8P12RNA, and hence considered as most reliable tree for discussion. The results confirmed the taxonomic position of *A. vulgaris* as sister species to *A. rufus* and recovered the monophyly of the Arionoidea superfamily (Arionidae + Philomycidae) with high support values (BPP = 1.00). A well-supported sister group relationship between Arionoidea and Orthalicoidea was recovered (BPP = 0.98) for the first time. However, previous studies using different datasets and sampling of taxa have proposed different sister groups with Arionoidea superfamily.

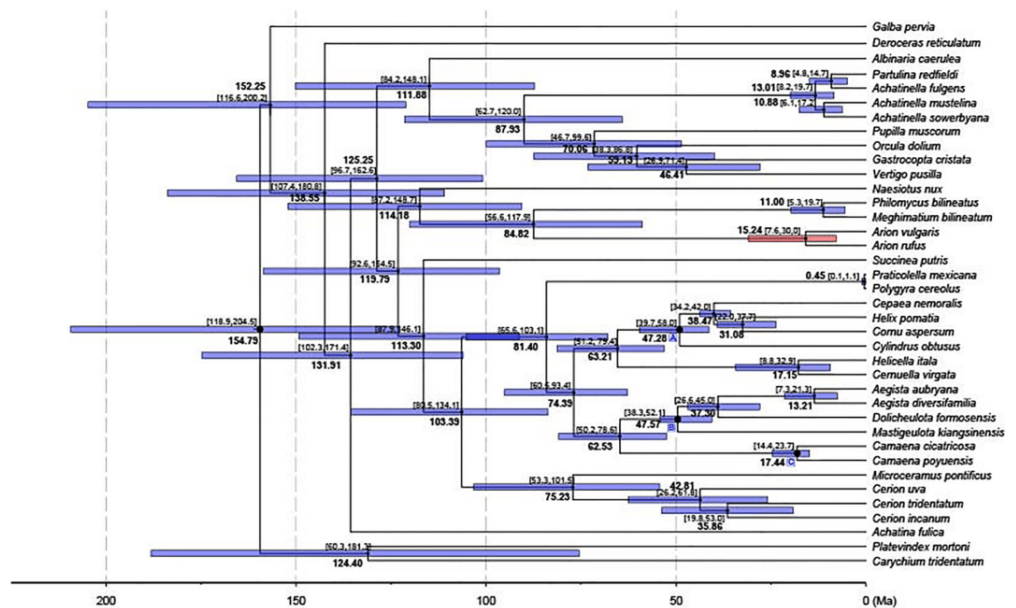


**Figure 4** Stylommatophoran phylogenetic tree constructed under BI using the dataset 8P12RNA. *Carychium tridentatum* (Ellobioidea), *Platevindex mortoni* (Systellommatophora) and *Galba pervia* (Hygrophila) were used as outgroup. Nodes are labelled with numbers refer to hypothetical ancestral mitogenome organizations inferred by MLGO.

Full-size DOI: 10.7717/peerj.8603/fig-4

Wade, Mordan & Naggs (2006) have found the superfamily Limacoidea as sister group to the superfamily Arionoidea using 160 stylommatophoran species, however they used only 823 nucleotides from rRNA gene-cluster. Holznagel, Colgan & Lydeard (2010) have proposed a sister group relationship between Arionoidea and Limacoidea + Zonitoidea based on the 28S rRNA sequences using seven species from Stylommatophora. The sister group relationships between Arionoidea and Urocoptoidea + Enoidea + Helicoidea (Jörger et al., 2010) or Limacoidea + (Succineoidea + Helicoidea) (Dayrat et al., 2011) have also been suggested by previous studies using relatively longer DNA sequences, however in both studies, only five stylommatophoran species were included for phylogenetic analyses. Furthermore, Xie et al. (2019b) have proposed sister group relationship between Arionoidea and Succineoidea using only amino acid dataset of mitochondrial PCGs, and stated it might be an artefact of poor taxon sampling.

In the phylogenetic tree obtained from 8P12RNA under BI approach, the monophyly of all included families and superfamilies were also supported with high support values except for the superfamily Helicoidea which supported with a low nodal support (BPP = 0.82) (Fig. 4). Arionoidea + Orthalicoidea clade was recovered as sister group to Succineoidea + (Urocoptoidea + (Polygyroidea + Helicoidea)). The tree (Fig. 4) also recovered *Deroceras*



**Figure 5** Dated phylogenetic tree. The axis on the bottom refers to million years. Letters in the boxes refers to external calibration points. The split between *A. vulgaris* and *A. rufus* was shown with a red bar, while the remainings were shown with blue bars.

Full-size [DOI: 10.7717/peerj.8603/fig-5](https://doi.org/10.7717/peerj.8603/fig-5)

*reticulatum* (Limacoidea: Agriolimacidae) at the most basal placement and did not support the monophyly of the suborder Helicina similar to the tree in Yang *et al.* (2019).

A chronogram for Stylommatophora divergence times based on the obtained tree topology is shown in Fig. 5. According to our divergence time analysis, the crown age of stylommatophorans was estimated as 138.55 Ma (95% CI [180.8–107.4 Ma]) corresponding to Early Cretaceous. Our estimated times for initial diversification of Stylommatophora are slightly older but broadly congruent with the fossil records and previous studies (Tillier, Masselot & Tillert, 1996; Jörger *et al.*, 2010; Dinapoli & Klussmann-Kolb, 2010). Although Solem & Yochelson (1979) suggested a Paleozoic origin for Stylommatophora, the widely accepted fossil records with recognizable taxa began from Late Cretaceous (Bandel & Riedel, 1994). The Cretaceous origin of stylommatophoran species was also suggested by sequence studies of 28S rDNA fragments by Tillier, Masselot & Tillert (1996), of combined data of 18S, 28S, 16S rDNA and COI by Dinapoli & Klussmann-Kolb (2010) and Jörger *et al.* (2010). The diversification of the stylommatophoran species may have been influenced by the explosive radiation of angiosperms and speciation by host-switching during Cretaceous (Friis, Pedersen & Crane, 2010).

The split time of *Achatina fulica* from other stylommatophoran species was inferred as 131.91 Ma in Early Cretaceous. The splits of the superfamilies Orthalicoidea and Arionoidea, of Succineoidea from Urocoptoidea + (Polygyroidea + Helicoidea), and of Clausilioidea + (Pupilloidea + Achatinelloidea) were dated to 114.18 Ma (95% CI [148.7–87.2 Ma]), 113.30 Ma (95% CI [146.1–87.9 Ma]) and 111.88 Ma (95% CI [148.1–84.2 Ma]), respectively, coinciding to the beginning of the Albian (Early Cretaceous). The

crown ages of the superfamilies Arionoidea, Urocoptoidea, Helicoidea and Pupilloidea were estimated corresponding to Late Cretaceous (84.82, 75.23, 74.39 and 70.06 Ma, respectively). The split of the two *Arion* species and the crown age of Achatinelloidea species were dated to 15.24 Ma (95% CI [30.0–7.6 Ma]) and 13.01 Ma (95% CI [19.7–8.2 Ma]), respectively, corresponding to the Miocene. The divergence time of *A. vulgaris* and *A. rufus* corresponds to one of Earth's most recent, global warming events, the Mid-Miocene Climatic Optimum (MMCO, 17–14.75 Ma) (Böhme, 2003). The MMCO is thought to have contributed to floristic and faunistic diversity across the world and so to animal-plant interactions, correlating with the rise in temperature (Barnosky & Carrasco, 2002; Vicentini et al., 2008; Tolley, Chase & Forest, 2008). The change of plant diversity, emergence of new host plants and the relative warm period may have triggered the diversification of the two *Arion* species. The divergence time of two polygyroid species was inferred as 0.45 Ma (95% CI [1.1–0.1 Ma]), in the Pleistocene.

### Selective pressures on stylommatophoran mitogenomes

The  $\omega$  value for each of the 13 PCGs was inferred under one-ratio model using PAML and presented in Table 4. All of the  $\omega$  values were extremely low ( $\omega < 1$ ), ranging between 0.0129 for *COX1* and 0.2198 for *ATP8*, reflecting that all genes were under strong purifying selection consistent with the general mitogenome evolution pattern in animals (Rand, 2001; Bazin, Glemin & Galtier, 2006). Although purifying selection is the predominant selective force shaping stylommatophoran mitogenomes, the comparison of the null neutral model and alternative branch-specific positive selection model revealed six of the PCGs (*ATP6*, *COX2*, *COX3*, *ND2*, *ND4* and *ND5*) have variation in  $\omega$  values along different branches. The variability in  $\omega$  values indicated different selective forces acting on each gene as well as each branch. A more sensitive branch-site method, aBSREL, providing three states for each branch and allowing each site to evolve under any kind of the value ( $<1$ , 1 or  $>1$ ) (Smith et al., 2015), was used for evaluating and confirming the selective forces across lineages determined by PAML analysis. All of the branches in the stylommatophoran phylogeny were tested with aBSREL analysis for each PCG, and the genes detected as under episodic diversifying selection were different from the results of branch-site model of PAML (Table 5) except for *COX3* and *ND4*. The aBSREL analyses discovered episodic diversifying selection in *ATP8* (at the branch leading to *Microceramus pontificus*), *COX1* (at the branch leading to *Achatinella mustelina*), *COX3* (at the branch leading to Arionoidea and the branch leading to *Philomycus bilineatus*), *ND3* (at the branch leading to *Helicella itala*), *ND4* (at the branch leading to *Succinea putris*) and *ND6* (at the branch leading to *Vertigo pusilla*). Due to their important function, mitochondrial genes might have a few positively selected sites and the signatures of purifying selection likely mask those of positive selection (Meiklejohn, Montooth & Rand, 2007; Da Fonseca et al., 2008). Therefore, two different methods were used to detect positive selection in addition to BEB analysis: FUBAR which estimates the rates of nonsynonymous and synonymous substitutions at each codon in a phylogeny, and MEME which estimates the probability for a codon to have experienced episodic positive selection and allows the  $\omega$  ratio to vary across branches and codons. BEB analysis identified eight positively selected codons in total in three genes

**Table 4** Likelihood ratios of PAML analysis showing different selective pressures on the mitochondrial PCGs in *Stylommatophora*.

Models <sup>a</sup>		A		B		C		A–B		B–C	
Gene	$\omega$	lnL <sup>b</sup>	Np <sup>c</sup>	lnL	Np	lnL	Np	LRT	P	LRT	P <sup>d</sup>
<i>ATP6</i>	0.0509	−19658.6890		−19377.7446		−19374.7961		561.8887	0.000	−5.896982	<b>0.015</b>
<i>ATP8</i>	0.2198	−6460.1406		−6355.5745		−6355.5745		209.1322	0.000	−0.000006	1.000
<i>COX1</i>	0.0129	−26428.6827		−26169.9332		−26169.9332		517.4990	0.000	0.000000	1.000
<i>COX2</i>	0.0393	−16022.1910		−15851.6973		−15848.6152		340.9874	0.000	−6.164250	<b>0.013</b>
<i>COX3</i>	0.0317	−18617.4781		−18345.8456		−18341.7284		543.2650	0.000	−8.234404	<b>0.004</b>
<i>CYTB</i>	0.0416	−26893.7004		−26297.6496		−26297.6496		1192.1017	0.000	0.000082	0.992
<i>ND1</i>	0.0430	−24857.6972	76	−24528.7292	78	−24529.0150	79	657.9359	0.000	0.571480	0.450
<i>ND2</i>	0.0670	−31769.5279		−31502.5684		−31498.5036		533.9191	0.000	−8.129668	<b>0.004</b>
<i>ND3</i>	0.0607	−11403.3678		−11183.4729		−11183.4729		439.7899	0.000	0.000010	1.000
<i>ND4</i>	0.0511	−40498.0727		−40032.1913		−40026.0861		931.7628	0.000	−12.210500	<b>0.000</b>
<i>ND4L</i>	0.0727	−10118.1495		−10076.4001		−10075.2708		83.4989	0.000	−2.258408	0.133
<i>ND5</i>	0.0676	−51970.2413		−51088.0681		−51092.3346		1764.3464	0.000	8.533012	<b>0.003</b>
<i>ND6</i>	0.1009	−16916.1956		−16691.7006		−16691.7006		448.9899	0.000	0.000018	1.000

**Notes.**

Degrees of freedom = 1.

<sup>a</sup>A, All branches have one  $\omega$ ; B, All branches have same  $\omega = 1$ ; C, Each branch has its own  $\omega$ .

<sup>b</sup>The natural algorithm of the likelihood value.

<sup>c</sup>Number of parameters.

<sup>d</sup>Bold faced figure indicate the statistical significance ( $P < 0.05$ ).

(*ND2*, *ND4* and *ND5*), whereas FUBAR defined six positively selected codons in five genes (*ATP6*, *ATP8*, *COX2*, *CYTB* and *ND4L*). The MEME analysis found the signals of episodic positive selection at 22 codons in nine genes (*ATP6*, *ATP8*, *CYTB*, *ND2-6*, and *ND4L*). There was not any shared codon determined by all of the three analyses (Table 6). Only four codons in three genes were shared by the results of FUBAR and MEME analyses: 44th codon in *ATP8* gene, 12th codon in *CYTB* gene, and 13th and 57th codons in *ND4L* gene. Therefore, we focused only on these four codons in the TreeSAAP analyses. The positively selected substitution at codon 44 in *ATP8* gene was the change of TTA (Leu) to ATT (Ile) at branches leading to *M. kiangsinensis*, *Cerion incanum* and *Cerion uva*. This substitution was a radical chemical change with a magnitude category of 8 and had an impact on the increment of the equilibrium constant (ionization of COOH). The change at the codon 12 in *CYTB* gene was a conserved change with a magnitude category of 1 and was a substitution of TTG (Leu) to ATG (Met). The positively selected substitutions in *ND4L* gene were the change of ATT (Ile) to ATA (Met) at branch leading to *H. pomatia*, to GTT (Val) at branch leading to *C. nemoralis* at codon 13, and the change of TTT (Phe) to AAT (Asn) at branch leading to Arionidae family at codon 57. The substitution at the 13th codon was a radical change with a magnitude category of 8 altering the equilibrium constant (ionization of COOH), while that at the 57th codon was a radical change with a magnitude category of 7 and modifying the solvent accessibility of the protein.

Consequently, six positive selected genes (*ATP8*, *COX1*, *COX3*, *ND3*, *ND4* and *ND6*) detected by branch-specific aBSREL approach and three genes (*ATP8*, *CYTB* and *ND4L*) detected by codon-based BEB, FUBAR and MEME approaches were exposed to diversifying

**Table 5** Genes and branches detected to be exposed episodic diversifying selection using the aBSREL approach.

Gene	Number of selected branches ( $P < 0.05$ )	Taxon	$\omega$	Proportion of codons under selection
<i>ATP8</i>	1	<i>Microceramus pontificus</i>	288	0.460
<i>COX1</i>	1	<i>Achatinella mustelina</i>	2,180	0.086
<i>COX3</i>	2	Arionoidea	670	0.053
		<i>Philomyces bilineatus</i>	119	0.092
<i>ND3</i>	1	<i>Helicella itala</i>	49.4	0.220
<i>ND4</i>	1	<i>Succinea putris</i>	4.18	0.370
<i>ND6</i>	1	<i>Vertigo pusilla</i>	15.6	0.370

**Table 6** Genes/codons under diversifying or positive selection under codon-based models.

Gene	BEB	FUBAR	MEME
<i>ATP6</i>	–	4	44
<i>ATP8</i>	–	44	44, 57, 64, 92, 109
<i>COX1</i>	–	–	–
<i>COX2</i>	–	32	–
<i>COX3</i>	–	–	–
<i>CYTB</i>	–	12	12
<i>ND1</i>	–	–	–
<i>ND2</i>	188	–	14, 16, 174
<i>ND3</i>	–	–	27
<i>ND4</i>	109, 170, 192, 301, 386, 427	–	9, 99
<i>ND4L</i>	–	13, 57	13, 57, 109, 111
<i>ND5</i>	451	–	260, 501
<i>ND6</i>	–	–	109, 179, 183

selection. Four of these genes (*ND3*, *ND4*, *ND4L* and *ND6*) play an important role in oxidative phosphorylation and are subunits of NADH dehydrogenase (Complex I) which is the most complicated and largest proton pump of the respiratory chain coupling electron transfer from NADH to ubiquinone. In addition to its important role in energy production, it has been shown that complex I is implicated in the regulation of reactive oxygen species (ROS) (Sharma, Lu & Bai, 2009). Substitutions in this complex might have been favoured for increasing the efficiency of proton pumping or regulating the response to ROS depending varying amount of oxygen in the atmosphere and adaptation to conditions in new habitats (temperature, humidity, altitude) and/or hosts. *CYTB* gene encodes only mitogenome derived subunit of Complex III and catalyses reversible electron transfer from ubiquinol to cytochrome c (Da Fonseca et al., 2008). The positively selected sites in complexes I and III have been suggested to contribute to environmental adaptation in different groups such as mammals, birds, fishes and insects (Da Fonseca et al., 2008; Garvin, Bielawski & Gharrett, 2011; Garvin et al., 2014; Melo-Ferreira et al., 2014; Morales et al., 2015; Li et al., 2018). In the cytochrome c oxidase complex (Complex IV), *COX1*

protein catalyses electron transfer to the molecular oxygen; *COX2* and *COX3* belong to the catalytic core of the complex may act as a regulator. *ATP8* gene encodes the part of ATP synthase (Complex V) regulating the assembly of complex (*Da Fonseca et al., 2008*). The favoured substitutions in *COX3* and *ATP8* gene might have an impact on assembly of the complexes IV and V. The positively selected substitutions and random accumulation of variation in mitochondrial PCGs of stylommatophoran species thus seem to be adaptive and affecting mitochondrial ATP production or protection from ROS effects, however effects of substitutions should be examined in a larger sample by considering protein folding and three-dimensional structure of complexes.

### Gene rearrangements in stylommatophoran mitogenomes

The ancestral mitogenome organisation of each node in the phylogeny was inferred using the maximum likelihood approach. The organisation of the hypothetical ancestral Stylommatophora mitogenome (node: A34, Fig. 4) was identical with that of *Deroceras reticulatum* as well as those of *Albinaria caerulea*, *Cernuella virgata* and *Helicella itala*. The mitogenome of *Achatina fulica* has only experienced the transposition of *trnP* to the downstream of *trnA* compared to its most recent ancestral mitogenome organisation. The common ancestors of Clausilioidea + (Pupilloidea + Achatinelloidea) (node: A31, Fig. 4), Orthalicoidea + Arionoidea (node: A22, Fig. 4), and Succineoidea + (Urocoptoidea + (Polygyroidea + Helicoidea)) (node: A18, Fig. 4) maintained the same order of hypothetical ancestral stylommatophoran mitogenome. In the mitogenome of the most recent common ancestor (MRCA) of Pupilloidea + Achatinelloidea (node: A30, Fig. 4), the reversal of *trnW*, *trnG* and *trnH* genes occurred individually and were followed by the reversal of the cluster *trnW-trnG-trnH*. In the superfamily Pupilloidea, rearrangements of several tRNA genes were observed: the transposition of the cluster *trnD-trnC* to downstream of *trnW* in *Pupilla muscorum*, transpositions of cluster *trnH-trnG* to downstream of *trnW* and of *trnT* to upstream of *COX3* in *Orcula dolium*, transposition of *trnG* to downstream of *trnW* in *Vertigo pusilla* and the reversal of *trnQ* in *Gastrocopta cristata*. In the mitogenome of the MRCA of the superfamily Achatinelloidea (node: A26, Fig. 4), *trnF-COX2-trnY-trnH-trnG-trnW-trnQ-ATP8-trnN-ATP6-trnR-trnE-rnsS-trnM* gene cluster rearranged as *trnW-trnQ-ATP8-ATP6-trnR-trnE-rnsS-trnM-trnF-COX2-trnY-trnH-trnG-trnN* via tandem duplication random loss (TDRL) mechanism. The organisation of the mitogenomes of achatinelloid species nearly matched with the putative ancestral order, except for *Achatinella sowerbyana* which has a transposed position of *trnK* to downstream of *ATP8* and a second copy of *trnL2*, and for *Partulina redfieldi* which has the inversion of *trnE* and *trnN* genes.

The mitogenome of *Naesiotus nux* has almost the same organisation with its MRCA (node: A22, Fig. 4), except for the second inverted copy of *ND4L* located between *trnL1* and *trnP*. The MRCA of the superfamily Arionoidea (node: A21, Fig. 4) had also identical mitogenome organisation with the ancestor of Stylommatophora, and the MRCAs of the families Arionidae (node: A19, Fig. 4) and of Philomycidae (node: A20, Fig. 4) were derived from this ancestor. The mitogenome of node A19 had shuffled positions of *trnY* and *trnW*, and also transpositions of *trnE* to downstream of *trnQ* and of *rnsS-trnM* to upstream

of *trnQ*. Both *Arion* mitogenomes also shared this mitogenome organisation and the rearranged orders of *trnW-trnY* and *trnE-trnQ-rrnS-trnM-trnL2-ATP8-trnN-ATP6-trnR* clusters seem to be synapomorphies of this genus. The mitogenome organisation of the node A20 was quite different from those of other stylommatophoran species, which had rearranged positions of almost all genes between *COX1* and *trnI* via two-steps TDRL, and two Philomycidae species also had identical organisation except for *Philomyces bilineatus* had a second copy of *trnC* located downstream of the original copy.

The mitogenome of *Succinea putris* has experienced the transpositions of *trnF* to upstream of *trnD* and of *trnW* to upstream of *trnY*, and also reverse transposition of the cluster *trnW-trnY* to the upstream of *ND3*. The MRCA of the all urocoptoid species (node: A16, Fig. 4) only had the reversal of *trnQ* gene from minor to major strand and *M. pontificus* has also maintained the identical arrangement. The four step requiring rearranged gene cluster was identified in the mitogenome of the MRCA of the genus *Cerion* (node: A15, Fig. 4): (i) reversal of *trnV-rrnL-trnL1*, (ii) reversal of *trnP*, (iii) reversal of *trnA*, and (iv) reversal of the cluster *trnL1-rrnL-trnV-trnP-trnA*. The mitogenome organisation remained the same in all three *Cerion* species and the rearranged state of *trnA-trnP-trnV-rrnL-trnL1* cluster might be a synapomorphy for this genus.

The MRCAs of the polygroid species (node: A12, Fig. 4) and Camaenidae + Bradybaenidae (node: A5, Fig. 4), as well as the *Polygyra cereolus* and *Praticolella mexicana*, had the transposed position of *trnG-trnH* to the upstream of *trnY*. The rearrangement of this cluster as *trnG-trnH-trnY* could be suggested as a synapomorphy for Polygyroidea, but more sampling is required to confirm its status at superfamily level. In the superfamily Helicoidea, the MRCAs of Geomitridae (node: A9, Fig. 4) and Geomitridae + Helicidae (node: A10, Fig. 4) shared the identical mitogenome organisation with the MRCA of Stylommatophora. Both of the Geomitridae species had also same mitogenome organisation except for *ATP8* in *Cernuella virgata*, in which this gene was missing, however it seems to be likely a misannotation. The mitogenome of MRCA of Helicidae species (node: A8, Fig. 4) had experienced the transpositions of *trnP* and cluster *trnT-COX3* to downstream of *ND6* and to upstream of *trnS1*, respectively. The mitogenome organisations of *Helix pomatia*, *Cornu aspersum* and *Cepaea nemoralis* have not changed and *trnA-ND6-trnP* and *trnS2-trnT-COX3-trnS1* gene orders might be interpreted as synapomorphic for these three species. However, the individual reversals of *trnA*, *ND6* and *trnP* genes followed by reversal of the cluster *trnA-ND6-trnP*, and reversal of *trnS1* were observed in *Cylindrus obtusus* mitogenome. In the mitogenomes of the species of the family Camaenidae, only the transpositions of *trnD* and *trnY* to downstream of *COX2* and to upstream of *trnG* were found, respectively. The arrangement of the *trnC-trnF-COX2-trnD-trnY-trnG* cluster could be considered as a synapomorphy for camaenid species, however the taxonomic level of this synapomorphy need to be evaluated in a wider taxonomic range. In the family of Bradybaenidae, the MRCA mitogenome had experienced only the reversal of *trnW*. In addition to this rearrangement, *Aegista* species also have the transposition of *ND3* gene to the downstream of *trnW* and the rearranged position of *ND3-trnW* cluster appears to be a synapomorphy for the genus.



## CONCLUSIONS

The sequencing and annotation of the mitogenome of *A. vulgaris* and its comparison with other stylommatophoran mitogenomes allow us to denote several conclusions: (i) the mitogenome characteristics of *A. vulgaris* are mostly consistent with the reported stylommatophoran mitogenomes; (ii) rearrangement events are detected in the *trnW-trnY* and *trnE-trnQ-rrnS-trnM-trnL2-ATP8-trnN-ATP6-trnR* gene clusters which may be apomorphic for the genus *Arion*, but further investigations are necessary; (iii) stylommatophoran mitogenome sequence information without the saturated positions seems to be useful for reconstructing phylogeny and estimating divergence times, and the taxon set used should be expanded; (iv) although purifying selection is the dominant force in shaping the stylommatophoran mitogenomes, in the background, several codons or different branches have experienced diversifying selection suggesting adaptation to new environmental conditions.

## ADDITIONAL INFORMATION AND DECLARATIONS

### Funding

This study has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 764840. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Grant Disclosures

The following grant information was disclosed by the authors:  
European Union's Horizon 2020 research and innovation programme: 764840.

### Competing Interests

The authors declare there are no competing interests.

### Author Contributions

- Özgül Doğan and Zeyuan Chen conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Michael Schrödl conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.

### DNA Deposition

The following information was supplied regarding the deposition of DNA sequences:  
The raw data is available at GenBank: [MN607980](#).

### Data Availability

The following information was supplied regarding data availability:  
Raw data is available as [Supplemental Files](#).

## Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.8603#supplemental-information>.

## REFERENCES

- Ahn S-J, Martin R, Rao S, Choi M-Y. 2017.** The complete mitochondrial genome of the gray garden slug *Deroceras reticulatum* (Gastropoda: Pulmonata: Stylommatophora). *Mitochondrial DNA Part B* 2:255–256 DOI [10.1080/23802359.2017.1318677](https://doi.org/10.1080/23802359.2017.1318677).
- Anderson S, Bankier AT, Barrell BG, De Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG. 1981.** Sequence and organization of the human mitochondrial genome. *Nature* 290:457–465 DOI [10.1038/290457a0](https://doi.org/10.1038/290457a0).
- Arquez M, Colgan D, Castro LR. 2014.** Sequence and comparison of mitochondrial genomes in the genus *Nerita* (Gastropoda: Neritimorpha: Neritidae) and phylogenetic considerations among gastropods. *Marine Genomics* 15:45–54 DOI [10.1016/j.margen.2014.04.007](https://doi.org/10.1016/j.margen.2014.04.007).
- Ballard JW, Kreitman M. 1995.** Is mitochondrial DNA a strictly neutral marker? *Trends in Ecology & Evolution* 10:485–488.
- Bandel K. 1991.** Gastropods from brackish and fresh water of the Jurassic—Cretaceous transition (a systematic reevaluation). *Berliner Geowissenschaftliche Abhandlungen Reihe A Geologie und Palaeontologie* 134:9–55.
- Bandel K, Riedel F. 1994.** The Late Cretaceous gastropod fauna from Ajka (Bakony Mountains, Hungary): a revision. *Annalen des Naturhistorischen Museums in Wien* 96:1–65.
- Barker GM. 2009.** Gastropods on land: phylogeny, diversity and adaptive morphology. In: *The biology of terrestrial molluscs*. Wallingford: CABI Publishing DOI [10.1079/9780851993188.0001](https://doi.org/10.1079/9780851993188.0001).
- Barnosky AD, Carrasco MA. 2002.** Effects of Oligo-Miocene global climate changes on mammalian species richness in the northwestern quarter of the USA. *Evolutionary Ecology Research* 4:811–841.
- Bazin E, Glemin S, Galtier N. 2006.** Population size does not influence mitochondrial genetic diversity in animals. *Science* 312:570–572 DOI [10.1126/science.1122033](https://doi.org/10.1126/science.1122033).
- Bernt M, Bleidorn C, Braband A, Dambach J, Donath A, Fritzsich G, Golombek A, Hadrys H, Jühling F, Meusemann K, Middendorf M, Misof B, Perseke M, Podsiadlowski L, Von Reumont B, Schierwater B, Schlegel M, Schrödl M, Simon S, Stadler PF, Stöger I, Struck TH. 2013a.** A comprehensive analysis of bilaterian mitochondrial genomes and phylogeny. *Molecular Phylogenetics and Evolution* 69:352–364 DOI [10.1016/j.ympev.2013.05.002](https://doi.org/10.1016/j.ympev.2013.05.002).
- Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsich G, Pütz J, Middendorf M, Stadler PF. 2013b.** MITOS: improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution* 69:313–319 DOI [10.1016/j.ympev.2012.08.023](https://doi.org/10.1016/j.ympev.2012.08.023).

- Bernt M, Merkle D, Ramsch K, Fritzsche G, Perseke M, Bernhard D, Schlegel M, Stadler PF, Middendorf M. 2007. CREx: inferring genomic rearrangements based on common intervals. *Bioinformatics* **23**(21):2957–2958 DOI [10.1093/bioinformatics/btm468](https://doi.org/10.1093/bioinformatics/btm468).
- Böhme M. 2003. The miocene climatic optimum: evidence from ectothermic vertebrates of Central Europe. *Palaeogeography, Palaeoclimatology, Palaeoecology* **195**(3–4):389–401 DOI [10.1016/S0031-0182\(03\)00367-5](https://doi.org/10.1016/S0031-0182(03)00367-5).
- Boore JL. 1999. Animal mitochondrial genomes. *Nucleic Acids Research* **27**:1767–1780.
- Bouchet P, Rocroi J-P, Hausdorf B, Kaim A, Kano Y, Nützel A, Parkhaev P, Schrödl M, Strong EE. 2017. Revised classification, nomenclator and typification of gastropod and monoplacophoran families. *Malacologia* **61**:1–526 DOI [10.4002/040.061.0201](https://doi.org/10.4002/040.061.0201).
- Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**:i884–i890 DOI [10.1093/bioinformatics/bty560](https://doi.org/10.1093/bioinformatics/bty560).
- Da Fonseca RR, Johnson WE, O'Brien SJ, Ramos MJ, Antunes A. 2008. The adaptive evolution of the mammalian mitochondrial genome. *BMC Genomics* **9**:119 DOI [10.1186/1471-2164-9-119](https://doi.org/10.1186/1471-2164-9-119).
- Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* **9**:772–772 DOI [10.1038/nmeth.2109](https://doi.org/10.1038/nmeth.2109).
- Darty K, Denise A, Ponty Y. 2009. VARNA: interactive drawing and editing of the RNA secondary structure. *Bioinformatics* **25**:1974–1975 DOI [10.1093/bioinformatics/btp250](https://doi.org/10.1093/bioinformatics/btp250).
- Dayrat B, Conrad M, Balayan S, White TR, Albrecht C, Golding R, Gomes SR, Harsawych MG, De Frias Martins AM. 2011. Phylogenetic relationships and evolution of pulmonate gastropods (Mollusca): new insights from increased taxon sampling. *Molecular Phylogenetics and Evolution* **59**:425–437 DOI [10.1016/j.ympev.2011.02.014](https://doi.org/10.1016/j.ympev.2011.02.014).
- Deng P-J, Wang W-M, Huang X-C, Wu X-P, Xie G-L, Ouyang S. 2016. The complete mitochondrial genome of Chinese land snail *Mastigculota kiangsinensis* (Gastropoda: Pulmonata: Bradybaenidae). *Mitochondrial DNA* **27**:1441–1442 DOI [10.3109/19401736.2014.953083](https://doi.org/10.3109/19401736.2014.953083).
- De Rijk P, Wuyts J, De Wachter R. 2003. RnaViz 2: an improved representation of RNA secondary structure. *Bioinformatics* **19**:299–300.
- Dinapoli A, Klussmann-Kolb A. 2010. The long way to diversity—phylogeny and evolution of the Heterobranchia (Mollusca: Gastropoda). *Molecular Phylogenetics and Evolution* **55**:60–76 DOI [10.1016/j.ympev.2009.09.019](https://doi.org/10.1016/j.ympev.2009.09.019).
- Doyle J, Doyle J. 1987. A rapid isolation procedure for small amounts of leaf tissue. *Phytochemical Bulletin* **19**:11–15 DOI [10.2307/4119796](https://doi.org/10.2307/4119796).
- Friis EM, Pedersen KR, Crane PR. 2010. Diversity in obscurity: fossil flowers and the early history of angiosperms. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**:369–382 DOI [10.1098/rstb.2009.0227](https://doi.org/10.1098/rstb.2009.0227).
- Gaitán-Espitia JD, Nespolo RF, Opazo JC. 2013. The complete mitochondrial genome of the land snail *Cornu aspersum* (Helicidae: Mollusca): intra-specific divergence

- of protein-coding genes and phylogenetic considerations within euthyneura. *PLOS ONE* **8**:e67299 DOI [10.1371/journal.pone.0067299](https://doi.org/10.1371/journal.pone.0067299).
- Garvin MR, Bielawski JP, Gharrett AJ. 2011.** Positive darwinian selection in the piston that powers proton pumps in Complex I of the mitochondria of Pacific salmon. *PLOS ONE* **6**:e24127 DOI [10.1371/journal.pone.0024127](https://doi.org/10.1371/journal.pone.0024127).
- Garvin MR, Bielawski JP, Sazanov LA, Gharrett AJ. 2014.** Review and meta-analysis of natural selection in mitochondrial complex I in metazoans. *Journal of Zoological Systematics and Evolutionary Research* **53**:1–17 DOI [10.1111/jzs.12079](https://doi.org/10.1111/jzs.12079).
- González VL, Kayal E, Halloran M, Shrestha Y, Harasewych MG. 2016.** The complete mitochondrial genome of the land snail *Cerion incanum* (Gastropoda: Stylommatophora) and the phylogenetic relationships of Cerionidae within Panpulmonata. *Journal of Molluscan Studies* **82**:525–533 DOI [10.1093/mollus/eyw017](https://doi.org/10.1093/mollus/eyw017).
- Grande C, Templado J, Zardoya R. 2008.** Evolution of gastropod mitochondrial genome arrangements. *BMC Evolutionary Biology* **8**:61 DOI [10.1186/1471-2148-8-61](https://doi.org/10.1186/1471-2148-8-61).
- Gray MW. 1989.** Origin and evolution of mitochondrial DNA. *Annual Review of Cell Biology* **5**:25–50 DOI [10.1146/annurev.cb.05.110189.000325](https://doi.org/10.1146/annurev.cb.05.110189.000325).
- Greiner S, Lehwark P, Bock R. 2019.** OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Research* **47**:W59–W64 DOI [10.1093/nar/gkz238](https://doi.org/10.1093/nar/gkz238).
- Groenenberg DSJ, Harl J, Duijm E, Gittenberger E. 2017.** The complete mitogenome of *Orcula dolium* (Draparnaud, 1801); ultra-deep sequencing from a single long-range PCR using the Ion-Torrent PGM. *Hereditas* **154**:Article 7 DOI [10.1186/s41065-017-0028-2](https://doi.org/10.1186/s41065-017-0028-2).
- Groenenberg DSJ, Pirovano W, Gittenberger E, Schilthuizen M. 2012.** The complete mitogenome of *Cylindrus obtusus* (Helicidae, Ariantinae) using Illumina next generation sequencing. *BMC Genomics* **13**:Article 114 DOI [10.1186/1471-2164-13-114](https://doi.org/10.1186/1471-2164-13-114).
- Harasewych MG, González VL, Windsor AM, Halloran M. 2017.** The complete mitochondrial genome of *Cerion uva uva* (Gastropoda: Panpulmonata: Stylommatophora: Cerionidae). *Mitochondrial DNA Part B* **2**:159–160 DOI [10.1080/23802359.2017.1303343](https://doi.org/10.1080/23802359.2017.1303343).
- 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. *Systematic Biology* **54**:277–298 DOI [10.1080/10635150590947843](https://doi.org/10.1080/10635150590947843).
- Hatzoglou E, Rodakis GC, Lecanidou R. 1995.** Complete sequence and gene organization of the mitochondrial genome of the land snail *Albinaria coerulea*. *Genetics* **140**:1353–1366.
- Hausdorf B. 2000.** Biogeography of the Limacoidea sensu lato (Gastropoda: Stylommatophora): vicariance events and long-distance dispersal. *Journal of Biogeography* **27**:379–390 DOI [10.1046/j.1365-2699.2000.00403.x](https://doi.org/10.1046/j.1365-2699.2000.00403.x).
- He Z-P, Dai X-B, Zhang S, Zhi T-T, Lun Z-R, Wu Z-D, Yang T-B. 2016.** Complete mitochondrial genome of the giant African snail, *Achatina fulica* (Mollusca: Achatinidae): a novel location of putative control regions (CR) in the

- mitogenome within Pulmonate species. *Mitochondrial DNA* **27**:1084–1085  
DOI [10.3109/19401736.2014.930833](https://doi.org/10.3109/19401736.2014.930833).
- Holznagel WE, Colgan DJ, Lydeard C. 2010.** Pulmonate phylogeny based on 28S rRNA gene sequences: a framework for discussing habitat transitions and character transformation. *Molecular Phylogenetics and Evolution* **57**:1017–1025  
DOI [10.1016/j.ympev.2010.09.021](https://doi.org/10.1016/j.ympev.2010.09.021).
- Hu F, Lin Y, Tang J. 2014.** MLGO: phylogeny reconstruction and ancestral inference from gene-order data. *BMC Bioinformatics* **15**:354 DOI [10.1186/s12859-014-0354-6](https://doi.org/10.1186/s12859-014-0354-6).
- Huang C-W, Lin S-M, Wu W-L. 2015.** Mitochondrial genome sequences of land-snails *Aegista diversifamilia* and *Dolicheulota formosensis* (Gastropoda: Pulmonata: Stylommatophora). *Mitochondrial DNA Part A* **27**(4):2793–2795  
DOI [10.3109/19401736.2015.1053070](https://doi.org/10.3109/19401736.2015.1053070).
- Hunter SS, Settles ML, New DD, Parent CE, Gerritsen AT. 2016.** Mitochondrial genome sequence of the galápagos endemic land snail *Naesiotus nux*. *Genome Announcements* **4**:e01362–15 DOI [10.1128/genomeA.01362-15](https://doi.org/10.1128/genomeA.01362-15).
- Jörger KM, Stöger I, Kano Y, Fukuda H, Knebelsberger T, Schrödl M. 2010.** On the origin of Acochlidia and other enigmatic euthyneuran gastropods, with implications for the systematics of Heterobranchia. *BMC Evolutionary Biology* **10**:323  
DOI [10.1186/1471-2148-10-323](https://doi.org/10.1186/1471-2148-10-323).
- Kano Y, Brenzinger B, Nützel A, Wilson NG, Schrödl M. 2016.** Ringiculid bubble snails recovered as the sister group to sea slugs (Nudipleura). *Scientific Reports* **6**:30908  
DOI [10.1038/srep30908](https://doi.org/10.1038/srep30908).
- Katoh K, Standley DM. 2013.** MAFFT Multiple Sequence Alignment Software Version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**:772–780 DOI [10.1093/molbev/mst010](https://doi.org/10.1093/molbev/mst010).
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012.** Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**:1647–1649  
DOI [10.1093/bioinformatics/bts199](https://doi.org/10.1093/bioinformatics/bts199).
- Korábek O, Petrusek A, Rovatsos M. 2019.** The complete mitogenome of *Helix pomatia* and the basal phylogeny of Helicinae (Gastropoda, Stylommatophora, Helicidae). *ZooKeys* **827**:19–30 DOI [10.3897/zookeys.827.33057](https://doi.org/10.3897/zookeys.827.33057).
- Kumar S, Stecher G, Tamura K. 2016.** MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* **33**:1870–1874  
DOI [10.1093/molbev/msw054](https://doi.org/10.1093/molbev/msw054).
- Lanfear R, Calcott B, Ho SYW, Guindon S. 2012.** PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution* **29**:1695–1701 DOI [10.1093/molbev/mss020](https://doi.org/10.1093/molbev/mss020).
- Laslett D, Canbäck B. 2008.** ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. *Bioinformatics* **24**:172–175  
DOI [10.1093/bioinformatics/btm573](https://doi.org/10.1093/bioinformatics/btm573).

- Li X-D, Jiang G-F, Yan L-Y, Li R, Mu Y, Deng W-A. 2018. Positive selection drove the adaptation of mitochondrial genes to the demands of flight and high-altitude environments in grasshoppers. *Frontiers in Genetics* **9**:605 DOI [10.3389/fgene.2018.00605](https://doi.org/10.3389/fgene.2018.00605).
- Lin J-H, Zhou W, Ding H-L, Wang P, Ai H-M. 2016. The mitochondrial genome of the land snail *Cerutuella virgata* (Da Costa, 1778): the first complete sequence in the family Hygromiidae (Pulmonata, Stylommatophora). *ZooKeys* **589**:55–69 DOI [10.3897/zookeys.589.7637](https://doi.org/10.3897/zookeys.589.7637).
- Liu G-H, Wang S-Y, Huang W-Y, Zhao G-H, Wei S-J, Song H-Q, Xu M-J, Lin R-Q, Zhou D-H, Zhu X-Q. 2012. The complete mitochondrial genome of *Galba pervia* (Gastropoda: Mollusca), an intermediate host snail of *Fasciola* spp. *PLOS ONE* **7**:e42172 DOI [10.1371/journal.pone.0042172](https://doi.org/10.1371/journal.pone.0042172).
- Meiklejohn CD, Montooth KL, Rand DM. 2007. Positive and negative selection on the mitochondrial genome. *Trends in Genetics* **23**:259–263 DOI [10.1016/j.tig.2007.03.008](https://doi.org/10.1016/j.tig.2007.03.008).
- Melo-Ferreira J, Vilela J, Fonseca MM, Fonseca RRD, Boursot P, Alves PC. 2014. The elusive nature of adaptive mitochondrial DNA evolution of an arctic lineage prone to frequent introgression. *Genome Biology and Evolution* **6**:886–896 DOI [10.1093/gbe/evu059](https://doi.org/10.1093/gbe/evu059).
- Meng G, Li Y, Yang C, Liu S. 2019. MitoZ: a toolkit for animal mitochondrial genome assembly, annotation and visualization. *Nucleic Acids Research* **47**(11):e63 DOI [10.1093/nar/gkz173](https://doi.org/10.1093/nar/gkz173).
- Minton RL, Cruz MAMartinez, Farman ML, Perez KE. 2016. Two complete mitochondrial genomes from *Praticolella mexicana* Perez, 2011 (Polygyridae) and gene order evolution in Helicoidea (Mollusca, Gastropoda). *ZooKeys* **626**:137–154 DOI [10.3897/zookeys.626.9633](https://doi.org/10.3897/zookeys.626.9633).
- Morales HE, Pavlova A, Joseph L, Sunnucks P. 2015. Positive and purifying selection in mitochondrial genomes of a bird with mitonuclear discordance. *Molecular Ecology* **24**:2820–2837 DOI [10.1111/mec.13203](https://doi.org/10.1111/mec.13203).
- Mordan P, Wade C. 2008. Heterobranchia II: the pulmonata. In: *Phylogeny and evolution of the Mollusca*. Berkeley, California: University of California Press, 409–426 DOI [10.1525/california/9780520250925.003.0015](https://doi.org/10.1525/california/9780520250925.003.0015).
- Moritz C, Dowling TE, Brown WM. 1987. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annual Review of Ecology and Systematics* **18**:269–292 DOI [10.1146/annurev.es.18.110187.001413](https://doi.org/10.1146/annurev.es.18.110187.001413).
- Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Pond SLKosakovsky, Scheffler K. 2013. FUBAR: a fast, unconstrained Bayesian approximation for inferring selection. *Molecular Biology and Evolution* **30**:1196–1205 DOI [10.1093/molbev/mst030](https://doi.org/10.1093/molbev/mst030).
- Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Pond SLKosakovsky. 2012. Detecting individual sites subject to episodic diversifying selection. *PLOS Genetics* **8**:e1002764 DOI [10.1371/journal.pgen.1002764](https://doi.org/10.1371/journal.pgen.1002764).
- Ojala D, Montoya J, Attardi G. 1981. tRNA punctuation model of RNA processing in human mitochondria. *Nature* **290**:470–474 DOI [10.1038/290470a0](https://doi.org/10.1038/290470a0).

- Osca D, Templado J, Zardoya R. 2014. The mitochondrial genome of *Ifremeria nautili* and the phylogenetic position of the enigmatic deep-sea Abyssochrysoidea (Mollusca: Gastropoda). *Gene* 547:257–266 DOI 10.1016/j.gene.2014.06.040.
- Perna NT, Kocher TD. 1995. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. *Journal of Molecular Evolution* 41:353–358 DOI 10.1007/BF01215182.
- Ponder WF, Lindberg DR. 1997. Towards a phylogeny of gastropod molluscs: an analysis using morphological characters. *Zoological Journal of the Linnean Society* 119:83–265 DOI 10.1006/zjls.1996.0066.
- Price MR, Forsman ZH, Knapp I, Hadfield MG, Toonen RJ. 2016a. The complete mitochondrial genome of *Achatinella mustelina* (Gastropoda: Pulmonata: Stylommatophora). *Mitochondrial DNA Part B* 1:175–177 DOI 10.1080/23802359.2016.1149787.
- Price MR, Forsman ZH, Knapp IS, Toonen RJ, Hadfield MG. 2016b. The complete mitochondrial genome of *Achatinella sowerbyana* (Gastropoda: Pulmonata: Stylommatophora: Achatinellidae). *Mitochondrial DNA Part B* 1:666–668 DOI 10.1080/23802359.2016.1219631.
- Price MR, Forsman ZH, Knapp I, Toonen RJ, Hadfield MG. 2018. A comparison of mitochondrial genomes from five species in three genera suggests polyphyly in the subfamily Achatinellinae (Gastropoda: Pulmonata: Stylommatophora: Achatinellidae). *Mitochondrial DNA Part B* 3:611–612 DOI 10.1080/23802359.2018.1473737.
- Rambaut A. 2012. FigTree v1. 4.0. In: *A graphical viewer of phylogenetic trees*. Edinburgh: Institute of Evolutionary Biology University of Edinburgh.
- Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior summarization in Bayesian phylogenetics using tracer 1.7. *Systematic Biology* 67:901–904 DOI 10.1093/sysbio/syy032.
- Rand DM. 2001. The units of selection on mitochondrial DNA. *Annual Review of Ecology and Systematics* 32:415–448 DOI 10.1146/annurev.ecolsys.32.081501.114109.
- Razkin O, Gómez-Moliner BJ, Prieto CE, Martínez-Ortí A, Arrébola JR, Muñoz B, Chueca LJ, Madeira MJ. 2015. Molecular phylogeny of the western Palaearctic Helicoidea (Gastropoda, Stylommatophora). *Molecular Phylogenetics and Evolution* 83:99–117 DOI 10.1016/j.ympev.2014.11.014.
- Reyes A, Gissi C, Pesole G, Saccone C. 1998. Asymmetrical directional mutation pressure in the mitochondrial genome of mammals. *Molecular Biology and Evolution* 15(8):957–966 DOI 10.1093/oxfordjournals.molbev.a026011.
- Rice P, Longden L, Bleasby A. 2000. EMBOSS: the European molecular biology open software suite. *Trends in Genetics* 16(6):276–277 DOI 10.1016/S0168-9525(00)02024-2.
- Romero PE, Weigand AM, Pfenninger M. 2016. Positive selection on panpulmonate mitogenomes provide new clues on adaptations to terrestrial life. *BMC Evolutionary Biology* 16:164 DOI 10.1186/s12862-016-0735-8.
- Ronquist F, Teslenko M, Van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian

- phylogenetic inference and model choice across a large model space. *Systematic Biology* **61**:539–542 DOI [10.1093/sysbio/sys029](https://doi.org/10.1093/sysbio/sys029).
- Roth B, Poinar GO, Acra A, Acra F. 1996.** Probable pupillid land snail of Early Cretaceous (Hauterivian) age in amber from Lebanon. *Veliger* **39**:87–88.
- Sevigny JL, Kirouac LE, Thomas WK, Ramsdell JS, Lawlor KE, Sharifi O, Grewal S, Baysdorfer C, Curr K, Naimie AA, Okamoto K, Murray JA, Newcomb JM. 2015.** The mitochondrial genomes of the Nudibranch mollusks, *Melibe leonina* and *Tritonia diomedea*, and their impact on gastropod phylogeny. *PLOS ONE* **10**:e0127519 DOI [10.1371/journal.pone.0127519](https://doi.org/10.1371/journal.pone.0127519).
- Sharma L, Lu J, Bai Y. 2009.** Mitochondrial respiratory complex i: structure, function and implication in human diseases. *Current Medicinal Chemistry* **16**(10):1266–1277 DOI [10.2174/092986709787846578](https://doi.org/10.2174/092986709787846578).
- Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Pond SLKosakovsky. 2015.** Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Molecular Biology and Evolution* **32**:1342–1353 DOI [10.1093/molbev/msv022](https://doi.org/10.1093/molbev/msv022).
- Solem A, Yochelson EL. 1979.** North American Paleozoic land snails, with a summary of other Paleozoic non-marine snails. In: *Geological survey professional paper*.
- Song L, Florea L, Langmead B. 2014.** Lighter: fast and memory-efficient sequencing error correction without counting. *Genome Biology* **15**:Article 509 DOI [10.1186/s13059-014-0509-9](https://doi.org/10.1186/s13059-014-0509-9).
- Stamatakis A. 2014.** RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**:1312–1313 DOI [10.1093/bioinformatics/btu033](https://doi.org/10.1093/bioinformatics/btu033).
- Stöger I, Schrödl M. 2013.** Mitogenomics does not resolve deep molluscan relationships (yet?). *Molecular Phylogenetics and Evolution* **69**:376–392 DOI [10.1016/j.ympev.2012.11.017](https://doi.org/10.1016/j.ympev.2012.11.017).
- Sun BN, Wei LL, Shen HD, Wu HX, Wang DF. 2016.** Phylogenetic analysis of euthyneuran gastropods from sea to land mainly based on comparative mitogenomic of four species of Onchidiidae (Mollusca: Gastropoda: Pulmonata). *Mitochondrial DNA Part A* **27**:3075–3077 DOI [10.3109/19401736.2014.1003916](https://doi.org/10.3109/19401736.2014.1003916).
- Swofford DL. 2002.** *Phylogenetic analysis using parsimony (\*and other methods). Version 4.* Sunderland: Sinauer Associates.
- Tillier S, Masselmot M, Tillirt A. 1996.** Phylogenetic relationships of the Pulmonate Gastropods from rRNA sequences, and tempo and age of the Stylommatophoran radiation. In: Taylor JD, ed. *Origin and evolutionary radiation of the Mollusca*. Oxford: Oxford Press, 267–284.
- Tolley KA, Chase BM, Forest F. 2008.** Speciation and radiations track climate transitions since the Miocene Climatic Optimum: a case study of southern African chameleons. *Journal of Biogeography* **35**:1402–1414 DOI [10.1111/j.1365-2699.2008.01889.x](https://doi.org/10.1111/j.1365-2699.2008.01889.x).
- Uribe JE, Colgan D, Castro LR, Kano Y, Zardoya R. 2016a.** Phylogenetic relationships among superfamilies of Neritimorpha (Mollusca: Gastropoda). *Molecular Phylogenetics and Evolution* **104**:21–31 DOI [10.1016/j.ympev.2016.07.021](https://doi.org/10.1016/j.ympev.2016.07.021).



- Uribe JE, Kano Y, Templado J, Zardoya R. 2016b. Mitogenomics of Vetigastropoda: insights into the evolution of pallial symmetry. *Zoologica Scripta* 45:145–159 DOI 10.1111/zsc.12146.
- Uribe JE, Zardoya R, Puillandre N. 2018. Phylogenetic relationships of the conoidean snails (Gastropoda: Caenogastropoda) based on mitochondrial genomes. *Molecular Phylogenetics and Evolution* 127:898–906 DOI 10.1016/j.ympev.2018.06.037.
- Vaidya G, Lohman DJ, Meier R. 2011. SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. *Cladistics* 27:171–180 DOI 10.1111/j.1096-0031.2010.00329.
- Vicentini A, Barber JC, Aliscioni SS, Giussani LM, Kellogg EA. 2008. The age of the grasses and clusters of origins of C4 photosynthesis. *Global Change Biology* 14(12):2963–2977 DOI 10.1111/j.1365-2486.2008.01688.x.
- Wade CM, Mordan PB, Clarke B. 2001. A phylogeny of the land snails (Gastropoda: Pulmonata). *Proceedings of the Royal Society of London. Series B: Biological Sciences* 268:413–422 DOI 10.1098/rspb.2000.1372.
- Wade CM, Mordan PB, Naggs F. 2006. Evolutionary relationships among the Pulmonate land snails and slugs (Pulmonata, Stylommatophora). *Biological Journal of the Linnean Society* 87(4):593–610 DOI 10.1111/j.1095-8312.2006.00596.x.
- Wang P, Yang H, Zhou W, Hwang C, Zhang W, Qian Z. 2014. The mitochondrial genome of the land snail *Camaena cicatricosa* (Müller, 1774) (Stylommatophora, Camaenidae): the first complete sequence in the family Camaenidae. *ZooKeys* 451:33–48 DOI 10.3897/zookeys.451.8537.
- Weaver S, Shank SD, Spielman SJ, Li M, Muse SV, Kosakovsky Pond SL. 2018. Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. *Molecular Biology and Evolution* 35:773–777 DOI 10.1093/molbev/msx335.
- White TR, Conrad MM, Tseng R, Balayan S, Golding R, Martins AMdeFrias, Dayrat BA. 2011. Ten new complete mitochondrial genomes of pulmonates (Mollusca: Gastropoda) and their impact on phylogenetic relationships. *BMC Evolutionary Biology* 11:295 DOI 10.1186/1471-2148-11-295.
- Williams ST, Foster PG, Littlewood DTJ. 2014. The complete mitochondrial genome of a turbinid vetigastropod from MiSeq Illumina sequencing of genomic DNA and steps towards a resolved gastropod phylogeny. *Gene* 533:38–47 DOI 10.1016/j.gene.2013.10.005.
- Wolstenholme DR. 1992. Animal mitochondrial DNA: structure and evolution. *International Review of Cytology* 141:173–216 DOI 10.1016/S0074-7696(08)62066-5.
- Woolley S, Johnson J, Smith MJ, Crandall KA, McClellan DA. 2003. TreeSAAP: selection on amino acid properties using phylogenetic trees. *Bioinformatics* 19:671–672 DOI 10.1093/bioinformatics/btg043.
- Xie J, Feng J, Guo Y, Ye Y, Li J, Guo B. 2019a. The complete mitochondrial genome and phylogenetic analysis of *Nerita yoldii* (Gastropoda: Neritidae). *Mitochondrial DNA Part B* 4:1099–1100 DOI 10.1080/23802359.2019.1586485.

- Xie G-L, Köhler F, Huang X-C, Wu R-W, Zhou C-H, Ouyang S, Wu X-P. 2019b.** A novel gene arrangement among the Stylommatophora by the complete mitochondrial genome of the terrestrial slug *Meghimatium bilineatum* (Gastropoda, Arionoidea). *Molecular Phylogenetics and Evolution* **135**:177–184 DOI [10.1016/j.ympev.2019.03.002](https://doi.org/10.1016/j.ympev.2019.03.002).
- Yamazaki N, Ueshima R, Terrett JA, Yokobori SI, Kaifu M, Segawa R, Kobayashi T, Numachi KI, Ueda T, Nishikawa K, Watanabe K, Thomas RH. 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.
- Yang Z. 2007.** PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* **24**:1586–1591 DOI [10.1093/molbev/msm088](https://doi.org/10.1093/molbev/msm088).
- Yang X, Xie G-L, Wu X-P, Ouyang S. 2016.** The complete mitochondrial genome of Chinese land snail *Aegista aubryana* (Gastropoda: Pulmonata: Bradybaenidae). *Mitochondrial DNA Part A* **27**:3538–3539 DOI [10.3109/19401736.2015.1074207](https://doi.org/10.3109/19401736.2015.1074207).
- Yang T, Xu G, Gu B, Shi Y, Mzuka HL, Shen H. 2019.** The complete mitochondrial genome sequences of the *Philomycus bilineatus* (Stylommatophora: Philomycidae) and phylogenetic analysis. *Genes* **10**:Article 198 DOI [10.3390/genes10030198](https://doi.org/10.3390/genes10030198).
- Zapata F, Wilson NG, Howison M, Andrade SCS, Jörger KM, Schrödl M, Goetz FE, Giribet G, Dunn CW. 2014.** Phylogenomic analyses of deep gastropod relationships reject Orthogastropoda. *Proceedings of the Royal Society B: Biological Sciences* **281**:Article 20141739 DOI [10.1098/rspb.2014.1739](https://doi.org/10.1098/rspb.2014.1739).
- Zuker M. 2003.** Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research* **31**:3406–3415 DOI [10.1093/nar/gkg595](https://doi.org/10.1093/nar/gkg595).

Supplemental Tables and Figures for Publication I

**The complete mitogenome of *Arion vulgaris* Moquin-Tandon, 1855 (Gastropoda: Stylommatophora): mitochondrial genome architecture, evolution and phylogenetic considerations within Stylommatophora.**

Table S1. Summary of best fit partition schemes and nucleotide substitution models

	Partition scheme	Model
P1	<i>ATP6</i> 1st + <i>ND1</i> 1st + <i>ND2</i> 1st + <i>ND3</i> 1st + <i>ND4</i> 1st + <i>ND4L</i> 1st + <i>ND5</i> 1st + <i>ND6</i> 1st	GTR + I + G
P2	<i>ATP6</i> 2nd + <i>CYTB</i> 2nd + <i>ND1</i> 2nd + <i>ND2</i> 2nd + <i>ND3</i> 2nd + <i>ND4</i> 2nd + <i>ND4L</i> 2nd + <i>ND5</i> 2nd + <i>ND6</i> 2nd	GTR + I + G
P3	<i>ATP6</i> 3rd + <i>COX1</i> 3rd + <i>COX2</i> 3rd + <i>COX3</i> 3rd + <i>CYTB</i> 3rd + <i>ND1</i> 3rd + <i>ND2</i> 3rd + <i>ND3</i> 3rd + <i>ND4</i> 3rd + <i>ND4L</i> 3rd + <i>ND5</i> 3rd + <i>ND6</i> 3rd	HKY + I + G
P4	<i>ATP8</i> 1st + <i>ATP8</i> 2nd + <i>ATP8</i> 3rd + <i>rrnL</i> + <i>rrnS</i> + <i>trnA</i> + <i>trnR</i> + <i>trnN</i> + <i>trnD</i> + <i>trnC</i> + <i>trnQ</i> + <i>trnE</i> + <i>trnG</i> + <i>trnH</i> + <i>trnI</i> + <i>trnL1</i> + <i>trnL2</i> + <i>trnK</i> + <i>trnM</i> + <i>trnF</i> + <i>trnP</i> + <i>trnS1</i> + <i>trnS2</i> + <i>trnT</i> + <i>trnW</i> + <i>trnY</i> + <i>trnV</i>	GTR + I + G
P5	<i>COX1</i> 1st + <i>COX2</i> 1st + <i>COX3</i> 1st + <i>CYTB</i> 1st	GTR + I + G
P6	<i>COX1</i> 2nd + <i>COX2</i> 2nd + <i>COX3</i> 2nd	GTR + I + G

Table S2. Regression of the pairwise distances of the first and second codon positions of each PCG, all positions of RNA genes and the first, second, first and second, and third codon positions of the concatenated mitochondrial PCGs.

Gene	Regression	R	Average GTR distance
<i>ATP6</i>	$y = 0.0626x + 0.274$	0.7819	1.5624
<i>ATP8</i>	$y = 0.0001x + 0.3919$	0.5420	1067.7146
<i>COX1</i>	$y = 0.6084x + 0.0221$	0.9935	0.1611
<i>COX2</i>	$y = 0.1909x + 0.1527$	0.8960	0.5734
<i>COX3</i>	$y = 0.3148x + 0.0879$	0.9620	0.4283
<i>CYTB</i>	$y = 0.3148x + 0.0921$	0.9594	0.4709
<i>ND1</i>	$y = 0.1695x + 0.1643$	0.9133	0.9260
<i>ND2</i>	$y = 9E-06x + 0.4495$	0.4323	2040.3701
<i>ND3</i>	$y = 2E-05x + 0.3981$	0.5514	1127.3637
<i>ND4</i>	$y = 0.0642x + 0.2721$	0.7938	1.7590
<i>ND4L</i>	$y = 2E-05x + 0.4225$	0.5594	4150.3960
<i>ND5</i>	$y = 0.0465x + 0.3076$	0.7014	1.7954
<i>ND6</i>	$y = 0.0002x + 0.3434$	0.6349	937.2557
rRNAs	$y = 0.0499x + 0.2856$	0.7631	2.0819
tRNAs	$y = 0.0236x + 0.3405$	0.6491	2.5675
1st codon position	$y = 0.0712x + 0.2618$	0.8148	1.7962
2nd codon position	$y = 0.2507x + 0.1118$	0.9456	0.5990
1st and 2nd codon position	$y = 0.1812x + 0.1606$	0.9168	0.9114
3rd codon position	$y = 0.0272x + 0.0951$	0.4202	17.2012

Table S3. Nucleotide compositions and skewness values of whole mitogenomes of stylommatophoran species

Species	T%	C%	A%	G%	A+T%	AT-skew	GC-skew
<i>Achatina fulica</i>	35.47	17.10	27.97	19.46	63.44	-0.1183	0.0645
<i>Achatinella fulgens</i>	42.55	9.27	36.38	11.79	78.94	-0.0782	0.1194
<i>Achatinella mustelina</i>	42.76	8.83	37.31	11.10	80.07	-0.0680	0.1140
<i>Achatinella sowerbyana</i>	42.61	9.18	36.51	11.70	79.12	-0.0770	0.1202
<i>Aegista aubryana</i>	37.86	14.45	31.32	16.36	69.18	-0.0946	0.0620
<i>Aegista diversifamilia</i>	38.59	13.26	32.48	15.66	71.07	-0.0860	0.0830
<i>Albinaria caerulea</i>	37.90	13.81	32.75	15.54	70.65	-0.0728	0.0591
<i>Arion rufus</i>	37.14	14.61	32.16	16.09	69.30	-0.0720	0.0485
<i>Arion vulgaris</i>	37.75	14.26	32.45	15.54	70.20	-0.0756	0.0431
<i>Camaena cicatricose</i>	37.90	13.47	31.90	16.72	69.80	-0.0860	0.1077
<i>Camaena poyuensis</i>	38.09	13.31	31.29	17.31	69.38	-0.0980	0.1304
<i>Cepaea nemoralis</i>	33.63	18.94	26.16	21.26	59.79	-0.1249	0.0577
<i>Cerion incanum</i>	35.97	15.83	29.75	18.45	65.72	-0.0948	0.0765
<i>Cerion tridentatum costellata</i>	36.08	15.62	28.18	20.11	64.27	-0.1229	0.1257
<i>Cerion uva</i>	34.45	17.29	28.29	19.98	62.73	-0.0982	0.0721
<i>Cernuella virgata</i>	36.89	15.59	29.07	18.46	65.96	-0.1186	0.0843
<i>Cornu aspersum</i>	39.15	13.61	30.72	16.52	69.87	-0.1207	0.0966
<i>Cylindrus obtusus</i>	35.76	16.61	25.78	21.86	61.53	-0.1622	0.1367
<i>Deroceras reticulatum</i>	39.14	12.17	31.04	17.65	70.17	-0.1154	0.1838
<i>Dolicheulota formosensis</i>	41.81	13.12	28.38	16.70	70.18	-0.1914	0.1199
<i>Gastrocopta cristata</i>	38.39	13.62	30.80	17.19	69.19	-0.1096	0.1159
<i>Helicella itala</i>	37.28	15.27	28.94	18.51	66.22	-0.1260	0.0958
<i>Helix pomatia</i>	37.41	15.07	29.60	17.92	67.01	-0.1166	0.0862
<i>Mastigeulota kiangsinesis</i>	37.91	14.38	29.48	18.22	67.40	-0.1251	0.1176
<i>Meghimatium bilineatum</i>	39.55	13.92	31.89	14.64	71.44	-0.1072	0.0251
<i>Microceramus pontificus</i>	39.06	12.29	32.88	15.77	71.94	-0.0860	0.1238
<i>Naesiotus nux</i>	39.69	12.03	33.57	14.71	73.26	-0.0834	0.1004
<i>Orcula dolium</i>	35.77	16.28	30.21	17.74	65.98	-0.0844	0.0431
<i>Partulina redfieldi</i>	42.44	8.87	37.23	11.46	79.67	-0.0654	0.1274
<i>Philomycus bilineatus</i>	39.44	13.69	32.68	14.20	72.11	-0.0938	0.0182
<i>Polygyra cereolus</i>	39.78	12.58	28.85	18.78	68.64	-0.1592	0.1978
<i>Praticolella mexicana</i>	39.42	13.01	28.57	19.00	67.99	-0.1596	0.1869
<i>Pupilla muscorum</i>	39.27	12.93	32.52	15.27	71.79	-0.0941	0.0829
<i>Succinea putris</i>	43.06	10.87	33.94	12.13	77.00	-0.1185	0.0551
<i>Vertigo pusilla</i>	39.66	12.27	32.57	15.49	72.23	-0.0982	0.1159

Table S4. Nucleotide and amino acid identities of mitochondrial protein coding genes among stylommatophoran and between *Arion* species

Genes	Among Stylommatophora		Between <i>Arion</i> species	
	Nucleotide identity (%)	Amino acid identity (%)	Nucleotide identity (%)	Amino acid identity (%)
<i>ATP6</i>	20.39	22.36	79.39	81.36
<i>ATP8</i>	34.73	28.57	74.24	68.18
<i>COX1</i>	41.94	56.45	87.56	97.45
<i>COX2</i>	32.30	35.14	86.94	96.40
<i>COX3</i>	36.69	42.38	86.92	94.23
<i>CYTB</i>	29.14	34.76	84.87	89.42
<i>ND1</i>	22.78	25.15	84.78	90.00
<i>ND2</i>	19.18	17.71	79.14	78.93
<i>ND3</i>	24.04	21.09	78.84	80.00
<i>ND4</i>	16.87	16.67	83.30	85.81
<i>ND4L</i>	28.72	25.16	84.00	83.00
<i>ND5</i>	18.02	19.83	82.33	84.52
<i>ND6</i>	15.20	11.92	83.23	79.62

Table S5. Codon usages of the mitochondrial PCGs of stylommatophoran species.

Codon	<i>A. vulgaris</i>		<i>A. rufus</i>		<i>A. fulica</i>		<i>A. fulgens</i>		<i>A. mustelina</i>		<i>A. sowerbyana</i>		<i>A. aubryana</i>		A. <i>diversifamilia</i>		<i>A. caerulea</i>	
	Count	RSCU	Count	RSCU	Count	RSCU	Count	RSCU	Count	RSCU	Count	RSCU	Count	RSCU	Count	RSCU	Count	RSCU
UUU(F)	226	1.64	215	1.55	218	1.48	345	1.93	358	1.95	354	1.90	263	1.67	269	1.74	241	1.81
UUC(F)	50	0.36	63	0.45	75	0.52	12	0.07	9	0.05	18	0.10	52	0.33	40	0.26	25	0.19
UUA(L)	389	3.89	347	3.64	215	2.31	456	4.59	472	4.77	465	4.77	321	3.23	363	3.70	341	3.38
UUG(L)	36	0.36	54	0.57	73	0.79	29	0.28	23	0.23	33	0.33	75	0.75	42	0.43	62	0.63
CUU(L)	85	0.85	71	0.74	96	1.04	77	0.78	74	0.75	68	0.69	90	0.90	96	0.98	83	0.83
CUC(L)	8	0.08	14	0.15	35	0.38	5	0.05	1	0.01	3	0.02	18	0.18	10	0.11	14	0.14
CUA(L)	71	0.71	73	0.77	104	1.12	27	0.27	18	0.18	16	0.16	74	0.74	68	0.69	79	0.79
CUG(L)	11	0.11	13	0.14	34	0.37	3	0.03	7	0.06	3	0.03	20	0.19	8	0.08	23	0.24
AUU(I)	230	1.74	218	1.71	192	1.49	335	1.94	326	1.96	309	1.91	229	1.79	231	1.82	267	1.76
AUC(I)	35	0.26	37	0.29	66	0.51	10	0.06	7	0.04	16	0.09	27	0.21	23	0.18	36	0.24
AUA(M)	213	1.66	213	1.61	157	1.44	285	1.83	282	1.87	289	1.83	215	1.75	237	1.79	213	1.62
AUG(M)	43	0.34	51	0.39	61	0.56	26	0.17	20	0.13	27	0.17	31	0.25	29	0.21	49	0.38
GUU(V)	71	1.17	79	1.24	124	1.50	74	1.48	81	1.64	86	1.63	105	1.48	112	1.64	90	1.57
GUC(V)	16	0.26	11	0.17	51	0.62	10	0.22	4	0.08	7	0.11	22	0.31	7	0.09	20	0.35
GUA(V)	110	1.82	117	1.84	99	1.20	108	2.16	105	2.15	107	2.03	116	1.63	124	1.80	98	1.69
GUG(V)	45	0.74	47	0.74	57	0.69	7	0.14	7	0.12	12	0.23	42	0.58	33	0.48	22	0.38
UCU(S)	99	2.13	84	1.84	86	1.69	108	2.38	96	1.97	92	1.95	55	1.33	73	1.71	77	1.79
UCC(S)	30	0.65	32	0.70	39	0.76	10	0.22	13	0.27	7	0.15	20	0.48	14	0.33	16	0.35
UCA(S)	52	1.12	58	1.27	98	1.90	64	1.41	81	1.67	83	1.76	59	1.40	65	1.52	68	1.58
UCG(S)	9	0.19	10	0.22	20	0.39	7	0.13	4	0.08	1	0.02	18	0.43	9	0.21	14	0.32
CCU(P)	54	1.16	46	1.01	55	1.55	68	2.31	70	2.37	72	2.44	66	1.87	66	1.80	49	1.48
CCC(P)	15	0.32	18	0.39	16	0.45	3	0.10	1	0.03	1	0.03	17	0.48	10	0.27	22	0.67
CCA(P)	72	1.55	77	1.68	53	1.49	43	1.46	44	1.49	43	1.46	51	1.45	60	1.63	49	1.52
CCG(P)	40	0.86	41	0.90	18	0.51	4	0.14	3	0.10	3	0.07	7	0.20	10	0.30	10	0.33
ACU(T)	80	1.71	70	1.43	62	1.41	73	1.97	74	1.90	73	1.96	87	1.72	94	1.98	77	1.66
ACC(T)	15	0.32	30	0.61	39	0.89	3	0.08	4	0.10	7	0.16	29	0.55	16	0.32	23	0.50



ACA(T)	82	1.75	83	1.69	60	1.36	66	1.78	73	1.87	68	1.80	61	1.21	70	1.49	73	1.58
ACG(T)	10	0.21	13	0.27	16	0.34	7	0.16	5	0.13	3	0.08	26	0.51	10	0.21	12	0.26
GCU(A)	88	1.91	83	1.83	85	1.94	73	2.21	74	2.21	72	2.17	85	1.73	78	1.72	91	1.89
GCC(A)	27	0.59	35	0.77	29	0.67	3	0.09	4	0.12	7	0.18	21	0.43	14	0.31	23	0.50
GCA(A)	55	1.20	44	0.97	43	0.99	51	1.55	52	1.55	49	1.50	64	1.32	72	1.59	66	1.37
GCG(A)	1	0.04	4	0.15	17	0.39	5	0.15	4	0.12	5	0.15	25	0.52	17	0.38	12	0.25
UAU(Y)	126	1.40	129	1.35	79	1.05	176	1.85	191	1.89	174	1.82	120	1.53	142	1.72	148	1.59
UAC(Y)	54	0.60	62	0.65	72	0.95	14	0.15	10	0.11	17	0.18	36	0.47	23	0.28	38	0.41
CAU(H)	42	1.15	62	1.59	42	1.14	60	1.82	62	1.85	64	1.88	57	1.58	61	1.69	61	1.65
CAC(H)	31	0.85	16	0.41	31	0.86	7	0.18	5	0.15	4	0.12	16	0.42	10	0.31	13	0.35
CAA(Q)	53	1.66	50	1.82	21	0.91	48	1.88	43	1.87	47	1.84	43	1.72	42	1.58	42	1.61
CAG(Q)	11	0.34	5	0.18	25	1.09	3	0.12	3	0.13	4	0.16	7	0.28	10	0.42	10	0.39
AAU(N)	106	1.64	113	1.61	56	1.29	146	1.87	150	1.84	142	1.80	98	1.50	118	1.71	107	1.54
AAC(N)	23	0.36	27	0.39	31	0.71	10	0.13	13	0.16	16	0.20	33	0.50	20	0.29	33	0.46
AAA(K)	62	1.51	62	1.53	48	1.25	109	1.74	118	1.82	105	1.75	77	1.67	81	1.65	73	1.70
AAG(K)	20	0.49	19	0.47	29	0.75	16	0.26	12	0.18	16	0.25	16	0.33	17	0.35	13	0.30
GAU(D)	49	1.66	45	1.58	46	1.44	49	1.75	48	2.00	52	1.89	47	1.57	55	1.74	47	1.52
GAC(D)	10	0.34	12	0.42	18	0.56	7	0.25	0	0.00	3	0.11	13	0.43	8	0.26	16	0.48
GAA(E)	56	1.42	45	1.18	33	0.85	66	1.69	72	1.92	65	1.78	48	1.37	43	1.19	48	1.32
GAG(E)	23	0.58	31	0.82	43	1.15	12	0.31	3	0.08	8	0.22	22	0.63	29	0.81	25	0.68
UGU(C)	31	1.55	40	1.70	46	1.30	39	1.90	42	1.91	39	1.77	36	1.57	39	1.59	34	1.55
UGC(C)	9	0.45	7	0.30	23	0.70	3	0.10	3	0.09	5	0.23	10	0.43	10	0.41	10	0.45
UGA(W)	69	1.48	72	1.50	48	1.07	77	1.88	75	1.79	72	1.71	60	1.36	70	1.57	68	1.55
UGG(W)	24	0.52	24	0.50	42	0.93	5	0.12	9	0.21	12	0.29	29	0.64	20	0.43	20	0.45
CGU(R)	25	1.82	28	1.81	16	1.07	14	1.17	16	1.25	12	0.96	26	1.58	20	1.27	17	1.39
CGC(R)	10	0.73	7	0.45	8	0.57	0	0.00	0	0.00	3	0.24	9	0.55	4	0.25	4	0.33
CGA(R)	17	1.24	21	1.35	18	1.29	33	2.67	29	2.42	34	2.72	20	1.21	25	1.59	22	1.80
CGG(R)	3	0.22	6	0.39	16	1.07	3	0.17	4	0.33	1	0.08	10	0.67	14	0.89	7	0.49
AGU(S)	71	2.09	58	1.76	33	0.65	52	1.15	69	1.42	55	1.15	75	1.81	73	1.71	49	1.14
AGC(S)	21	0.62	26	0.79	23	0.45	8	0.18	9	0.19	5	0.11	20	0.48	20	0.45	29	0.65

AGA(S)	34	1.00	31	0.94	51	1.00	101	2.25	109	2.24	114	2.44	72	1.71	64	1.50	64	1.48
AGG(S)	10	0.29	17	0.52	59	1.16	13	0.29	8	0.16	20	0.42	16	0.36	23	0.56	30	0.70
GGU(G)	71	1.30	76	1.41	74	1.17	81	1.58	85	1.71	74	1.51	90	1.48	75	1.23	73	1.33
GGC(G)	19	0.35	21	0.39	35	0.56	5	0.10	5	0.10	9	0.18	23	0.38	16	0.24	23	0.44
GGA(G)	70	1.28	76	1.41	48	0.76	99	1.95	101	2.05	95	1.94	81	1.31	98	1.57	79	1.44
GGG(G)	14	0.32	28	0.62	95	1.51	20	0.37	7	0.14	18	0.37	51	0.84	59	0.96	43	0.79

Table S5. (continued) Codon usages of the mitochondrial PCGs of stylommatophoran species.

Codon	<i>C. cicatricose</i>		<i>C. poyuensis</i>		<i>C. nemoralis</i>		<i>C. incanum</i>		<i>C. tridentatum</i>		<i>C. uva</i>		<i>C. virgata</i>		<i>C. aspersum</i>		<i>C. obtusus</i>	
	Count	RSCU	Count	RSCU	Count	RSCU	Count	RSCU	Count	RSCU	Count	RSCU	Count	RSCU	Count	RSCU	Count	RSCU
UUU(F)	251	1.79	250	1.79	216	1.47	198	1.51	224	1.68	195	1.52	251	1.67	278	1.72	265	1.66
UUC(F)	30	0.21	30	0.21	77	0.53	64	0.49	42	0.32	61	0.48	49	0.33	46	0.28	55	0.34
UUA(L)	374	3.83	384	3.89	164	1.66	263	2.77	265	2.74	226	2.37	335	3.04	332	3.48	269	2.47
UUG(L)	72	0.73	74	0.75	103	1.05	72	0.76	99	1.02	87	0.91	82	0.75	61	0.64	86	0.79
CUU(L)	61	0.62	61	0.62	101	1.04	87	0.92	94	0.96	88	0.93	88	0.80	96	1.01	124	1.14
CUC(L)	5	0.05	8	0.08	55	0.56	29	0.29	30	0.31	30	0.31	30	0.27	22	0.23	35	0.32
CUA(L)	62	0.65	49	0.51	104	1.06	94	0.99	69	0.71	101	1.06	101	0.91	55	0.58	107	0.98
CUG(L)	12	0.12	16	0.15	62	0.64	26	0.27	23	0.25	40	0.42	26	0.24	5	0.05	33	0.30
AUU(I)	243	1.75	224	1.75	143	1.53	231	1.63	202	1.58	179	1.45	209	1.74	246	1.88	178	1.71
AUC(I)	34	0.25	33	0.25	44	0.47	52	0.37	53	0.42	69	0.55	31	0.26	16	0.12	30	0.29
AUA(M)	186	1.63	192	1.67	107	1.28	189	1.45	178	1.47	157	1.37	159	1.51	182	1.78	101	1.29
AUG(M)	42	0.37	38	0.33	60	0.72	72	0.55	65	0.53	73	0.63	52	0.49	23	0.22	56	0.71
GUU(V)	112	1.57	117	1.55	101	1.44	100	1.36	120	1.49	127	1.48	125	1.55	120	1.95	112	1.49
GUC(V)	13	0.18	20	0.27	42	0.58	26	0.35	33	0.41	46	0.52	33	0.40	20	0.33	31	0.41
GUA(V)	117	1.64	124	1.63	85	1.19	121	1.65	95	1.19	103	1.19	116	1.44	87	1.41	78	1.04
GUG(V)	44	0.62	42	0.54	56	0.79	47	0.64	72	0.90	70	0.81	49	0.61	20	0.31	79	1.05
UCU(S)	66	1.60	65	1.58	62	1.34	101	1.96	105	2.11	96	1.88	59	1.29	99	2.19	68	1.52
UCC(S)	10	0.24	13	0.32	42	0.87	38	0.73	33	0.64	29	0.55	21	0.46	18	0.40	33	0.74
UCA(S)	55	1.34	56	1.36	34	0.73	86	1.65	64	1.28	72	1.39	62	1.39	55	1.22	36	0.80
UCG(S)	14	0.34	13	0.32	30	0.64	14	0.27	16	0.32	20	0.37	14	0.31	8	0.18	18	0.40
CCU(P)	59	1.70	64	1.83	56	1.53	68	1.88	60	1.59	70	1.75	73	1.91	75	2.08	59	1.51
CCC(P)	13	0.37	10	0.29	33	0.88	25	0.69	46	1.19	44	1.10	30	0.77	9	0.25	42	1.05
CCA(P)	48	1.38	49	1.43	33	0.90	46	1.24	34	0.90	38	0.95	33	0.86	46	1.28	31	0.79
CCG(P)	20	0.55	16	0.46	25	0.68	7	0.19	12	0.32	8	0.20	17	0.46	14	0.39	25	0.64
ACU(T)	86	1.63	77	1.49	64	1.42	82	1.84	95	2.17	62	1.51	78	1.57	88	1.70	60	1.19
ACC(T)	20	0.36	20	0.39	27	0.60	30	0.67	29	0.64	36	0.89	30	0.59	29	0.56	35	0.70
ACA(T)	79	1.50	81	1.55	59	1.29	49	1.12	38	0.87	49	1.20	70	1.42	70	1.34	60	1.19

ACG(T)	27	0.51	30	0.58	31	0.69	16	0.36	14	0.32	17	0.41	21	0.42	21	0.40	46	0.92
GCU(A)	96	2.13	87	1.90	96	1.44	98	2.05	94	1.80	78	1.51	90	1.81	90	2.07	82	1.43
GCC(A)	10	0.24	22	0.48	57	0.85	33	0.70	36	0.70	51	0.99	39	0.79	20	0.44	36	0.65
GCA(A)	51	1.13	49	1.07	72	1.06	47	0.99	51	0.99	57	1.10	51	1.03	49	1.15	55	0.96
GCG(A)	22	0.49	25	0.55	43	0.64	12	0.25	27	0.52	21	0.41	18	0.37	16	0.34	55	0.96
UAU(Y)	134	1.58	139	1.58	107	1.15	107	1.37	111	1.38	98	1.25	148	1.47	148	1.55	121	1.37
UAC(Y)	36	0.42	36	0.42	79	0.85	49	0.63	49	0.63	59	0.75	53	0.53	43	0.45	56	0.63
CAU(H)	61	1.39	68	1.53	51	1.17	55	1.38	55	1.36	36	0.97	49	1.33	61	1.58	44	1.14
CAC(H)	27	0.61	21	0.47	36	0.83	23	0.62	26	0.64	38	1.03	25	0.67	16	0.42	33	0.86
CAA(Q)	52	1.79	55	1.86	39	1.28	40	1.27	42	1.24	29	1.06	42	1.43	31	1.24	33	1.16
CAG(Q)	7	0.21	4	0.14	22	0.72	23	0.73	25	0.76	25	0.94	17	0.57	20	0.76	23	0.84
AAU(N)	109	1.58	111	1.59	49	1.16	66	1.47	68	1.48	55	1.09	75	1.27	99	1.56	57	1.27
AAC(N)	29	0.42	29	0.41	36	0.84	23	0.53	23	0.52	46	0.91	43	0.73	29	0.44	33	0.73
AAA(K)	65	1.67	70	1.69	35	1.01	57	1.37	53	1.25	42	1.04	49	1.25	55	1.50	35	1.03
AAG(K)	13	0.33	13	0.31	34	0.99	26	0.63	33	0.75	39	0.96	30	0.75	18	0.50	33	0.97
GAU(D)	48	1.60	51	1.67	29	0.97	33	1.14	46	1.32	39	1.15	46	1.42	53	1.63	36	1.16
GAC(D)	12	0.40	10	0.33	31	1.03	25	0.86	23	0.68	29	0.85	18	0.58	12	0.37	26	0.84
GAA(E)	48	1.35	43	1.25	30	1.09	48	1.12	33	0.84	35	0.95	55	1.33	49	1.53	34	1.01
GAG(E)	23	0.65	26	0.75	25	0.91	38	0.88	44	1.16	39	1.05	27	0.67	16	0.47	33	0.99
UGU(C)	33	1.49	40	1.63	46	1.14	33	1.31	29	1.41	21	1.08	38	1.49	52	1.46	36	1.47
UGC(C)	10	0.51	9	0.37	34	0.86	17	0.69	12	0.59	18	0.92	13	0.51	20	0.54	13	0.53
UGA(W)	57	1.34	57	1.30	57	1.25	64	1.33	62	1.33	55	1.09	49	1.16	73	1.62	47	1.15
UGG(W)	29	0.66	31	0.70	34	0.75	33	0.67	31	0.67	46	0.91	36	0.84	17	0.38	35	0.85
CGU(R)	22	1.33	20	1.18	31	1.32	16	1.14	17	1.15	16	1.03	26	1.55	33	2.00	30	1.69
CGC(R)	7	0.42	9	0.53	22	0.94	8	0.57	3	0.14	9	0.58	10	0.65	10	0.63	13	0.73
CGA(R)	22	1.33	18	1.06	18	0.77	22	1.57	18	1.22	22	1.42	14	0.84	17	1.06	12	0.68
CGG(R)	16	0.91	21	1.24	23	0.98	10	0.71	22	1.49	16	0.97	17	0.97	5	0.31	16	0.90
AGU(S)	78	1.90	82	1.99	69	1.47	33	0.63	38	0.76	33	0.63	77	1.70	85	1.88	78	1.74
AGC(S)	27	0.66	22	0.53	53	1.13	23	0.46	29	0.56	35	0.69	23	0.50	31	0.69	47	1.05
AGA(S)	46	1.09	36	0.90	49	1.05	72	1.38	59	1.18	68	1.31	53	1.17	43	0.95	27	0.60

AGG(S)	34	0.83	42	1.02	36	0.77	47	0.90	57	1.14	60	1.18	53	1.17	23	0.51	52	1.16
GGU(G)	94	1.67	88	1.59	74	1.38	60	1.07	55	0.97	49	0.90	101	1.60	107	2.06	81	1.29
GGC(G)	21	0.37	40	0.72	57	1.06	25	0.45	23	0.41	44	0.79	39	0.62	22	0.43	59	0.94
GGA(G)	47	0.84	42	0.74	40	0.74	87	1.55	62	1.11	36	0.65	43	0.69	59	1.13	21	0.34
GGG(G)	62	1.12	53	0.95	44	0.82	52	0.93	86	1.52	94	1.67	69	1.09	20	0.39	88	1.44

---

Table S5. (continued) Codon usages of the mitochondrial PCGs of stylommatophoran species.

Codon	<i>D. reticulatum</i>		<i>D. formosensis</i>		<i>G. cristata</i>		<i>H. itala</i>		<i>H. pomatia</i>		<i>M. kiangsinensis</i>		<i>M. bilineatum</i>		<i>M. pontificus</i>		<i>N. nux</i>	
	Count	RSCU	Count	RSCU	Count	RSCU	Count	RSCU	Count	RSCU	Count	RSCU	Count	RSCU	Count	RSCU	Count	RSCU
UUU(F)	289	1.82	291	1.82	302	1.80	235	1.68	263	1.71	264	1.77	261	1.81	295	1.88	288	1.73
UUC(F)	29	0.18	29	0.18	34	0.20	44	0.32	44	0.29	34	0.23	27	0.19	20	0.12	45	0.27
UUA(L)	393	4.07	335	3.33	341	3.32	309	2.95	308	3.00	326	3.25	352	3.48	324	3.51	386	3.80
UUG(L)	68	0.69	78	0.77	69	0.67	92	0.88	75	0.73	83	0.83	57	0.56	78	0.84	57	0.57
CUU(L)	64	0.66	125	1.24	122	1.19	85	0.81	90	0.88	88	0.88	118	1.16	83	0.90	85	0.85
CUC(L)	5	0.05	10	0.11	16	0.16	21	0.20	36	0.35	20	0.19	14	0.15	7	0.06	14	0.14
CUA(L)	42	0.42	48	0.48	60	0.58	95	0.91	87	0.85	72	0.71	60	0.59	51	0.55	55	0.54
CUG(L)	10	0.10	7	0.07	9	0.09	26	0.25	21	0.20	14	0.14	7	0.06	12	0.13	10	0.10
AUU(I)	274	1.88	215	1.85	255	1.86	202	1.71	176	1.64	199	1.78	216	1.77	270	1.79	302	1.82
AUC(I)	18	0.12	17	0.15	20	0.14	34	0.29	39	0.36	25	0.22	27	0.23	31	0.21	29	0.18
AUA(M)	189	1.62	185	1.60	134	1.52	159	1.51	166	1.58	166	1.58	156	1.68	230	1.67	200	1.78
AUG(M)	44	0.38	46	0.40	42	0.48	51	0.49	44	0.42	46	0.42	30	0.32	46	0.33	25	0.22
GUU(V)	135	1.86	140	1.89	112	1.77	121	1.57	120	1.71	159	2.06	104	1.86	111	1.59	84	1.46
GUC(V)	13	0.18	16	0.20	20	0.32	29	0.38	23	0.33	16	0.19	9	0.17	10	0.16	8	0.14
GUA(V)	112	1.54	107	1.43	87	1.38	105	1.36	88	1.26	81	1.05	88	1.58	112	1.62	118	2.03
GUG(V)	30	0.41	35	0.47	34	0.54	53	0.69	49	0.70	55	0.70	22	0.40	43	0.62	21	0.37
UCU(S)	88	1.98	114	2.52	88	1.84	72	1.60	62	1.42	73	1.77	99	2.39	104	2.17	119	2.31
UCC(S)	16	0.36	12	0.26	20	0.40	26	0.58	20	0.43	20	0.46	17	0.41	14	0.29	20	0.37
UCA(S)	60	1.33	49	1.10	49	1.04	52	1.17	52	1.17	48	1.16	56	1.35	75	1.59	62	1.18
UCG(S)	10	0.24	7	0.13	10	0.23	7	0.16	23	0.54	20	0.46	8	0.18	22	0.46	8	0.17
CCU(P)	72	2.09	75	2.25	68	1.93	68	1.89	62	1.88	68	1.88	74	2.27	64	1.91	90	2.52
CCC(P)	14	0.41	5	0.15	22	0.63	30	0.83	16	0.48	16	0.41	20	0.60	18	0.54	14	0.39
CCA(P)	46	1.30	44	1.30	42	1.21	36	1.03	44	1.33	53	1.46	34	1.02	49	1.49	34	0.92
CCG(P)	7	0.20	10	0.30	8	0.23	9	0.25	10	0.30	9	0.25	4	0.11	3	0.06	6	0.17
ACU(T)	69	1.76	100	2.31	69	1.70	74	1.64	86	1.74	72	1.54	79	1.97	72	1.91	85	2.24
ACC(T)	14	0.36	18	0.42	18	0.44	23	0.51	25	0.51	25	0.53	18	0.46	18	0.48	14	0.37

ACA(T)	68	1.73	47	1.09	62	1.53	66	1.46	66	1.33	64	1.37	53	1.33	52	1.40	45	1.18
ACG(T)	7	0.15	8	0.18	13	0.32	18	0.40	21	0.42	26	0.56	10	0.25	8	0.21	8	0.21
GCU(A)	107	2.38	122	2.57	109	2.04	72	1.48	94	1.93	62	1.28	86	2.11	85	2.27	104	2.38
GCC(A)	20	0.42	16	0.34	36	0.69	40	0.83	18	0.37	64	1.32	21	0.50	8	0.21	15	0.37
GCA(A)	36	0.82	44	0.93	56	1.05	55	1.15	56	1.15	44	0.91	51	1.23	48	1.28	39	0.90
GCG(A)	17	0.38	8	0.17	12	0.22	26	0.54	27	0.55	23	0.49	7	0.16	9	0.24	15	0.35
UAU(Y)	113	1.52	140	1.59	109	1.48	133	1.44	148	1.53	138	1.60	129	1.67	127	1.57	146	1.73
UAC(Y)	36	0.48	36	0.41	38	0.52	52	0.56	46	0.47	35	0.40	25	0.33	35	0.43	22	0.27
CAU(H)	59	1.71	49	1.49	52	1.53	49	1.39	55	1.33	55	1.46	65	1.65	66	1.78	64	1.73
CAC(H)	10	0.29	17	0.51	16	0.47	22	0.61	29	0.67	20	0.54	14	0.35	8	0.22	10	0.27
CAA(Q)	42	1.35	38	1.49	43	1.59	48	1.55	39	1.59	44	1.69	49	1.77	53	1.63	41	1.71
CAG(Q)	20	0.65	13	0.51	10	0.41	14	0.45	10	0.41	8	0.31	7	0.23	12	0.37	7	0.29
AAU(N)	103	1.75	103	1.72	90	1.49	86	1.31	88	1.52	100	1.65	112	1.75	103	1.63	129	1.71
AAC(N)	16	0.25	17	0.28	31	0.51	46	0.69	29	0.48	21	0.35	16	0.25	23	0.37	22	0.29
AAA(K)	72	1.71	60	1.38	81	1.78	55	1.48	44	1.21	66	1.47	79	1.60	81	1.64	92	1.84
AAG(K)	12	0.29	27	0.62	10	0.22	20	0.52	29	0.79	23	0.53	20	0.40	18	0.36	8	0.16
GAU(D)	52	1.73	52	1.70	51	1.65	33	1.10	42	1.39	46	1.48	40	1.63	46	1.48	48	1.65
GAC(D)	8	0.27	9	0.30	10	0.35	26	0.90	18	0.61	16	0.52	9	0.37	16	0.52	10	0.35
GAA(E)	55	1.33	49	1.35	59	1.51	47	1.25	42	1.24	49	1.34	51	1.53	48	1.07	64	1.52
GAG(E)	27	0.67	23	0.65	20	0.49	29	0.75	26	0.76	23	0.66	16	0.47	42	0.93	20	0.48
UGU(C)	35	1.63	40	1.67	29	1.37	46	1.64	42	1.35	44	1.80	36	1.66	39	1.70	35	1.79
UGC(C)	8	0.37	8	0.33	13	0.63	10	0.36	20	0.65	5	0.20	8	0.34	7	0.30	4	0.21
UGA(W)	69	1.38	65	1.41	65	1.57	49	1.22	57	1.31	56	1.27	65	1.52	77	1.57	69	1.51
UGG(W)	31	0.62	27	0.59	18	0.43	33	0.78	30	0.69	33	0.73	21	0.48	21	0.43	22	0.49
CGU(R)	16	1.21	33	2.17	9	0.71	20	1.29	23	1.50	21	1.33	17	1.47	16	1.18	15	1.15
CGC(R)	1	0.08	1	0.07	5	0.39	10	0.65	10	0.69	7	0.38	5	0.49	3	0.16	1	0.08
CGA(R)	26	1.96	16	1.08	27	2.12	17	1.10	20	1.19	18	1.14	18	1.63	27	2.12	31	2.38
CGG(R)	10	0.75	10	0.68	10	0.78	16	0.97	10	0.63	18	1.14	5	0.41	7	0.55	6	0.38
AGU(S)	65	1.44	81	1.75	66	1.38	75	1.71	74	1.67	61	1.48	49	1.19	39	0.81	34	0.66
AGC(S)	7	0.13	12	0.26	16	0.33	29	0.63	46	1.04	21	0.51	18	0.43	12	0.25	13	0.25

AGA(S)	73	1.62	59	1.29	90	1.88	53	1.19	42	0.95	59	1.43	59	1.42	74	1.55	111	2.13
AGG(S)	40	0.89	31	0.68	43	0.90	43	0.97	35	0.79	30	0.73	26	0.63	42	0.88	48	0.93
GGU(G)	91	1.39	112	1.98	70	1.17	96	1.54	109	1.98	78	1.21	77	1.54	49	1.00	46	0.89
GGC(G)	12	0.18	16	0.27	20	0.32	29	0.45	30	0.55	34	0.53	10	0.20	7	0.14	15	0.29
GGA(G)	69	1.06	48	0.85	75	1.26	55	0.88	43	0.78	72	1.12	69	1.38	83	1.65	104	2.00
GGG(G)	88	1.36	51	0.90	75	1.26	70	1.12	38	0.69	73	1.14	44	0.87	61	1.21	42	0.82

---



Table S5. (continued) Codon usages of the mitochondrial PCGs of stylommatophoran species.

Codon	<i>O. dolium</i>		<i>P. redfieldi</i>		<i>P. bilineatus</i>		<i>P. cereolus</i>		<i>P. mexicana</i>		<i>P. muscorum</i>		<i>S. putris</i>		<i>V. pusilla</i>	
	Count	RSCU	Count	RSCU	Count	RSCU	Count	RSCU	Count	RSCU	Count	RSCU	Count	RSCU	Count	RSCU
UUU(F)	246	1.59	346	1.93	274	1.72	278	1.72	281	1.71	339	1.76	381	1.88	330	1.86
UUC(F)	62	0.41	12	0.07	44	0.28	46	0.28	48	0.29	46	0.24	25	0.12	25	0.14
UUA(L)	256	2.48	439	4.52	387	3.68	361	3.30	358	3.29	354	3.62	398	4.32	397	4.01
UUG(L)	83	0.80	22	0.23	55	0.52	109	1.00	111	1.01	42	0.42	44	0.48	49	0.49
CUU(L)	107	1.04	75	0.77	111	1.05	95	0.87	94	0.87	96	0.98	62	0.68	77	0.78
CUC(L)	36	0.35	1	0.01	5	0.05	16	0.14	14	0.13	10	0.11	5	0.05	18	0.18
CUA(L)	94	0.90	44	0.45	68	0.64	60	0.55	60	0.55	77	0.79	38	0.41	49	0.49
CUG(L)	44	0.43	3	0.02	7	0.06	16	0.15	17	0.16	7	0.07	5	0.05	4	0.04
AUU(I)	195	1.52	311	1.92	239	1.84	205	1.77	205	1.77	257	1.77	359	1.93	264	1.83
AUC(I)	62	0.48	13	0.08	21	0.16	26	0.23	27	0.23	34	0.23	13	0.07	25	0.17
AUA(M)	155	1.44	278	1.84	173	1.68	127	1.49	127	1.45	176	1.67	222	1.78	168	1.73
AUG(M)	60	0.56	23	0.16	33	0.32	44	0.51	48	0.55	35	0.33	29	0.22	26	0.27
GUU(V)	88	1.30	86	1.58	94	1.56	146	2.11	142	2.04	111	1.86	94	2.15	130	2.11
GUC(V)	31	0.46	9	0.17	10	0.18	14	0.20	18	0.26	20	0.34	10	0.25	13	0.21
GUA(V)	81	1.20	111	2.02	107	1.80	83	1.20	87	1.25	88	1.50	65	1.49	88	1.43
GUG(V)	72	1.05	13	0.24	27	0.45	34	0.49	33	0.46	18	0.30	5	0.11	16	0.26
UCU(S)	101	1.99	112	2.44	111	2.57	73	1.64	72	1.62	81	1.66	129	2.68	91	1.79
UCC(S)	31	0.60	8	0.17	16	0.35	30	0.67	30	0.68	16	0.33	20	0.39	14	0.28
UCA(S)	55	1.07	74	1.61	55	1.26	26	0.58	25	0.56	82	1.70	64	1.33	78	1.54
UCG(S)	20	0.39	3	0.04	8	0.19	5	0.11	5	0.11	13	0.27	10	0.21	5	0.10
CCU(P)	60	1.90	70	2.39	87	2.47	65	2.05	68	2.13	85	2.49	81	2.79	55	1.79
CCC(P)	29	0.89	3	0.10	10	0.28	16	0.50	16	0.47	12	0.36	8	0.28	8	0.26
CCA(P)	23	0.73	40	1.37	42	1.16	29	0.91	29	0.87	33	0.95	26	0.90	56	1.82
CCG(P)	16	0.48	4	0.14	3	0.09	17	0.54	17	0.53	7	0.21	1	0.03	4	0.13
ACU(T)	68	1.66	72	1.87	88	1.91	88	2.06	85	2.00	68	1.92	86	2.34	75	2.04
ACC(T)	29	0.72	10	0.26	21	0.45	33	0.77	34	0.80	18	0.51	9	0.24	9	0.24

ACA(T)	57	1.42	68	1.76	73	1.57	33	0.77	33	0.78	49	1.38	49	1.36	59	1.58
ACG(T)	8	0.20	4	0.11	3	0.06	17	0.40	18	0.42	7	0.20	3	0.05	5	0.13
GCU(A)	95	1.84	70	2.09	94	2.01	98	2.34	95	2.30	87	2.00	79	2.47	105	2.13
GCC(A)	46	0.89	8	0.24	21	0.45	31	0.75	34	0.82	23	0.55	10	0.34	25	0.51
GCA(A)	44	0.85	52	1.55	62	1.35	23	0.55	21	0.51	55	1.26	36	1.16	56	1.14
GCG(A)	22	0.43	4	0.12	9	0.19	16	0.36	16	0.36	8	0.18	1	0.03	10	0.22
UAU(Y)	95	1.16	178	1.85	139	1.73	152	1.65	152	1.66	107	1.47	156	1.76	107	1.53
UAC(Y)	69	0.84	14	0.15	22	0.27	33	0.35	31	0.34	38	0.53	21	0.24	33	0.47
CAU(H)	44	1.24	65	1.86	65	1.59	56	1.49	57	1.50	52	1.55	55	1.71	48	1.50
CAC(H)	27	0.76	5	0.14	17	0.41	20	0.51	20	0.50	16	0.45	9	0.29	16	0.50
CAA(Q)	36	1.31	46	1.76	59	1.90	59	1.73	56	1.70	43	1.76	44	1.80	46	1.61
CAG(Q)	20	0.69	7	0.24	3	0.10	9	0.27	10	0.30	7	0.24	5	0.20	10	0.39
AAU(N)	86	1.33	137	1.77	114	1.67	101	1.70	98	1.68	98	1.53	137	1.72	94	1.56
AAC(N)	43	0.67	18	0.23	23	0.33	18	0.30	20	0.32	30	0.47	22	0.28	26	0.44
AAA(K)	94	1.66	100	1.65	85	1.71	62	1.59	62	1.57	85	1.52	82	1.67	82	1.73
AAG(K)	20	0.34	21	0.35	14	0.29	16	0.41	17	0.43	27	0.48	16	0.33	13	0.27
GAU(D)	40	1.18	48	1.81	43	1.56	43	1.62	42	1.62	43	1.56	56	1.84	48	1.57
GAC(D)	29	0.82	5	0.19	12	0.44	10	0.38	10	0.38	12	0.44	5	0.16	13	0.43
GAA(E)	55	1.46	69	1.82	52	1.51	53	1.58	53	1.58	59	1.38	66	1.63	68	1.61
GAG(E)	20	0.54	7	0.18	17	0.49	14	0.42	14	0.42	26	0.62	16	0.37	16	0.39
UGU(C)	36	1.36	40	1.90	42	1.83	39	1.59	36	1.54	35	1.63	38	1.81	34	1.55
UGC(C)	17	0.64	3	0.10	4	0.17	10	0.41	10	0.46	8	0.37	4	0.19	10	0.45
UGA(W)	57	1.30	77	1.86	72	1.54	55	1.10	53	1.07	62	1.44	65	1.59	79	1.65
UGG(W)	31	0.70	7	0.14	21	0.46	44	0.90	46	0.93	23	0.56	17	0.41	17	0.35
CGU(R)	7	0.54	13	1.06	23	1.84	21	1.35	22	1.42	9	0.67	13	1.02	16	1.15
CGC(R)	8	0.62	0	0.00	4	0.32	3	0.19	3	0.19	3	0.22	0	0.00	3	0.15
CGA(R)	23	1.77	31	2.53	16	1.28	12	0.77	12	0.77	29	2.15	31	2.43	27	2.08
CGG(R)	14	1.08	5	0.41	7	0.56	26	1.68	25	1.61	13	0.96	7	0.55	8	0.62
AGU(S)	49	0.95	59	1.29	56	1.31	98	2.20	99	2.23	36	0.77	59	1.23	48	0.95
AGC(S)	36	0.70	4	0.09	12	0.28	23	0.54	23	0.52	25	0.52	12	0.25	20	0.37

AGA(S)	75	1.46	90	1.96	61	1.42	57	1.28	56	1.26	100	2.08	83	1.72	116	2.29
AGG(S)	43	0.84	18	0.39	27	0.63	44	0.99	46	1.01	33	0.66	9	0.19	35	0.69
GGU(G)	62	1.08	75	1.48	85	1.66	113	1.81	114	1.84	47	0.87	74	1.66	73	1.28
GGC(G)	36	0.62	7	0.14	7	0.12	26	0.42	27	0.43	20	0.35	7	0.16	12	0.21
GGA(G)	62	1.08	98	1.93	75	1.50	49	0.78	49	0.78	81	1.47	78	1.75	100	1.75
GGG(G)	72	1.22	23	0.45	36	0.73	62	0.99	59	0.94	72	1.31	20	0.43	43	0.75

---

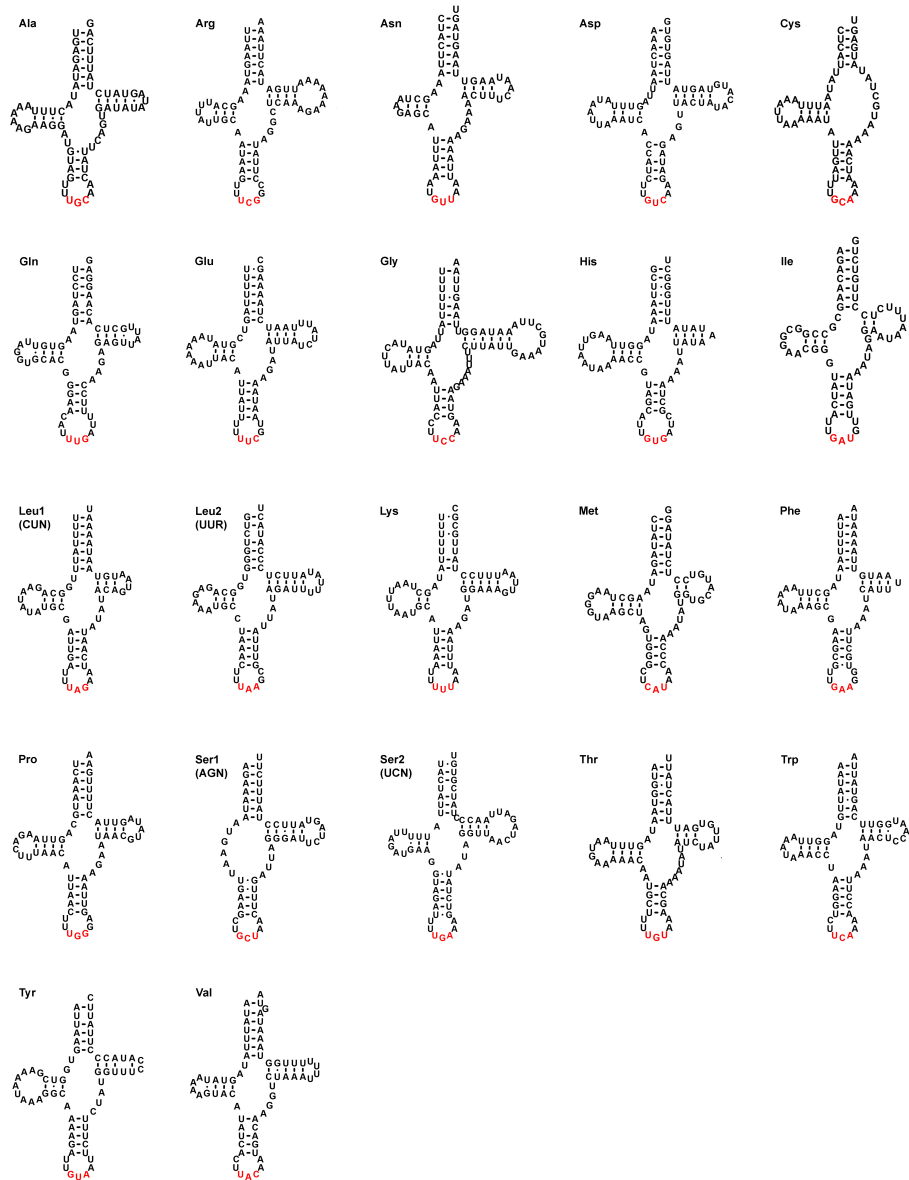


Figure S1. Predicted secondary structures for the 22 typical tRNA genes of *A. vulgaris* mitogenome. Watson–Crick pairs are indicated by lines and wobble GU pairs are indicated by dots.

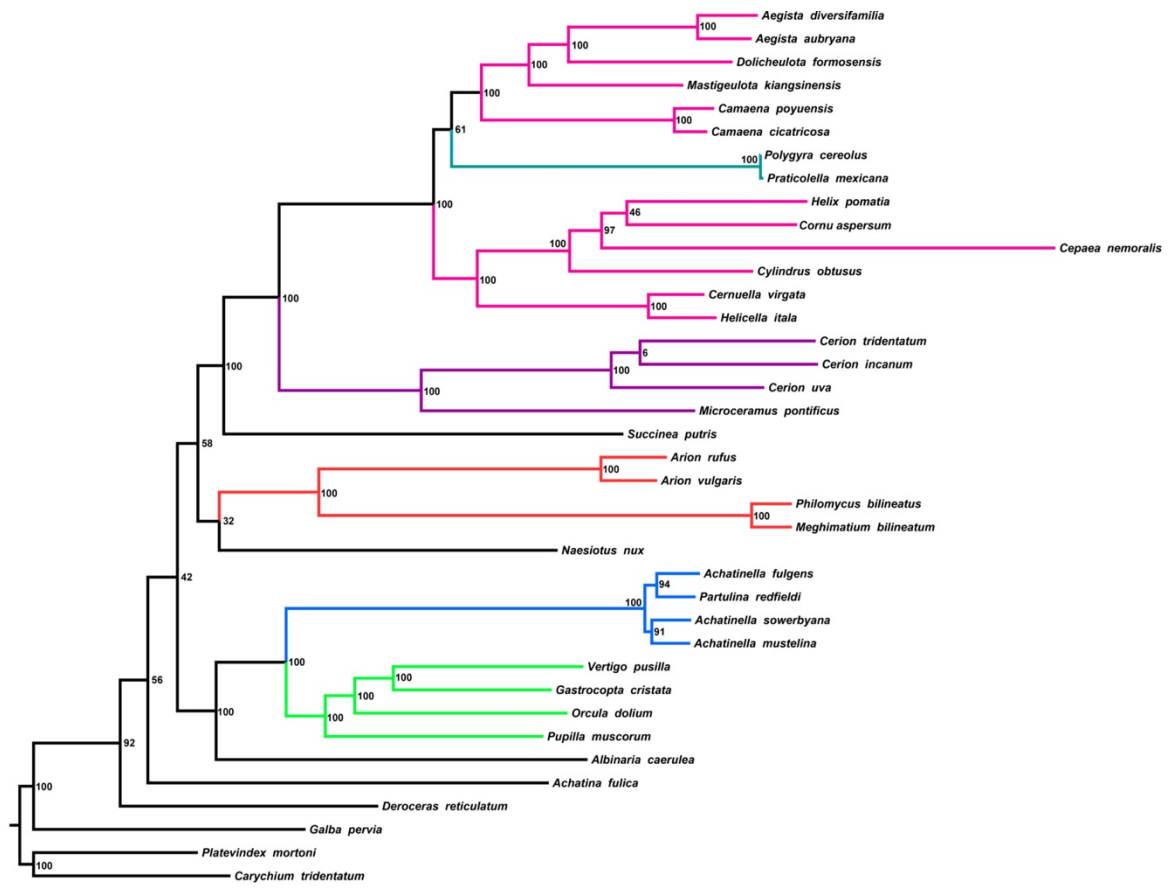


Figure S2. The phylogenetic tree constructed under ML using the dataset 8P12RNA.

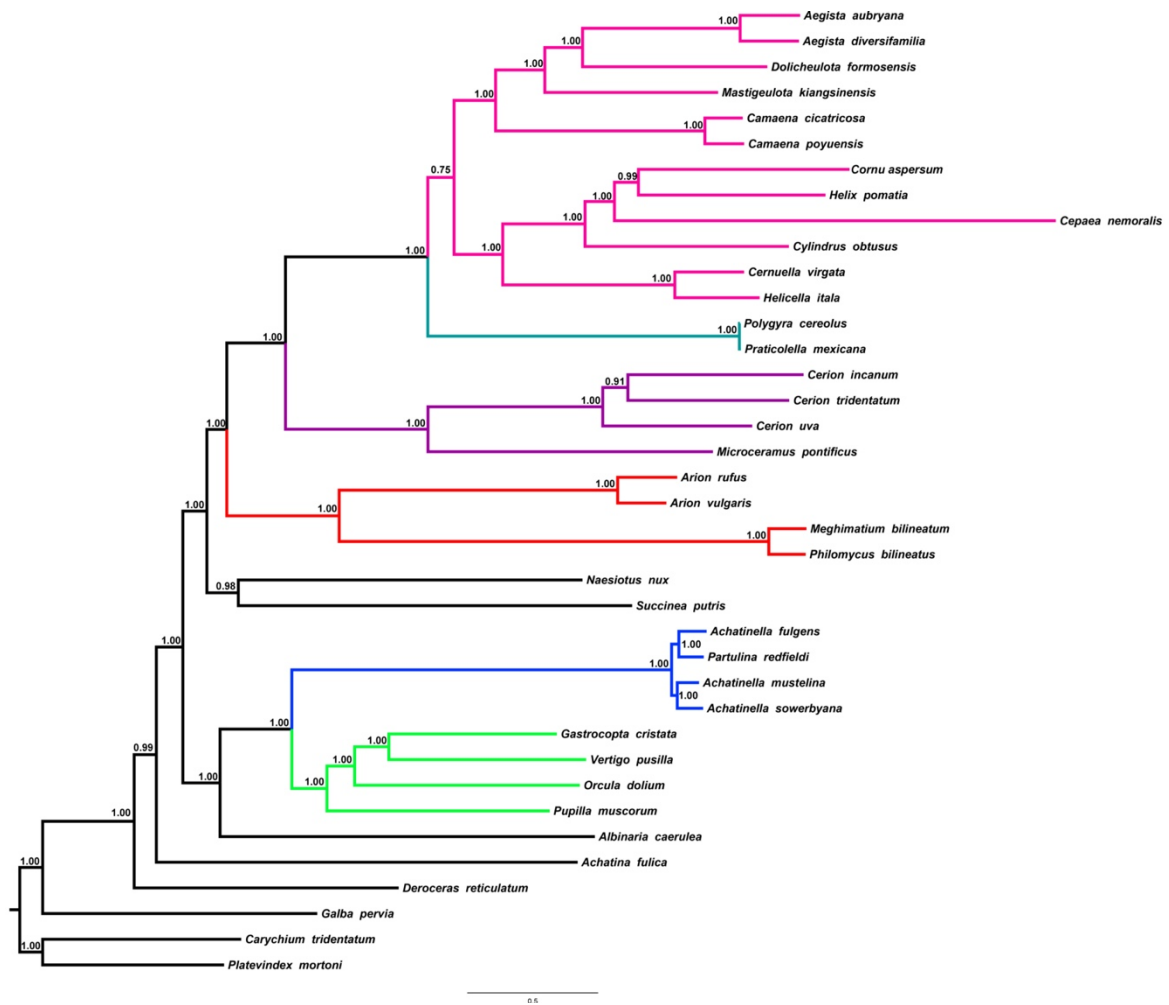


Figure S3. The phylogenetic tree constructed under BI using the dataset P123RNA.

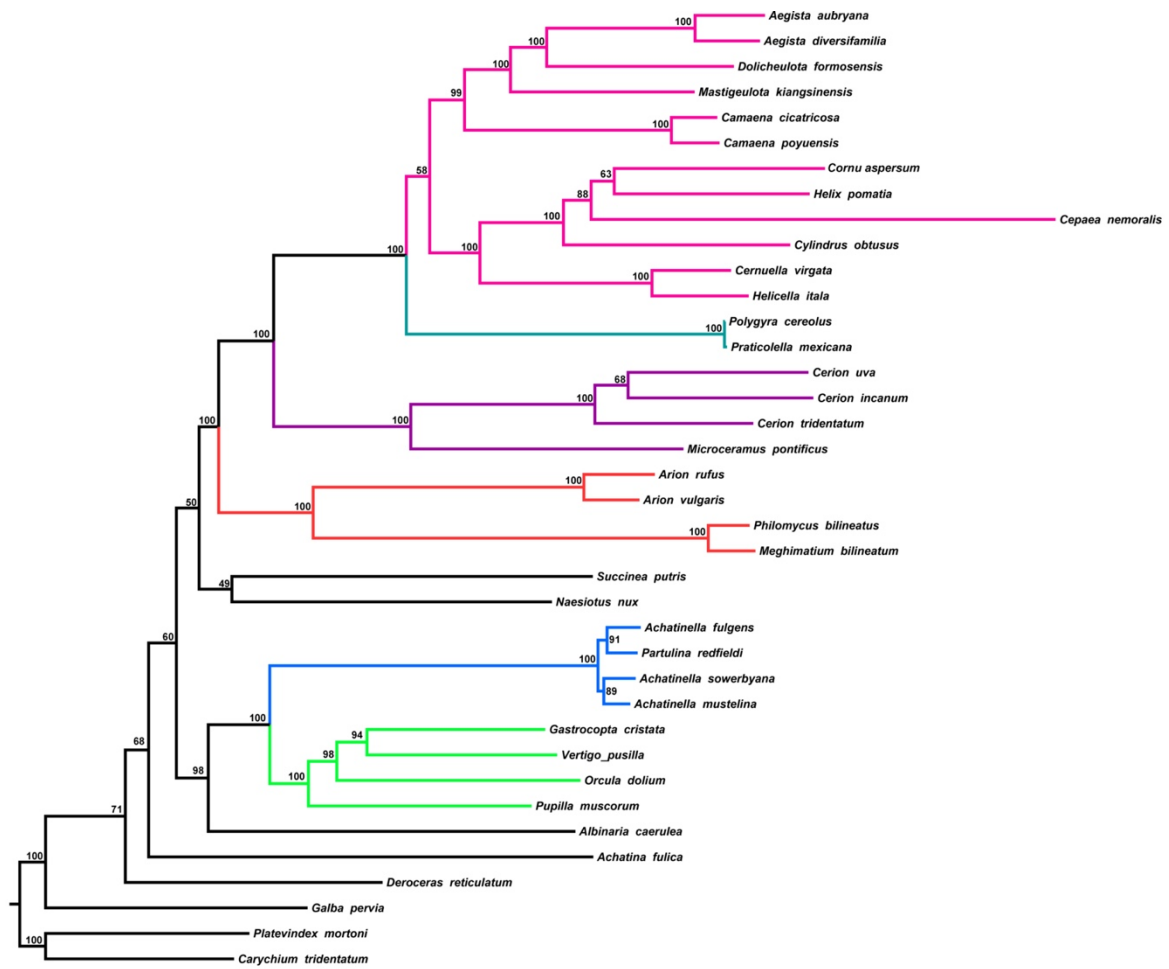


Figure S4. The phylogenetic tree constructed under ML using the dataset P123RNA

### 3.2 Publication II

**Zeyuan Chen**, Michael Schrödl (2022): How many single-copy orthologous genes from whole genomes reveal deep gastropod relationships? PeerJ 10:e13285

A pdf of the article is available at:

<http://doi.org/10.7717/peerj.8603>

This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, reproduction and adaptation in any medium and for any purpose provided that it is properly attributed. For attribution, the original author(s), title, publication source (PeerJ) and either DOI or URL of the article must be cited.



# How many single-copy orthologous genes from whole genomes reveal deep gastropod relationships?

Zeyuan Chen<sup>1,2</sup> and Michael Schrödl<sup>1,2,3</sup>

<sup>1</sup> Mollusca, SNSB-Bavarian State Collection of Zoology, Munich, Bavaria, Germany

<sup>2</sup> Department Biology II, Ludwig-Maximilians-Universität München, Munich, Bavaria, Germany

<sup>3</sup> GeoBio-Center LMU, Munich, Bavaria, Germany

## ABSTRACT

The Gastropoda contains 80% of existing mollusks and is the most diverse animal class second only to the Insecta. However, the deep phylogeny of gastropods has been controversial for a long time. Especially the position of Patellogastropoda is a major uncertainty. Morphology and some mitochondria studies concluded that Patellogastropoda is likely to be sister to all other gastropods (Orthogastropoda hypothesis), while transcriptomic and other mitogenomic studies indicated that Patellogastropoda and Vetigastropoda are sister taxa (Psilogastropoda). With the release of high-quality genomes, orthologous genes can be better identified and serve as powerful candidates for phylogenetic analysis. The question is, given the current limitations on the taxon sampling side, how many markers are needed to provide robust results. Here, we identified single-copy orthologous genes (SOGs) from 14 gastropods species with whole genomes available which cover five main gastropod subclasses. We generated different datasets from 395 to 1610 SOGs by allowing species missing in different levels. We constructed gene trees of each SOG, and inferred species trees from different collections of gene trees. We found as the number of SOGs increased, the inferred topology changed from Patellogastropoda being sister to all other gastropods to Patellogastropoda being sister to Vetigastropoda + Neomphalina (Psilogastropoda *s.l.*), with considerable support. Our study thus rejects the Orthogastropoda concept showing that the selection of the representative species and use of sufficient informative sites greatly influence the analysis of deep gastropod phylogeny.

**Subjects** Evolutionary Studies, Molecular Biology, Taxonomy

**Keywords** Phylogeny, Whole genomes, Gene tree, Mollusca, Patellogastropoda, Conflicting topologies, Orthogastropoda

## INTRODUCTION

Commonly known as snails and slugs, gastropods are the most common and relevant class within the phylum Mollusca. The earliest undisputed gastropods date from the Late Cambrian Period, around 500 million years ago, and now the total number of existing gastropod species is estimated to range from 63,000 to more than 100,000 (Bieler, 1992; Bouchet *et al.*, 2017), accounting for 80% of known mollusks. Gastropods have a worldwide distribution, from the near Arctic and Antarctic zones to the tropics (Crame, 2013) and

Submitted 19 January 2022

Accepted 28 March 2022

Published 18 April 2022

Corresponding author

Zeyuan Chen, Chen.Z@snsb.de

Academic editor

Kamil Steczkiewicz

Additional Information and  
Declarations can be found on  
page 9

DOI 10.7717/peerj.13285

© Copyright

2022 Chen and Schrödl

Distributed under

Creative Commons CC-BY 4.0

**OPEN ACCESS**

**How to cite this article** Chen Z, Schrödl M. 2022. How many single-copy orthologous genes from whole genomes reveal deep gastropod relationships?. PeerJ 10:e13285 DOI 10.7717/peerj.13285

ubiquitous in the oceans. They have colonized even extreme environments, such as the deep sea (e.g., faunas associated with hydrothermal vents) (Govenar, Fisher & Shank, 2015) and deserts (Greve et al., 2017). Gastropods are highly diverse with regard to their body forms, shells, and functions, and in many ways, they affect human life. Some gastropod species (such as conch, abalone, limpets, and whelks) are used as food, the shells of some species were used as ornaments or in making jewellery (Frýda, 2013). A couple of gastropod species are intermediate hosts of parasites and transmit human or animal diseases, such as schistosomiasis (Fried & Huffman, 2017). Several land and freshwater snails and slugs are highly invasive and competitive species, feeding on crops and vegetables and causing threats to local ecology and economy, e.g., the notorious Spanish slug, *Arion vulgaris* Moquin-Tandon, 1855 (Zemanova, Knop & Heckel, 2017), golden apple snail, *Pomacea canaliculata* Lamarck, 1822, etc. (Global Invasive Species Database, 2021). All these characteristics: long evolutionary history, wide adaptability, high diversity, potential invasiveness, competitiveness, and high relevance for human lives, together with their long and rich fossil record (Frýda, 2013), make gastropods a unique animal group for evolutionary, ecological, and biogeographical investigations. Some of the most relevant questions concern the transition from ocean to freshwater to terrestrial lifestyle, the rapid adaptation in a new environment, the radiation evolution, the changes of shell morphology such as towards limpets or its reduction, internalization or loss (Solem, 2020).

The basis of all the above issues relies on resolving the gastropod evolutionary history, and there is great progress in subgroups, e.g., in Vetigastropoda (Cunha et al., 2019) or within Heterobranchia (Brenzinger, Schrödl & Kano, 2021; Kano et al., 2016). However, the early gastropod phylogeny, i.e., the relationship between major gastropod groups, is still controversial. The extant Gastropoda are divided into seven main lineages, namely Patellogastropoda, Neomphalina, Cocculiniformia, Vetigastropoda, Neritimorpha, Caenogastropoda, and Heterobranchia. Usually, Heterobranchia and Caenogastropoda were recovered as sister groups (forming the clade Apogastropoda), and a close relationship of Neomphalina, Cocculiniformia, and Vetigastropoda has been supported (Lee et al., 2019). The main uncertainty refers to the position of Patellogastropoda (true limpets) as different data and analysis methods led to different conclusions (Zapata et al., 2014). Based on morphological data, Patellogastropoda has been recognized as the earliest-branching gastropod group, i.e., being the sister of all the remaining lineages together forming the clade Orthogastropoda (Golikov & Starobogatov, 1975; Haszprunar, 1988; Lindberg, 1988). The Orthogastropoda-Patellogastropoda topology has also been supported by Kocot et al. (2011) using 308 genes from transcriptome data. However, an extended sampling of transcriptomic markers favored a sister group relationship of Patellogastropoda and Vetigastropoda (combined clade Psilogastropoda), thus rejecting the monophyly of Orthogastropoda (Cunha & Giribet, 2019; Smith et al., 2011; Zapata et al., 2014). Early phylogenies based on complete mitochondrial genomes consistently recovered Patellogastropoda sister to Heterobranchia (Arquez, Colgan & Castro, 2014; Grande, Templado & Zardoya, 2008; Osca et al., 2014; Uribe et al., 2016), which was suspected to be a long branch artefact (Schrödl & Stöger, 2014; Stöger & Schrödl, 2013).

More recently, a deep gastropod split into Patellogastropoda and Orthogastropoda was recovered by expanding the Patellogastropoda sampling (Uribe *et al.*, 2019), however, this was still based on a limited number of relatively fast-evolving mitochondrial genes.

Ultimately, the increasing number of published high-quality gastropod whole nuclear genomes raised the possibility of identifying large numbers of putative orthologs across multiple species (Gomes-dos-Santos *et al.*, 2019). Orthologs are defined as genes sharing a common ancestor by speciation, in contrast to paralogs, which are duplicated copies arising through polyploidization or duplications (Gogarten & Olendzenski, 1999; Sonnhammer & Koonin, 2002). Orthologous genes, if in a single copy status, are so-called single-copy orthologous genes (SOGs), which imply that they have kept this status since the species' last common ancestor, when not considered rare cases, such as the differential loss of paralogs after whole-genome duplication or orthologous gene displacement (Creevey *et al.*, 2011). SOGs thus hold great information potential for phylogenetic reconstruction, especially where universal markers are not able to generate strong phylogenetic hypotheses (Sang, 2002; Wu *et al.*, 2006). Several newly published gastropod genome studies have tried to reconstruct gastropod phylogeny using hundreds of SOGs (Table 1); Lan *et al.* (2021) and Sun *et al.* (2020) recovered Psilogastropoda, while Chen *et al.* (2020) inferred Orthogastropoda (Table S1). Therefore, the conflict still remains when using SOGs in different taxon and gene sets with different species coverage. To explain the foundations leading to inconsistent topologies, we re-analyzed and expanded Chen *et al.*'s (2020) study. In our new analyses, establishing sets of SOGs we allowed species missing per SOG successively to a certain degree (0–20% of all species) and reconstructed the phylogenies. We found that as the number of species missing per SOG increases, the number of orthologous genes discovered also increases, and the final topology changed from Patellogastropoda-Orthogastropoda to Patellogastropoda sister to a combined clade of Vetigastropoda and Neomphalina (Psilogastropoda in a broader sense) with significantly increased support values.

## MATERIALS AND METHODS

### Identification of gastropod SOGs sets

Species were selected as in the *A. vulgaris* genome study (Chen *et al.*, 2020). These species include bivalve species, *Argopecten purpuratus* Lamarck, 1819 (Li *et al.*, 2018) and *Saccostrea glomerata* Gould, 1850 (Powell *et al.*, 2018) as the outgroup. The 14 gastropods cover five (of six) subclasses, specifically, 7 Heterobranchia: *A. vulgaris*, *Achatina fulica* Ferussac, 1821, *Ac. immaculata* Lamarck, 1822, *Biomphalaria glabrata* Say, 1818, *Radix auricularia* Linnaeus, 1758, *Aplysia californica* J. G. Cooper, 1863, *Elysia chlorotica* Gould, 1870, 4 Caenogastropoda: *P. canaliculata*, *Marisa cornuarietis* Linnaeus, 1758, *Lanistes nyassanus* Dohrn, 1865, *Conus consors* G. B. Sowerby I, 1833, 1 Vetigastropoda: *Haliotis rufescens* Swainson, 1822, 1 Neomphalina: *Chrysomallon squamiferum* C. Chen, Linse, Copley & Rogers, 2015 and 1 Patellogastropoda: *Lottia gigantea* G. B. Sowerby I, 1834 (Table 1). Protein coding genes of all species were downloaded and gene families were clustered using SonicParanoid v1.3.6 (Cosentino & Iwasaki, 2019) with default parameters.

**Table 1** Summary of published gastropods genome and deep gastropods phylogeny based on corresponding genome studies using single-copy orthologous genes (SOGs).

Sub-class	Species	GenBank assembly accession	Assembly level	Topology	Used of SOGs	References
H	<i>Arion vulgaris</i>	GCA_020796225.1	Chromosome	P,((H,C),(V,N))	223	Chen et al. (2020)
H	<i>Achatina immaculata</i>	GCA_009760885.1	Chromosome	P,(H,C)	229	Liu et al. (2020)
H	<i>Achatina fulica</i>	100647 (Gigadb)	Chromosome	(P,H)	675	Guo et al. (2019)
H	<i>Lymnaea stagnalis</i>	GCA_900036025.1	Contig	–	–	BANG, 2016, Unpublished data
H	<i>Aplysia californica</i>	GCA_000002075.2	Scaffold	–	–	Broad Institute, 2013, Unpublished data
H	<b><i>Biomphalaria glabrata</i></b>	GCA_000457365.1	Scaffold	–	–	Adema et al. (2017)
H	<i>Candidula unifasciata</i>	GCA_905116865.2	Scaffold	–	–	Chueca, Schell & Pfenninger (2021)
H	<i>Cepaea nemoralis</i>	GCA_014155875.1	Scaffold	–	–	Saenko et al. (2021)
H	<b><i>Elysia chlorotica</i></b>	GCA_003991915.1	Scaffold	–	–	Cai et al. (2019)
H	<i>Elysia marginata</i>	GCA_019649035.1	Scaffold	–	–	Maeda et al. (2021)
H	<i>Limacina bulimoides</i>	GCA_009866985.1	Scaffold	–	–	Choo et al. (2020)
H	<i>Physella acuta</i>	GCA_004329575.1	Scaffold	–	–	Ebbs, Loker & Brant (2018)
H	<i>Plakobranthus ocellatus</i>	GCA_019648995.1	Scaffold	–	–	Maeda et al. (2021)
H	<b><i>Radix Auricularia</i></b>	GCA_002072015.1	Scaffold	–	–	Schell et al. (2017)
C	<i>Lautoconus ventricosus</i>	GCA_018398815.1	Chromosome	–	–	Pardos-Blas et al. (2021)
C	<b><i>Pomacea canaliculata</i></b>	GCA_004794335.1	Chromosome	(P,V),(H,C)	1,357*	Sun et al. (2019)
C	<i>Alviniconcha marisindica</i>	GCA_018857735.1	Contig	–	–	HKUST, 2021, Unpublished data
C	<i>Batillaria attramentaria</i>	GCA_018292915.1	Contig	–	–	Ewha Womans University, 2021, Unpublished data
C	<i>Colubraria reticulata</i>	GCA_900004695.1	Contig	–	–	University of Konstanz, 2016, Unpublished data
C	<b><i>Marisa cornuarietis</i></b>	GCA_004794655.1	Contig	–	–	Sun et al. (2019)
C	<i>Phymorhynchus buccinoides</i>	GCA_017654935.1	Contig	–	–	BGI, 2021, Unpublished data
C	<i>Anentome Helena</i>	GCA_009936545.1	Scaffold	–	–	IRIDION GENOMES, 2020, Unpublished data
C	<i>Babylonia areolate</i>	GCA_011634625.1	Scaffold	–	–	Fisheries and Technical, Economic College, 2020, Unpublished data
C	<i>Conus betulinus</i>	GCA_016801955.1	Chromosome	–	–	Peng et al. (2021)
C	<b><i>Conus consors</i></b>	GCA_004193615.1	Scaffold	–	–	Andreson et al. (2019)
C	<i>Conus tribblei</i>	GCA_001262575.1	Scaffold	–	–	Barghi et al. (2016)
C	<b><i>Lanistes nyassanus</i></b>	GCA_004794575.1	Scaffold	–	–	Sun et al. (2019)
C	<i>Pomacea maculate</i>	GCA_004794325.1	Scaffold	–	–	Sun et al. (2019)
V	<i>Steromphala cineraria</i>	GCA_916613615.1	Chromosome	–	–	Wellcome Sanger Institute, 2021, Unpublished data
V	<i>Haliotis laevigata</i>	GCA_008038995.1	Scaffold	–	–	Botwright et al. (2019)
V	<i>Haliotis rubra</i>	GCA_003918875.1	Scaffold	–	–	Gan et al. (2019)
V	<b><i>Haliotis rufescens</i></b>	GCA_003343065.1	Scaffold	–	–	Masonbrink et al. (2019)
N	<i>Gigantopelta aegis</i>	GCA_016097555.1	Chromosome	(P,(V,N)),(H,C)	529	Lan et al. (2021)
N	<b><i>Chrysomallon squamiferum</i></b>	GCA_012295275.1	Chromosome	(P,(V,N)),(H,C)	1,375*	Sun et al. (2020)
N	<i>Dracogyra subfuscus</i>	GCA_016106625.1	Scaffold	–	–	Lan et al. (2021)
P	<b><i>Lottia gigantea</i></b>	GCA_000327385.1	Scaffold	–	–	DOE Joint Genome Institute, 2012, Unpublished data

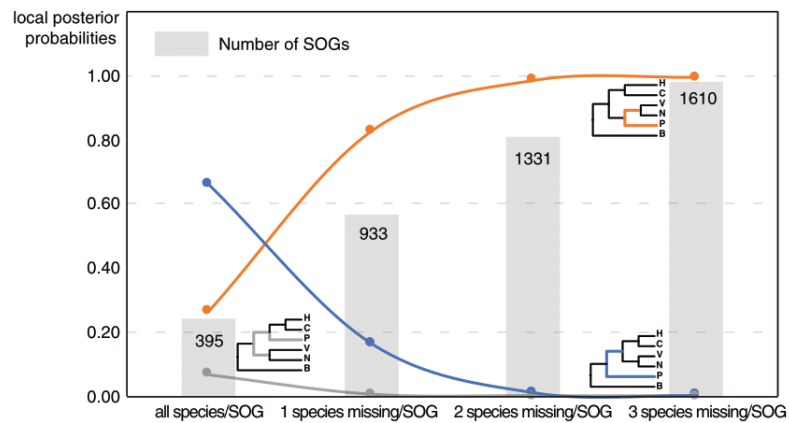
**Note:** Subclass H, C, V, N, P represents Heterobranchia, Caenogastropoda, Vetigastropoda, Neomphalina, and Patellogastropoda respectively. \* Represents SOGs that can be found in at least 60% of taxa. Data without citations have been replaced by the data submitter and data publication date in NCBI. The species in bold are the species used to infer the phylogeny in this article.



**Figure 1** The four data sets used to infer gastropod relationships. The first data set includes single-copy orthologous genes (SOGs) identified in all 16 species per SOG and results in 395 SOGs. 933, 1,331, 1,610 SOGs were identified while allowed 1–3 species missing per SOG respectively. The blue line represents SOGs identified by each species, and the black vertical line represents the missing of that gene in the corresponding species. The dots below represent the number of SOGs occupied by corresponding species under different degrees of species missing.

Full-size [DOI: 10.7717/peerj.13285/fig-1](https://doi.org/10.7717/peerj.13285/fig-1)

SonicParanoid is a graph-based orthology inference tool, given  $N$  input proteomes, SonicParanoid conducts all-vs-all protein alignment for  $N * (N - 1)$  between-proteome and  $N$  within-proteome pairs using MMseqs2 (Steinegger & Söding, 2017). Different SOGs data sets were extracted from the results by: (1) all 16 species, (2) allow one species missing per SOG, (3) allow two species missing per SOG, and (4) allow three species missing per SOG, resulting in a matrix with 395, 933, 1,331, 1,610 genes respectively (Fig. 1).



**Figure 2** The gastropod relationships constructed with different data sets and the corresponding posterior probabilities of different topology. H: Heterobranchia, C: Caenogastropoda, V: Vetigastropoda, N: Neomphalina, P: Patello gastropoda, B: Bivalve.

Full-size DOI: [10.7717/peerj.13285/fig-2](https://doi.org/10.7717/peerj.13285/fig-2)

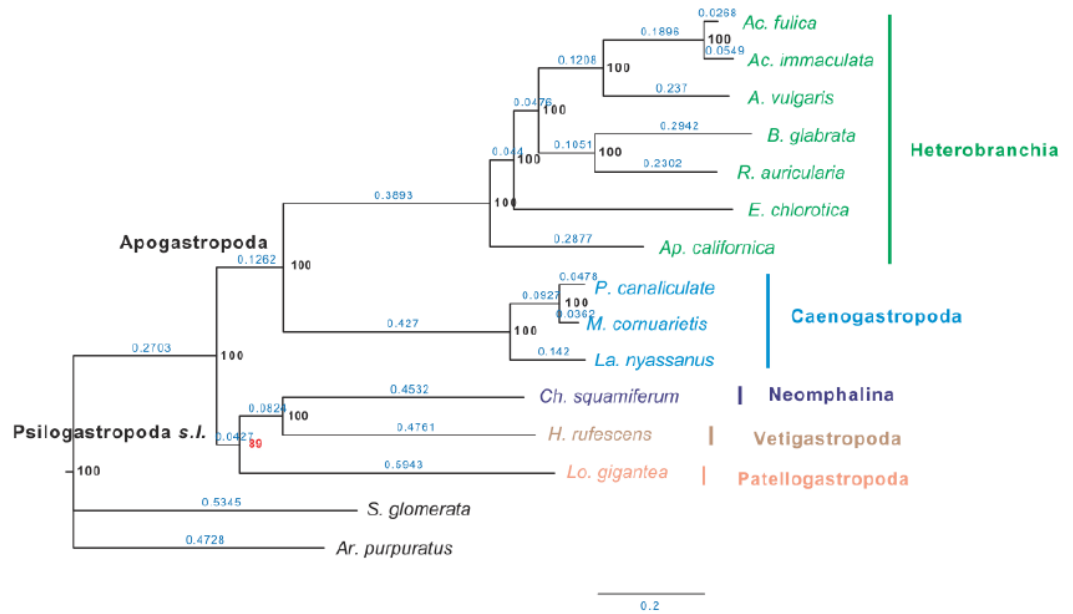
## Phylogenetic analysis

For each SOGs set, amino acid sequences were aligned using MUSCLE v3.8.1551 with the default parameter (Edgar, 2004), and maximum likelihood (ML) gene trees were inferred using RAXML v8.2.12 with the parameter: “-f a -m PROTGAMMAAUTO -k -x 271828 -N 100 -p 31415” respectively (Stamatakis, 2014). The best-scoring ML tree of each gene was extracted separately and merged as input for ASTRAL v5.7.1 to estimate species trees with quartet support and posterior probabilities (PP) (Zhang et al., 2020).

For the subsequent testing of the species tree, we removed the species *C. consors*, which has a significant lower number of SOGs compared with other species. We used the same method to identify SOGs from all the remaining species. Protein sequences were then concatenated and aligned by MUSCLE v3.8.1551, and the maximum likelihood trees were inferred using IQ-TREE v2.0.3 with the parameter: “-m MFP -mtree -b 100”, and with an automatically selected best model (Table S1) (Minh et al., 2020).

## RESULTS

Our main goal was to explore the relationships between Patello gastropoda and other gastropods using SOGs depending on the number of species and genes involved. A total of 395 SOGs were first identified from all 16 species (including 14 gastropods species covering five main subclasses and two bivalves as the outgroup) (Table 1, Fig. 1). The topology of the species tree inferred is as described in Chen et al. (2020): Patello gastropoda is sister to all other gastropods with a low support (PP = 0.66), and the alternative topologies of Patello gastropoda being sister to Vetigastropoda + Neomphalina (PP = 0.27), Patello gastropoda sister to Heterobranchia + Caenogastropoda (PP = 0.07) even less likely (Fig. 2, Fig. S1). However, when we allowed one species missing per SOG, we got a total of 933 SOGs with each species covered by an average of 899 genes (Fig. 1). Surprisingly, the species trees change to Patello gastropoda as sister to Vetigastropoda + Neomphalina as dominant (PP = 0.83), the possibility of Patello gastropoda sister to all



**Figure 3** Gastropod phylogeny inferred from 847 single-copy orthologous genes (SOGs) identified from 15 species except *C. consors*. Bootstrap support percentages are indicated for each internal branch. Branch length is marked with blue numbers. [Full-size !\[\]\(fd7fe780e8fd8eece60268c87d0c3e04\_img.jpg\) DOI: 10.7717/peerj.13285/fig-3](https://doi.org/10.7717/peerj.13285/fig-3)

other gastropods highly decreased (PP = 0.17), and the posterior probabilities of Patellogastropoda sister to Heterobranchia + Ceanogastropoda decreased to 0.005 (Fig. 2, Fig. S2). With more species missing per SOG, the datasets of gene trees used for the estimation of species trees increased, and the posterior probabilities of Patellogastropoda sister to Vetigastropoda + Neomphalina gradually increased to 1 while the probabilities of the two alternative topologies decreased near to 0 (Fig. 2, Figs. S3–S4). Thus, inconsistent gastropod trees appear to be related to the size of the selected data set.

Furthermore, when examining the gene coverage of each species we found that the species *C. consors* is missing in most of the SOGs (Fig. 1). For example, when allowing one species missing per SOG, *C. consors* only showed 718 SOGs, which is 21% lower than the average level (mean = 911) (Fig. 1); allowing two to three species missing per SOG, this value increased from 30% (*C. consors* = 893, mean = 1,271) to 33% (*C. consors* = 1,007, mean = 1,505) (Fig. 1). We thus suspected the consideration of *C. consors* may greatly reduce the number of identified 1:1 orthologous genes, thereby affecting the inference of the phylogenetic tree. Our results support this hypothesis: excluding *C. consors*, we got a total of 847 SOGs (767,180 sites) among the rest of the 15 species, increasing SOG numbers 2.15 times. Then we used the 847 SOGs for phylogenetic reconstruction using a concatenation method and compared it with [Chen et al. \(2020\)](#) who used 233 SOGs (158,094 sites) identified in all 16 species. In contrast to [Chen et al. \(2020\)](#), Patellogastropoda clustered as sister to Vetigastropoda plus Neomphalina with a highly increased support (bootstrap value 89 compared to 48) (Fig. 3).

## DISCUSSION

The lack of resolution and support in molecular phylogenies of early Paleozoic or even Precambrian groups, such as mollusks and gastropods, can be assigned to a plethora of potential problems regarding the sampling and selection of taxa, markers, and analyses (e.g., [Schrödl & Stöger, 2014](#)). One main issue is weak phylogenetic signal which may be eroded over time even in the most conservative genetic markers ([Wagele et al., 2009](#)). Selecting signal from noise by bioinformatic programs may benefit from expanding sets of suitable markers and taxa in, e.g., transcriptomic studies ([Kocot et al., 2011](#); [Zapata et al., 2014](#)); however, without having delivered consistent and fully reliable hypotheses in the case of early gastropod evolution yet. High-quality whole genomes now offer the greatest possible set of genetic markers, with massive single-copy orthologs being the prime candidates for success ([Creevey et al., 2011](#)), but there is still a trade-off with the sampling of taxa, the representation and balance of major subgroups, and the quality and quantity of genomic data available per species. As a case study, here we focused on the role of number and coverage of taxa per SOGs using available quality whole genomes of gastropods.

Although the definition of SOGs is very clear, it is still challenging to correctly identify and optimize single copy (1-to-1) orthologs gene sets, especially on a large species scale. As expected, the numbers of detected SOGs shared by all gastropod species decreased as the number of quarried species increased, which highly influenced the results of the phylogenetic inference ([Fig. 2](#)). Without knowing the historic truth there is no direct way of assessing which (if any) of the controversial trees is correct or how many genes are necessary to recover it. However, using larger sets of SOGs revealed higher supported topologies than pursuing full species coverage with less SOGs: in our study, the gastropod phylogeny reconstructed by 80% species coverage of SOGs had the highest posterior probability of the whole tree ([Fig. 2](#)).

Moreover, the selection of species also needs to be cautious. In our case, the inclusion of *C. consors* led to a significant decrease in the number of shared SOGs and influenced the following phylogenetic inference. This might be caused by the fragmentation and incomplete assembly of *C. consors* (scaffold N50 = 1,128 bp) ([Andreson et al., 2019](#)), which in future studies can be replaced with the newly published chromosome level genome of cone snails, such as *Conus betulinus* Linnaeus, 1758 ([Peng et al., 2021](#)) and *Lautoconus ventricosus* Gmelin, 1791 ([Pardos-Blas et al., 2021](#)).

Furthermore, the currently available genomes constitute a far-from-optimal taxon set for resolving deep gastropod phylogeny: they are a small sample representing a large group, and those that are available come primarily from a rarefied pool ([Sigwart et al., 2021](#)). Such as *L. gigantea*, which is the only available Patellogastropoda, although it showed long-branch attraction and challenged deep gastropod mitogenomics ([Stöger & Schrödl, 2013](#); [Uribe et al., 2019](#)). Whole genomes of Neritimorpha and Cocculiniformia are entirely missing and other major groups such as Neomphalina and Vetigastropoda rather than being represented by single or few members should be much more densely sampled, including the entire diversity of early branching subclades. As shown here, robust



reconstruction of deep gastropod relationships will depend on using sufficient numbers of well-assembled whole genomes from a large and balanced taxon set.

The small and species fully covered SOGs data set used here and by [Chen et al. \(2020\)](#) just weakly supported Orthogastropoda. Our larger SOGs sets allowing 1–3 species missing per SOG, however, rejected Orthogastropoda and recovered Patellogastropoda as sister to a combined clade of Neomphalina and Vetigastropoda. This is in accordance with SOGs studies by [Sun et al. \(2020\)](#) (1,375 SOGs) and [Lan et al. \(2021\)](#) (529 SOGs) and supports large-scale transcriptomics by [Cunha & Giribet \(2019\)](#) who recovered Patellogastropoda sister to Vetigastropoda and called this clade the Psilogastropoda. Because they did not include Neomphalina into their transcriptomic study and established Psilogastropoda as the most inclusive clade containing Patellogastropoda and Vetigastropoda, our combined clade of Neomphalina plus Vetigastropoda as sister to Patellogastropoda supports and expands their Psilogastropoda hypothesis.

## CONCLUSIONS

Striving for complete species covered SOGs in whole genomic studies lowered the total number of SOGs, while less stringent coverage largely expanded the set of SOGs, changed the topologies, and improved node supports. Only if using sufficient SOGs from an optimized taxon sampling, our study supports Patellogastropoda as a sister group to Vetigastropoda + Neomphalina, instead of Patellogastropoda as sister to all other gastropods. Future whole genome phylogenies are promising because of their unique potential to exploit the full genetic signal but should go for larger numbers of SOGs from more complete and better-balanced taxon sets and further explore the trade-offs of numbers, selection, and coverage of genes and species.

## ADDITIONAL INFORMATION AND DECLARATIONS

### Funding

This study has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 764840. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Grant Disclosures

The following grant information was disclosed by the authors:  
European Union's Horizon 2020: 764840.

### Competing Interests

The authors declare that they have no competing interests.

### Author Contributions

- Zeyuan Chen conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.

- Michael Schrödl conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.

### Data Availability

The following information was supplied regarding data availability:

The raw data is available in the [Supplemental File](#).

### Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.13285#supplemental-information>.

## REFERENCES

- Adema CM, Hillier LW, Jones CS, Loker ES, Knight M, Minx P, Oliveira G, Raghavan N, Shedlock A, do Amaral LR, Arican-Goktas HD, Assis JG, Baba EH, Baron OL, Bayne CJ, Bickham-Wright U, Biggar KK, Blouin M, Bonning BC, Botka C, Bridger JM, Buckley KM, Buddenborg SK, Lima Caldeira R, Carleton J, Carvalho OS, Castillo MG, Chalmers IW, Christensens M, Clifton S, Cosseau C, Coustau C, Cripps RM, Cuesta-Astroz Y, Cummins SF, di Stephano L, Dinguirard N, Duval D, Emrich S, Feschotte C, Feyereisen R, FitzGerald P, Fronick C, Fulton L, Galinier R, Gava SG, Geusz M, Geyer KK, Giraldo-Calderon GI, de Souza Gomes M, Gordy MA, Gourbal B, Grunau C, Hanington PC, Hoffmann KF, Hughes D, Humphries J, Jackson DJ, Jannotti-Passos LK, de Jesus Jeremias W, Jobling S, Kamel B, Kapusta A, Kaur S, Koene JM, Kohn AB, Lawson D, Lawton SP, Liang D, Limpanont Y, Liu S, Lockyer AE, Lovato TL, Ludolf F, Magrini V, McManus DP, Medina M, Misra M, Mitta G, Mkoji GM, Montague MJ, Montelongo C, Moroz LL, Munoz-Torres MC, Niazi U, Noble LR, Oliveira FS, Pais FS, Papenfuss AT, Peace R, Pena JJ, Pila EA, Quelais T, Raney BJ, Rast JP, Rollinson D, Rosse IC, Rotgans B, Routledge EJ, Ryan KM, Scholte LLS, Storey KB, Swain M, Tennesen JA, Tomlinson C, Trujillo DL, Volpi EV, Walker AJ, Wang T, Wannaporn I, Warren WC, Wu XJ, Yoshino TP, Yusuf M, Zhang SM, Zhao M, Wilson RK. 2017. Whole genome analysis of a schistosomiasis-transmitting freshwater snail. *Nature Communication* 8(1):15451 DOI 10.1038/ncomms15451.
- 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* 870:223 DOI 10.1101/590695.
- Arquez M, Colgan D, Castro LR. 2014. Sequence and comparison of mitochondrial genomes in the genus *Nerita* (Gastropoda: Neritimorpha: Neritidae) and phylogenetic considerations among gastropods. *Marine Genomics* 15(11):45–54 DOI 10.1016/j.margen.2014.04.007.
- Barghi N, Concepcion GP, Olivera BM, Lluisma AO. 2016. Structural features of conopeptide genes inferred from partial sequences of the *Conus tribblei* genome. *Molecular Genetics and Genomics* 291(1):411–422 DOI 10.1007/s00438-015-1119-2.
- Bieler R. 1992. Gastropod phylogeny and systematics. *Annual Review of Ecology and Systematics* 23(1):311–338 DOI 10.1146/annurev.ecolsys.23.1.311.
- Botwright NA, Zhao M, Wang T, McWilliam S, Colgrave ML, Hlinka O, Li S, Suwansa-Ard S, Subramanian S, McPherson L, King H, Reverter A, Cook MT, McGrath A, Elliott NG, Cummins SF. 2019. Greenlip Abalone (*Haliotis laevigata*) genome and protein analysis provides insights into maturation and spawning. *G3 Genes | Genomes | Genetics* 9(10):3067–3078 DOI 10.1534/g3.119.400388.

- Bouchet P, Rocroi JP, Hausdorf B, Kaim A, Kano Y, Nützel A, Parkhaev P, Schrödl M, Strong EE. 2017. Revised classification, nomenclator and typification of gastropod and monoplacophoran families. *Malacologia* **61**(1–2):1–526 DOI [10.4002/040.061.0201](https://doi.org/10.4002/040.061.0201).
- Brenzinger B, Schrödl M, Kano Y. 2021. Origin and significance of two pairs of head tentacles in the radiation of euthyneuran sea slugs and land snails. *Scientific Reports* **11**(1):21016 DOI [10.1038/s41598-021-99172-5](https://doi.org/10.1038/s41598-021-99172-5).
- Cai H, Li Q, Fang X, Li J, Curtis NE, Altenburger A, Shibata T, Feng M, Maeda T, Schwartz JA, Shigenobu S, Lundholm N, Nishiyama T, Yang H, Hasebe M, Li S, Pierce SK, Wang J. 2019. A draft genome assembly of the solar-powered sea slug *Elysia chlorotica*. *Scientific Data* **6**(1):190022 DOI [10.1038/sdata.2019.22](https://doi.org/10.1038/sdata.2019.22).
- Chen Z, Doğan Ö, Guiguelmoni N, Guichard A, Schrödl M. 2020. The de novo genome of the Spanish slug *Arion vulgaris* Moquin-Tandon, 1855 (Gastropoda: Panpulmonata): massive expansion of transposable elements in a major pest species. *bioRxiv* DOI [10.1101/2020.11.30.403303](https://doi.org/10.1101/2020.11.30.403303).
- Choo LQ, Bal TMP, Choquet M, Smolina I, Ramos-Silva P, Marletaz F, Kopp M, Hoarau G, Peijnenburg K. 2020. Novel genomic resources for shelled pteropods: a draft genome and target capture probes for *Limacina bulimoides*, tested for cross-species relevance. *BMC Genomics* **21**(1):11 DOI [10.1186/s12864-019-6372-z](https://doi.org/10.1186/s12864-019-6372-z).
- Chueca LJ, Schell T, Pfenninger M. 2021. De novo genome assembly of the land snail *Candidula unifasciata* (Mollusca: Gastropoda). *G3 Genes|Genomes|Genetics* **11**(8):jkab180 DOI [10.1093/g3journal/jkab180](https://doi.org/10.1093/g3journal/jkab180).
- Cosentino S, Iwasaki W. 2019. SonicParanoid: fast, accurate and easy orthology inference. *Bioinformatics* **35**(1):149–151 DOI [10.1093/bioinformatics/bty631](https://doi.org/10.1093/bioinformatics/bty631).
- Crame JA. 2013. Early cenozoic differentiation of polar marine faunas. *PLOS ONE* **8**(1):e54139 DOI [10.1371/journal.pone.0054139](https://doi.org/10.1371/journal.pone.0054139).
- Creevey CJ, Muller J, Doerks T, Thompson JD, Arendt D, Bork P. 2011. Identifying single copy orthologs in Metazoa. *PLOS Computational Biology* **7**(12):e1002269 DOI [10.1371/journal.pcbi.1002269](https://doi.org/10.1371/journal.pcbi.1002269).
- Cunha TJ, Giribet G. 2019. A congruent topology for deep gastropod relationships. *Proceedings of the Royal Society B* **286**(1898):20182776 DOI [10.1098/rspb.2018.2776](https://doi.org/10.1098/rspb.2018.2776).
- Cunha TJ, Lemer S, Bouchet P, Kano Y, Giribet G. 2019. Putting keyhole limpets on the map: phylogeny and biogeography of the globally distributed marine family Fissurellidae (Vetigastropoda, Mollusca). *Molecular Phylogenetics and Evolution* **135**:249–269 DOI [10.1016/j.ympev.2019.02.008](https://doi.org/10.1016/j.ympev.2019.02.008).
- EBBS ET, LOKER ES, BRANT SV. 2018. Phylogeography and genetics of the globally invasive snail *Physa acuta* Draparnaud 1805, and its potential to serve as an intermediate host to larval digenetic trematodes. *BMC Ecology and Evolution* **18**(1):103 DOI [10.1186/s12862-018-1208-z](https://doi.org/10.1186/s12862-018-1208-z).
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**(5):1792–1797 DOI [10.1093/nar/gkh340](https://doi.org/10.1093/nar/gkh340).
- Fried B, Huffman JE. 2017. Helminthic diseases: foodborne trematode infections. In: Cockerham WC, ed. *International Encyclopedia of Public Health*. Second Edition. Kidlington: Oxford Elsevier, 327–332.
- Frýda J. 2013. Fossil invertebrates: gastropods. In: *Reference Module in Earth Systems and Environmental Sciences*.
- Gan HM, Tan MH, Austin CM, Sherman CDH, Wong YT, Strugnell J, Gervis M, McPherson L, Miller AD. 2019. Best foot forward: nanopore long reads, hybrid meta-assembly, and haplotig

- purging optimizes the first genome assembly for the southern hemisphere blacklip abalone (*Haliotis rubra*). *Frontiers in Genetics* **10**:889 DOI [10.3389/fgene.2019.00889](https://doi.org/10.3389/fgene.2019.00889).
- Global Invasive Species Database.** 2021. Species profile: *Pomacea canaliculata*. Available at <http://www.iucngisd.org/gisd/species.php?sc=135> (accessed 06 April 2022).
- Gogarten JP, Olendzenski L.** 1999. Orthologs, paralogs and genome comparisons. *Current Opinion in Genetics & Development* **9**(6):630–636 DOI [10.1016/S0959-437X\(99\)00029-5](https://doi.org/10.1016/S0959-437X(99)00029-5).
- Golikov AN, Starobogatov Y.** 1975. Systematics of prosobranch gastropods. *Malacologia* **15**:185–232.
- Gomes-dos-Santos A, Lopes-Lima M, Castro LFC, Froufe E.** 2019. Molluscan genomics: the road so far and the way forward. *Hydrobiologia* **847**(7):1705–1726 DOI [10.1007/s10750-019-04111-1](https://doi.org/10.1007/s10750-019-04111-1).
- Govenar B, Fisher CR, Shank TM.** 2015. Variation in the diets of hydrothermal vent gastropods. *Deep Sea Research Part II: Topical Studies in Oceanography* **121**:193–201 DOI [10.1016/j.dsr2.2015.06.021](https://doi.org/10.1016/j.dsr2.2015.06.021).
- Grande C, Templado J, Zardoya R.** 2008. Evolution of gastropod mitochondrial genome arrangements. *BMC Ecology and Evolution* **8**(1):61 DOI [10.1186/1471-2148-8-61](https://doi.org/10.1186/1471-2148-8-61).
- Greve C, Haase M, Hutterer R, Rödder D, Ihlow F, Misof B.** 2017. Snails in the desert: species diversification of *Theba* (Gastropoda: Helicidae) along the Atlantic coast of NW Africa. *Ecology and Evolution* **7**(14):5524–5538 DOI [10.1002/ece3.3138](https://doi.org/10.1002/ece3.3138).
- Guo Y, Zhang Y, Liu Q, Huang Y, Mao G, Yue Z, Abe EM, Li J, Wu Z, Li S, Zhou X, Hu W, Xiao N.** 2019. A chromosomal-level genome assembly for the giant African snail *Achatina fulica*. *Gigascience* **8**(10):6 DOI [10.1093/gigascience/giz124](https://doi.org/10.1093/gigascience/giz124).
- Haszprunar G.** 1988. On the origin and evolution of major gastropod groups, with special reference to the streptoneura. *Journal of Molluscan Studies* **54**(4):367–441 DOI [10.1093/mollus/54.4.367](https://doi.org/10.1093/mollus/54.4.367).
- Kano Y, Brenzinger B, Nutzelt A, Wilson NG, Schrodler M.** 2016. Ringiculid bubble snails recovered as the sister group to sea slugs (Nudipleura). *Scientific Reports* **6**(1):30908 DOI [10.1038/srep30908](https://doi.org/10.1038/srep30908).
- Kocot KM, Cannon JT, Todt C, Citarella MR, Kohn AB, Meyer A, Santos SR, Schander C, Moroz LL, Lieb B, Halanych KM.** 2011. Phylogenomics reveals deep molluscan relationships. *Nature* **477**(7365):452–456 DOI [10.1038/nature10382](https://doi.org/10.1038/nature10382).
- Lan Y, Sun J, Chen C, Sun Y, Zhou Y, Yang Y, Zhang W, Li R, Zhou K, Wong WC, Kwan YH, Cheng A, Bougouffa S, Van Dover CL, Qiu JW, Qian PY.** 2021. Hologenome analysis reveals dual symbiosis in the deep-sea hydrothermal vent snail *Gigantopelta aegis*. *Nature Communications* **12**(1):1165 DOI [10.1038/s41467-021-21450-7](https://doi.org/10.1038/s41467-021-21450-7).
- Lee H, Chen W, Puillandre N, Aznar-Cormano L, Tsai MH, Samadi S.** 2019. Incorporation of deep-sea and small-sized species provides new insights into gastropods phylogeny. *Molecular Phylogenetics and Evolution* **135**:136–147 DOI [10.1016/j.ympev.2019.03.003](https://doi.org/10.1016/j.ympev.2019.03.003).
- 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**(4):147 DOI [10.1093/gigascience/giy031](https://doi.org/10.1093/gigascience/giy031).
- Lindberg DR.** 1988. The Patellogastropoda. In: Ponder WF, Warén A, eds. *Prosobranch Phylogeny: Malacological Review*. Ann Arbor: The Museum of Zoology, the University of Michigan by the Society for Experimental and Descriptive Malacology, 35–63.
- Liu C, Ren Y, Li Z, Hu Q, Yin L, Qiao X, Zhang Y, Xing L, Xi Y, Jiang F, Wang S, Huang C, Liu B, Wang H, Liu H, Wan F, Qian W, Fan W.** 2020. Giant African snail genomes provide insights into molluscan whole-genome duplication and aquatic-terrestrial transition. *Molecular Ecology Resources* **21**(2):478–494 DOI [10.1111/1755-0998.13261](https://doi.org/10.1111/1755-0998.13261).

- Maeda T, Takahashi S, Yoshida T, Shimamura S, Takaki Y, Nagai Y, Toyoda A, Suzuki Y, Arimoto A, Ishii H, Satoh N, Nishiyama T, Hasebe M, Maruyama T, Minagawa J, Obokata J, Shigenobu S. 2021. Chloroplast acquisition without the gene transfer in kleptoplastic sea slugs, *Plakobranchnus ocellatus*. *Elife* 10:4197 DOI 10.7554/eLife.60176.
- Masonbrink RE, Purcell CM, Boles SE, Whitehead A, Hyde JR, Seetharam AS, Severin AJ. 2019. An annotated genome for *Haliotis rufescens* (Red Abalone) and resequenced green, pink, pinto, black, and white abalone species. *Genome Biology and Evolution* 11(2):431–438 DOI 10.1093/gbe/evz006.
- Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* 37(5):1530–1534 DOI 10.1093/molbev/msaa015.
- 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 Evolutionary Biology* 14(1):197 DOI 10.1186/s12862-014-0197-9.
- Pardos-Blas JR, Irisarri I, Abalde S, Afonso CML, Tenorio MJ, Zardoya R. 2021. The genome of the venomous snail *Lautoconus ventricosus* sheds light on the origin of conotoxin diversity. *Gigascience* 10(5):842 DOI 10.1093/gigascience/giab037.
- Peng C, Huang Y, Bian C, Li J, Liu J, Zhang K, You X, Lin Z, He Y, Chen J, Lv Y, Ruan Z, Zhang X, Yi Y, Li Y, Lin X, Gu R, Xu J, Yang J, Fan C, Yao G, Chen JS, Jiang H, Gao B, Shi Q. 2021. The first *Conus* genome assembly reveals a primary genetic central dogma of conopeptides in *C. betulinus*. *Cell Discovery* 7(1):11 DOI 10.1038/s41421-021-00244-7.
- 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(6):655–665 DOI 10.1093/dnares/dsy032.
- Saenko SV, Groenenberg DSJ, Davison A, Schilthuizen M. 2021. The draft genome sequence of the grove snail *Cepaea nemoralis*. *G3: Genes|Genomes|Genetics* 11(2):2442 DOI 10.1093/g3journal/jkaa071.
- Sang T. 2002. Utility of low-copy nuclear gene sequences in plant phylogenetics. *Critical Reviews in Biochemistry and Molecular Biology* 37(3):121–147 DOI 10.1080/10409230290771474.
- Schell T, Feldmeyer B, Schmidt H, Greshake B, Tills O, Truebano M, Rundle SD, Paule J, Ebersberger I, Pfenninger M. 2017. An annotated draft genome for *Radix auricularia* (Gastropoda, Mollusca). *Genome Biology and Evolution* 9(3):585–592 DOI 10.1093/gbe/evx032.
- Schrödl M, Stöger I. 2014. A review on deep molluscan phylogeny: old markers, integrative approaches, persistent problems. *Journal of Natural History* 48(45–48):2773–2804 DOI 10.1080/00222933.2014.963184.
- Sigwart JD, Lindberg DR, Chen C, Sun J. 2021. Molluscan phylogenomics requires strategically selected genomes. *Philosophical Transactions of the Royal Society B* 376(1825):20200161 DOI 10.1098/rstb.2020.0161.
- Smith S, Wilson N, Goetz F, Feehery C, Andrade Sónia CS, Rouse GW, Giribet G, Dunn CW. 2011. Resolving the evolutionary relationships of molluscs with phylogenomic tools. *Nature* 480:364–367 DOI 10.1038/nature10526.
- Solem GA. 2020. Gastropod. Available at <https://www.britannica.com/animal/gastropod> (accessed 13 January 2022).
- Sonnhammer E, Koonin E. 2002. Orthology, paralogy and proposed classification for paralog subtypes. *Trends in Genetics* 18(12):619–620 DOI 10.1016/S0168-9525(02)02793-2.

- Stamatakis A.** 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30(9)**:1312–1313 DOI [10.1093/bioinformatics/btu033](https://doi.org/10.1093/bioinformatics/btu033).
- Steinegger M, Söding J.** 2017. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology* **35(11)**:1026–1028 DOI [10.1038/nbt.3988](https://doi.org/10.1038/nbt.3988).
- Stöger I, Schrödl M.** 2013. Mitogenomics does not resolve deep molluscan relationships (yet?). *Molecular Phylogenetics and Evolution* **69(2)**:376–392 DOI [10.1016/j.ympev.2012.11.017](https://doi.org/10.1016/j.ympev.2012.11.017).
- Sun J, Chen C, Miyamoto N, Li R, Sigwart JD, Xu T, Sun Y, Wong WC, Ip JCH, Zhang W, Lan Y, Bissessor D, Watsuji TO, Watanabe HK, Takaki Y, Ikeo K, Fujii N, Yoshitake K, Qiu JW, Takai K, Qian PY.** 2020. The scaly-foot snail genome and implications for the origins of biomineralised armour. *Nature Communications* **11(1)**:1657 DOI [10.1038/s41467-020-15522-3](https://doi.org/10.1038/s41467-020-15522-3).
- Sun J, Mu H, Ip JCH, Li R, Xu T, Accorsi A, Alvarado AS, Ross E, Lan Y, Sun Y, Castro-Vazquez A, Vega IA, Heras H, Ituarte S, Bocslaer BV, Hayes KA, Cowie RH, Zhao Z, Zhang Y, Qian PY, Qiu JW.** 2019. Signatures of divergence, invasiveness, and terrestrialization revealed by four apple snail genomes. *Molecular Biology and Evolution* **36(7)**:1507–1520 DOI [10.1093/molbev/msz084](https://doi.org/10.1093/molbev/msz084).
- Uribe JE, Irisarri I, Templado J, Zardoya R.** 2019. New patellogastropod mitogenomes help counteracting long-branch attraction in the deep phylogeny of gastropod mollusks. *Molecular Phylogenetics and Evolution* **133**:12–23 DOI [10.1016/j.ympev.2018.12.019](https://doi.org/10.1016/j.ympev.2018.12.019).
- Uribe JE, Kano Y, Templado J, Zardoya R.** 2016. Mitogenomics of Vetigastropoda: insights into the evolution of pallial symmetry. *Zoologica Scripta* **45(2)**:145–159 DOI [10.1111/zsc.12146](https://doi.org/10.1111/zsc.12146).
- Wagele JW, Letsch H, Klussmann-Kolb A, Mayer C, Misof B, Wagele H.** 2009. Phylogenetic support values are not necessarily informative: the case of the Serialia hypothesis (a mollusk phylogeny). *Frontiers in Zoology* **6(1)**:12 DOI [10.1186/1742-9994-6-12](https://doi.org/10.1186/1742-9994-6-12).
- Wu F, Mueller LA, Cruzillat D, Petiard V, Tanksley SD.** 2006. Combining bioinformatics and phylogenetics to identify large sets of single-copy orthologous genes (COSII) for comparative, evolutionary and systematic studies: a test case in the euasterid plant clade. *Genetics* **174(3)**:1407–1420 DOI [10.1534/genetics.106.062455](https://doi.org/10.1534/genetics.106.062455).
- Zapata F, Wilson NG, Howison M, Andrade SC, Jörger KM, Schrödl M, Goetz FE, Giribet G, Dunn CW.** 2014. Phylogenomic analyses of deep gastropod relationships reject Orthogastropoda. *Proceedings of the Royal Society B: Biological Sciences* **281(1794)**:20141739 DOI [10.1098/rspb.2014.1739](https://doi.org/10.1098/rspb.2014.1739).
- Zemanova MA, Knop E, Heckel G.** 2017. Introgressive replacement of natives by invading Arion pest slugs. *Scientific Reports* **7(1)**:14908 DOI [10.1038/s41598-017-14619-y](https://doi.org/10.1038/s41598-017-14619-y).
- Zhang C, Scornavacca C, Molloy EK, Mirarab S.** 2020. ASTRAL-Pro: quartet-based species-tree inference despite paralogy. *Molecular Biology and Evolution* **37(11)**:3292–3307 DOI [10.1093/molbev/msaa139](https://doi.org/10.1093/molbev/msaa139).

Supplemental Tables and Figures for Publication II

**How many single-copy orthologous genes from whole genomes reveal deep gastropod relationships?**

Table S1. List of models sorted by BIC scores. The best-fit model according to BIC: JTT+F+R5. AIC, w-AIC: Akaike information criterion scores and weights. AICc, w-AICc: Corrected AIC scores and weights. BIC, w-BIC: Bayesian information criterion scores and weights. Plus signs denote the 95% confidence sets. Minus signs denote significant exclusion.

Model	LogL	AIC	w-AIC	AICc	w-AICc	BIC	w-BIC
JTT+F+R5	-9617668.9	19235445.82	+0.12	19235445.8	+0.12	19236069.55	+1
JTT+F+R6	-9617664.9	19235441.83	+0.88	19235441.8	+0.88	19236088.66	-7E-05
LG+F+R6	-9617709.1	19235530.26	-5.52E-20	19235530.3	-5.52E-20	19236177.09	-4E-24
LG+F+R5	-9617725.7	19235559.35	-2.67E-26	19235559.4	-2.67E-26	19236183.08	-2E-25
JTTDCMut+F+R5	-9617961.7	19236031.35	-8.58E-129	19236031.4	-8.59E-129	19236655.07	-7E-128
JTTDCMut+F+R6	-9617957.8	19236027.5	-5.86E-128	19236027.5	-5.86E-128	19236674.33	-4E-132
WAG+F+R5	-9632763.6	19265635.28	-0	19265635.3	-0	19266259.01	-0
WAG+F+R6	-9632758.7	19265629.37	-0	19265629.4	-0	19266276.2	-0
VT+F+R5	-9634284.1	19268676.19	-0	19268676.2	-0	19269299.92	-0
VT+F+R6	-9634282.7	19268677.47	-0	19268677.5	-0	19269324.3	-0
JTT+R5	-9639564.2	19279198.49	-0	19279198.5	-0	19279602.75	-0
JTT+R6	-9639560.2	19279194.34	-0	19279194.3	-0	19279621.71	-0
JTTDCMut+R5	-9639889.1	19279848.28	-0	19279848.3	-0	19280252.55	-0
JTTDCMut+R6	-9639885.5	19279845.01	-0	19279845	-0	19280272.38	-0
rtREV+F+R5	-9644512.4	19289132.85	-0	19289132.9	-0	19289756.57	-0
rtREV+F+R6	-9644502.4	19289116.77	-0	19289116.8	-0	19289763.6	-0
LG+R5	-9646067.8	19292205.66	-0	19292205.7	-0	19292609.93	-0
LG+R6	-9646056.6	19292187.19	-0	19292187.2	-0	19292614.56	-0
LG+R4	-9646466.7	19292999.48	-0	19292999.5	-0	19293380.65	-0
VT+R5	-9648879.4	19297828.73	-0	19297828.7	-0	19298233	-0



VT+R6	-9648878.1	19297830.14	-0	19297830.1	-0	19298257.51	-0
LG+I+G4	-9649218.2	19298494.43	-0	19298494.4	-0	19298829.4	-0
LG+G4	-9651046.4	19302148.89	-0	19302148.9	-0	19302472.31	-0
LG+R3	-9651370	19302801.95	-0	19302802	-0	19303160.02	-0
WAG+R5	-9653905.2	19307880.38	-0	19307880.4	-0	19308284.65	-0
WAG+R6	-9653900.2	19307874.35	-0	19307874.4	-0	19308301.72	-0
Dayhoff+F+R6	-9662727	19325565.98	-0	19325566	-0	19326212.8	-0
Dayhoff+F+R5	-9662744.5	19325596.92	-0	19325596.9	-0	19326220.64	-0
DCMut+F+R6	-9662790.8	19325693.56	-0	19325693.6	-0	19326340.39	-0
DCMut+F+R5	-9662808.2	19325724.43	-0	19325724.4	-0	19326348.16	-0
mtInv+F+R6	-9671163.6	19342439.25	-0	19342439.3	-0	19343086.08	-0
mtInv+F+R5	-9671195	19342498.02	-0	19342498	-0	19343121.75	-0
cpREV+F+R6	-9684203.1	19368518.1	-0	19368518.1	-0	19369164.93	-0
cpREV+F+R5	-9684222.5	19368553.09	-0	19368553.1	-0	19369176.81	-0
PMB+F+R5	-9693167.3	19386442.61	-0	19386442.6	-0	19387066.34	-0
PMB+F+R6	-9693165	19386441.97	-0	19386442	-0	19387088.8	-0
LG+R2	-9696350.1	19392758.18	-0	19392758.2	-0	19393093.15	-0
Blosum62+F+R5	-9697606.1	19395320.19	-0	19395320.2	-0	19395943.92	-0
Blosum62+F+R6	-9697604.4	19395320.71	-0	19395320.7	-0	19395967.54	-0
mtMet+F+R6	-9710893.7	19421899.45	-0	19421899.5	-0	19422546.28	-0
mtMet+F+R5	-9710951.4	19422010.89	-0	19422010.9	-0	19422634.61	-0
Dayhoff+R6	-9711459.9	19422993.79	-0	19422993.8	-0	19423421.16	-0
Dayhoff+R5	-9711476.2	19423022.32	-0	19423022.3	-0	19423426.58	-0
DCMut+R6	-9711641.1	19423356.17	-0	19423356.2	-0	19423783.54	-0
DCMut+R5	-9711656.9	19423383.85	-0	19423383.9	-0	19423788.12	-0

mtZOA+F+R6	-9714727.9	19429567.84	-0	19429567.9	-0	19430214.67	-0
mtZOA+F+R5	-9714990.9	19430089.89	-0	19430089.9	-0	19430713.62	-0
rtREV+R5	-9722048.7	19444167.32	-0	19444167.3	-0	19444571.59	-0
rtREV+R6	-9722039.1	19444152.24	-0	19444152.2	-0	19444579.61	-0
PMB+R5	-9725572.9	19451215.83	-0	19451215.8	-0	19451620.09	-0
PMB+R6	-9725569.1	19451212.16	-0	19451212.2	-0	19451639.53	-0
cpREV+R5	-9728657.4	19457384.89	-0	19457384.9	-0	19457789.16	-0
cpREV+R6	-9728646.3	19457366.68	-0	19457366.7	-0	19457794.05	-0
mtREV+F+R6	-9732086.7	19464285.49	-0	19464285.5	-0	19464932.32	-0
mtREV+F+R5	-9732158.3	19464424.53	-0	19464424.5	-0	19465048.25	-0
Blosum62+R5	-9733212.2	19466494.45	-0	19466494.4	-0	19466898.71	-0
Blosum62+R6	-9733209.4	19466492.76	-0	19466492.8	-0	19466920.13	-0
FLU+F+R6	-9734824.9	19469761.81	-0	19469761.8	-0	19470408.64	-0
FLU+F+R5	-9735056.2	19470220.38	-0	19470220.4	-0	19470844.1	-0
HIVb+F+R6	-9753619.1	19507350.27	-0	19507350.3	-0	19507997.1	-0
HIVb+F+R5	-9753668.6	19507445.14	-0	19507445.1	-0	19508068.87	-0
FLU+R6	-9787731.1	19575536.17	-0	19575536.2	-0	19575963.54	-0
FLU+R5	-9787972.7	19576015.44	-0	19576015.4	-0	19576419.71	-0
mtART+F+R6	-9791886.5	19583885.01	-0	19583885	-0	19584531.84	-0
mtART+F+R5	-9792490.8	19585089.63	-0	19585089.6	-0	19585713.36	-0
mtVer+F+R6	-9812605.2	19625322.36	-0	19625322.4	-0	19625969.19	-0
mtVer+F+R5	-9812686.5	19625480.94	-0	19625481	-0	19626104.67	-0
HIVb+R6	-9836157.5	19672389.08	-0	19672389.1	-0	19672816.44	-0
HIVb+R5	-9836193	19672456.01	-0	19672456	-0	19672860.27	-0
LG+I	-9853377.3	19706810.56	-0	19706810.6	-0	19707133.97	-0

mtMAM+F+R6	-9874496.7	19749105.4	-0	19749105.4	-0	19749752.23	-0
mtMAM+F+R5	-9875083.9	19750275.82	-0	19750275.8	-0	19750899.54	-0
HIVw+F+R6	-9955266.9	19910645.83	-0	19910645.8	-0	19911292.66	-0
HIVw+F+R5	-9955368.8	19910845.67	-0	19910845.7	-0	19911469.4	-0
mtZOA+R6	-10028697	20057467.72	-0	20057467.7	-0	20057895.09	-0
mtZOA+R5	-10028812	20057693.08	-0	20057693.1	-0	20058097.35	-0
mtMet+R6	-10106713	20213500.58	-0	20213500.6	-0	20213927.95	-0
mtMet+R5	-10106771	20213612.8	-0	20213612.8	-0	20214017.07	-0
LG	-10111112	20222278.9	-0	20222278.9	-0	20222590.76	-0
mtREV+R6	-10125829	20251731.26	-0	20251731.3	-0	20252158.63	-0
mtREV+R5	-10125882	20251834.98	-0	20251835	-0	20252239.25	-0
HIVw+R6	-10143980	20288033.12	-0	20288033.1	-0	20288460.49	-0
HIVw+R5	-10144105	20288279.62	-0	20288279.6	-0	20288683.89	-0
mtInv+R6	-10171857	20343787.74	-0	20343787.7	-0	20344215.11	-0
mtInv+R5	-10171879	20343827.44	-0	20343827.4	-0	20344231.71	-0
mtART+R6	-10186387	20372848.9	-0	20372848.9	-0	20373276.27	-0
mtART+R5	-10187267	20374603.08	-0	20374603.1	-0	20375007.35	-0
mtVer+R6	-10194988	20390050.56	-0	20390050.6	-0	20390477.93	-0
mtVer+R5	-10195102	20390273.7	-0	20390273.7	-0	20390677.97	-0
mtMAM+R6	-10283473	20567019.46	-0	20567019.5	-0	20567446.83	-0
mtMAM+R5	-10283866	20567801.1	-0	20567801.1	-0	20568205.37	-0

---

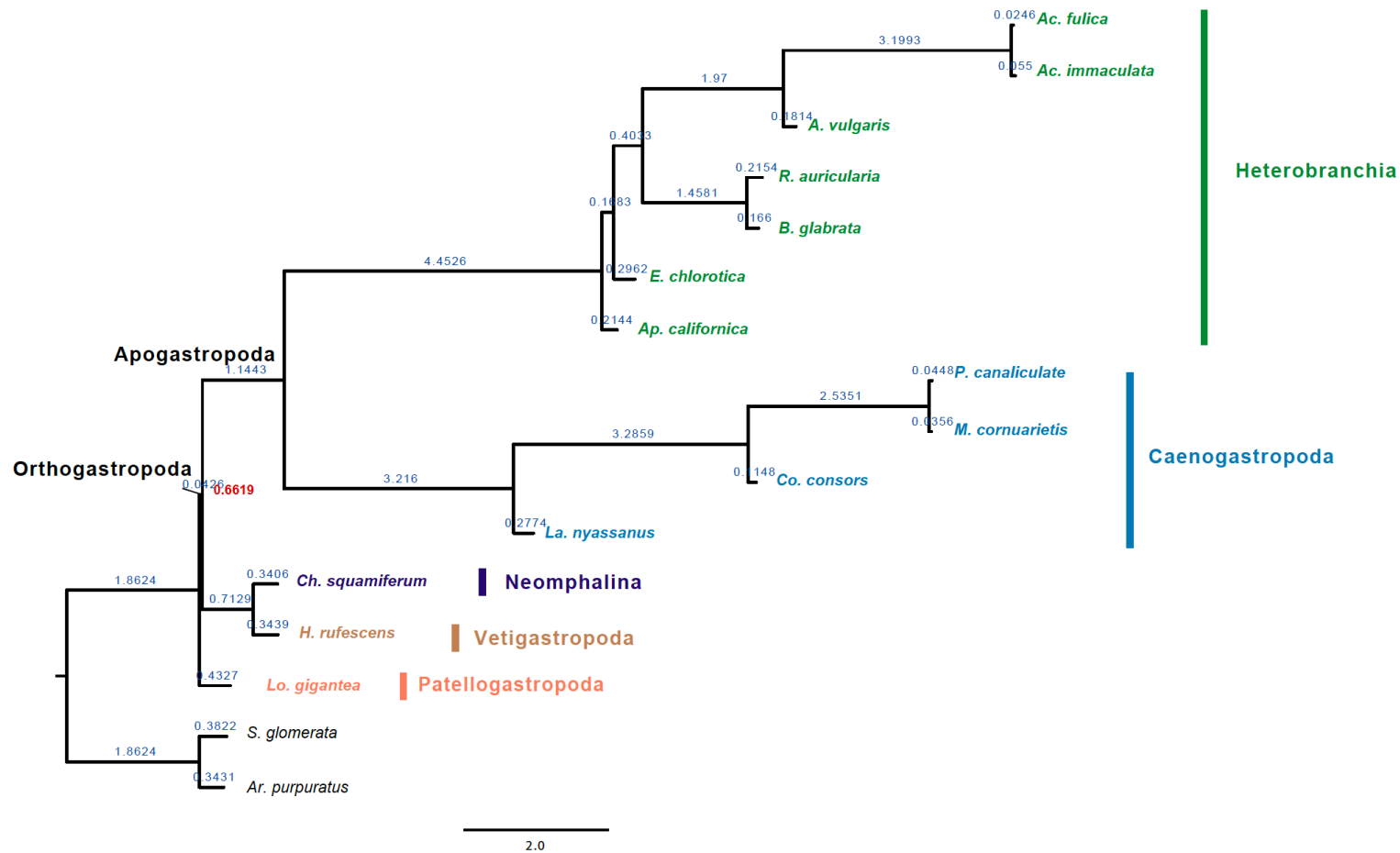


Fig S1. Gastropod phylogeny inferred from 395 single-copy orthologous genes (SOGs) using ASTRAL. posterior probability (PP) is highlighted in red. Branch length is marked with blue numbers.

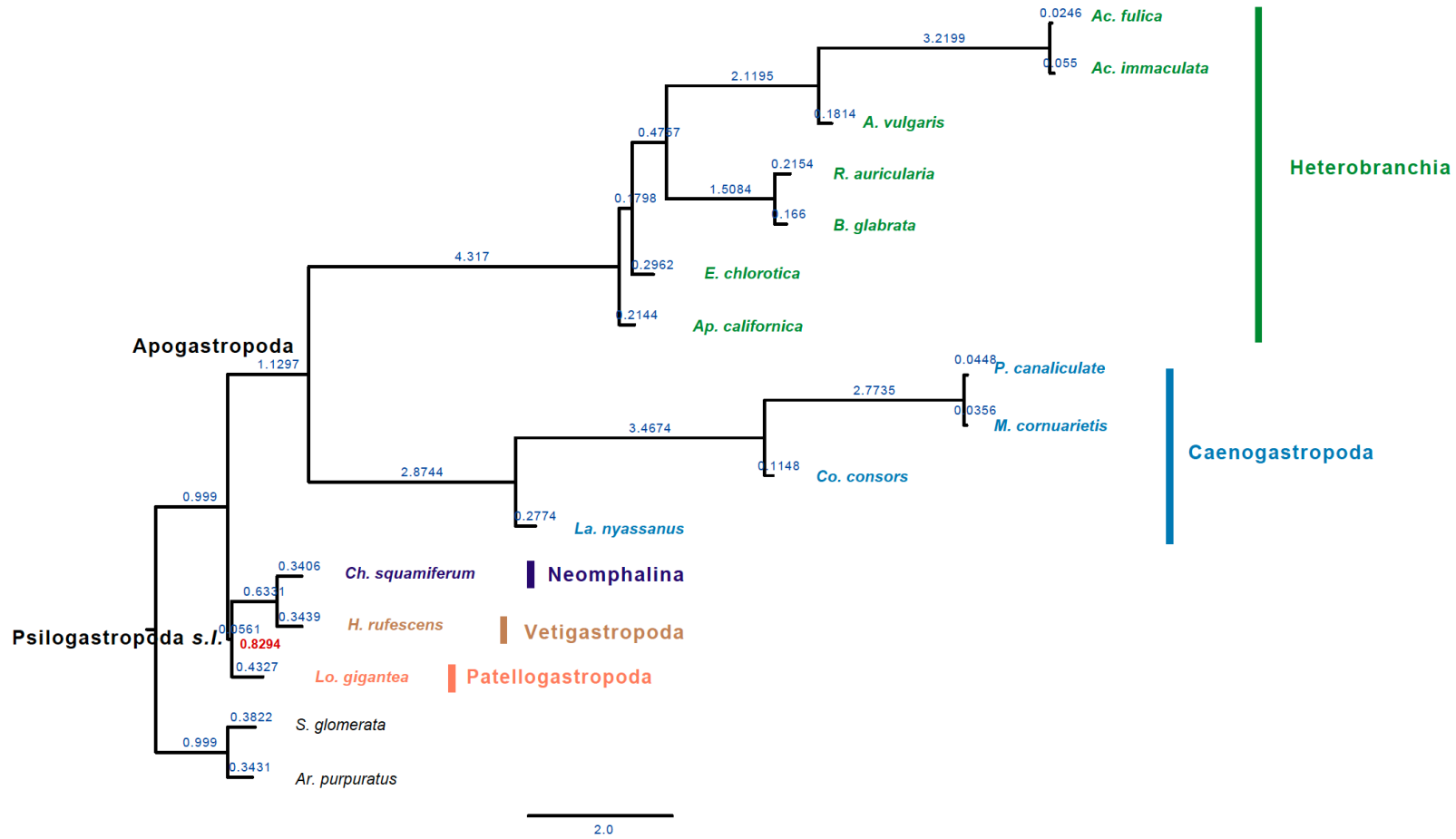


Fig S2. Gastropod phylogeny inferred from 933 SOGs using ASTRAL. posterior probability (PP) is highlighted in red. Branch length is marked with blue numbers.

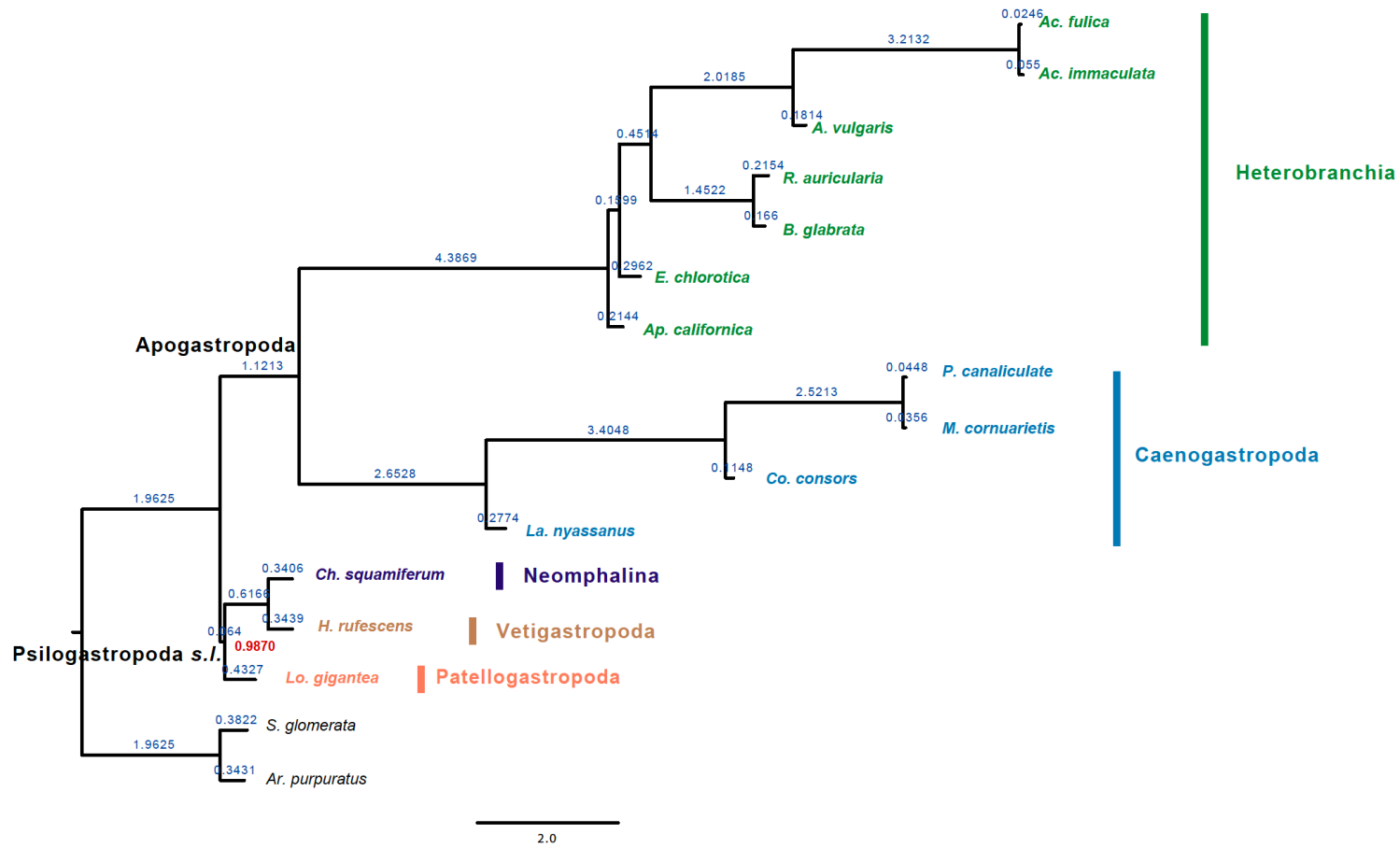


Fig S3. Gastropod phylogeny inferred from 1331 SOGs using ASTRAL. posterior probability (PP) is highlighted in red. Branch length is marked with blue numbers.

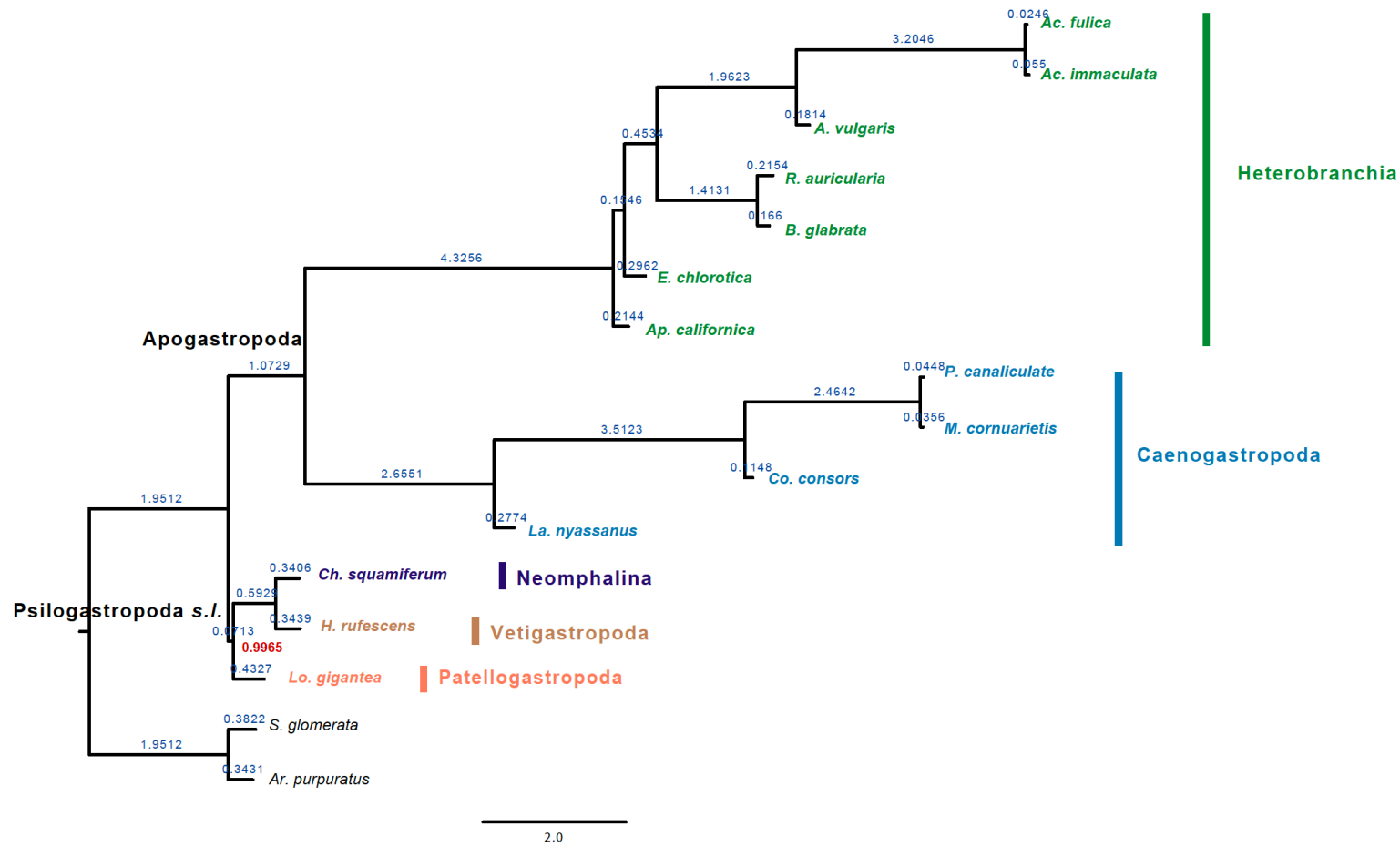


Fig S4. Gastropod phylogeny inferred from 1610 SOGs using ASTRAL. posterior probability (PP) is highlighted in red. Branch length is marked with blue numbers.

### 3.3 Publication III

**Zeyuan Chen**, Özgül Doğan, Nadège Guiguelmoni, Anne Guichard & Michael Schrödl (2022): Pulmonate slug evolution is reflected in the *de novo* genome of *Arion vulgaris* Moquin-Tandon, 1855. *Scientific Reports* 12, 14226

A pdf of the article is available at:

<https://www.nature.com/articles/s41598-022-18099-7.pdf>.

Open Access. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made.





OPEN

## Pulmonate slug evolution is reflected in the de novo genome of *Arion vulgaris* Moquin-Tandon, 1855

Zeyuan Chen<sup>1,2</sup>✉, Özgül Doğan<sup>3</sup>, Nadège Guiguelmoni<sup>4</sup>, Anne Guichard<sup>5,6</sup> & Michael Schrödl<sup>1,2,7</sup>

Stylommatophoran pulmonate land slugs and snails successfully completed the water-to-land transition from an aquatic ancestor and flourished on land. Of the 30,000 estimated species, very few genomes have so far been published. Here, we assembled and characterized a chromosome-level genome of the “Spanish” slug, *Arion vulgaris* Moquin-Tandon, 1855, a notorious pest land slug in Europe. Using this reference genome, we conclude that a whole-genome duplication event occurred approximately 93–109 Mya at the base of Stylommatophora and might have promoted land invasion and adaptive radiation. Comparative genomic analyses reveal that genes related to the development of kidney, blood vessels, muscle, and nervous systems had expanded in the last common ancestor of land pulmonates, likely an evolutionary response to the terrestrial challenges of gravity and water loss. Analyses of *A. vulgaris* gene families and positively selected genes show the slug has evolved a stronger ability to counteract the greater threats of external damage, radiation, and water loss lacking a protective shell. Furthermore, a recent burst of long interspersed elements in the genome of *A. vulgaris* might affect gene regulation and contribute to rapid phenotype changes in *A. vulgaris*, which might be conducive to its rapid adaptation and invasiveness.

Land slugs and snails (Mollusca: Gastropoda), which are often abundant in gardens, forests, fields, and orchards, are, for the most part, classified as stylommatophoran pulmonates. They have radiated into about 30,000 species, have highly successfully colonized habitats from polar regions to the tropics, and some are well-known invasive species or pests across the world<sup>1–5</sup>. Stylommatophoran pulmonates are among the few representatives of mollusks that have colonized the terrestrial environment. The changes in the physical and chemical properties of the environment are immense for animals moving from aquatic to terrestrial environments, and these changes could affect all possible life processes, from respiration and excretion to methods of movement, the functioning of sense organs, and reproduction<sup>6</sup>. Overcoming drought, for example, is one of the biggest challenges in water-land transition<sup>7</sup>. Compared to land snails, the lack of a protective shell in land slugs seems to have further increased the difficulty in coping with external stimuli, predators, sun exposure, and drought. Land slugs have evolved certain innovations, such as defense by chemical compounds or behavior, to counteract these challenges<sup>1,8</sup>. However, the lack of shell also gives advantages such as reduced weight and lower energy costs, reduced dependence on calcium uptake, better mobility, and ability to occupy small spaces. Recently, comparative genomics methods have provided key perspectives for revealing the process of water-land transition and illuminated adaptive mechanisms<sup>9–11</sup>. With the rapid development of genome sequencing, several land snail genomes have been published (Supplementary data 1), however, the genomic resources for land slugs are still lacking.

In recent years, the notorious “Spanish” slug, *Arion vulgaris* Moquin-Tandon, 1855, has attracted widespread attention due to its invasiveness and negative impact on the economy, ecology, health, and social system<sup>12</sup>. As a major defoliator of plants, *A. vulgaris* causes serious damage in orchard cultivation, gardens, and agriculture resulting in financial losses<sup>13–15</sup>. *Arion vulgaris* also transmits plant pathogens, contaminates silage, and might

<sup>1</sup>SNSB-Bavarian State Collection of Zoology, Münchhausenstr. 21, 81247 Munich, Germany. <sup>2</sup>Department Biology II, Ludwig-Maximilians-Universität, Planegg-Martinsried, 82152 Munich, Germany. <sup>3</sup>Department of Molecular Biology and Genetics, Faculty of Science, Sivas Cumhuriyet University, Sivas, Turkey. <sup>4</sup>Evolutionary Biology and Ecology, Université Libre de Bruxelles, 1050 Brussels, Belgium. <sup>5</sup>INRAE, Agrocampus Ouest, Université de Rennes, IGEPP, 35650 Le Rheu, France. <sup>6</sup>Univ. Rennes, CNRS, Inria, IRISA-UMR 6074, 35000 Rennes, France. <sup>7</sup>GeoBio-Center LMU, 80333 Munich, Germany. ✉email: Chen.Z@snsb.de

cause health problems in animals<sup>16,17</sup>. It also outcompetes native slug species and reduces biodiversity<sup>18</sup>. Delivering Alien Invasive Species Inventoried for Europe (DAISIE) has listed *A. vulgaris* as one of the 100 worst alien species in Europe<sup>19</sup>, and it is the only land gastropod in the list. Although recent studies disputed the origin and invasiveness of *A. vulgaris* based on the genetic diversity patterns of mitochondrial and nuclear loci<sup>20–22</sup>, its outstanding adaptability and mass occurrences are undeniable.

Here, we assembled and annotated the first land slug genome—*A. vulgaris*. By comparing *A. vulgaris* with two stylommatophoran land snails, and stylommatophoran species with other aquatic or marine gastropods, the well-annotated genome provides a broader perspective to decipher the water-land transition process of stylommatophoran species. The *A. vulgaris* genome also provides insights into how shell-less *A. vulgaris* adapted to terrestrial environments and the underlying molecular mechanisms (e.g., whole-genome duplication (WGD), small scale gene duplication, transposable elements explosion). Moreover, the high-quality genome provides an important reference for future research on *A. vulgaris* population genetics and mollusk evolutionary trajectories, e.g., the loss and evolution of mollusk shells<sup>23</sup>.

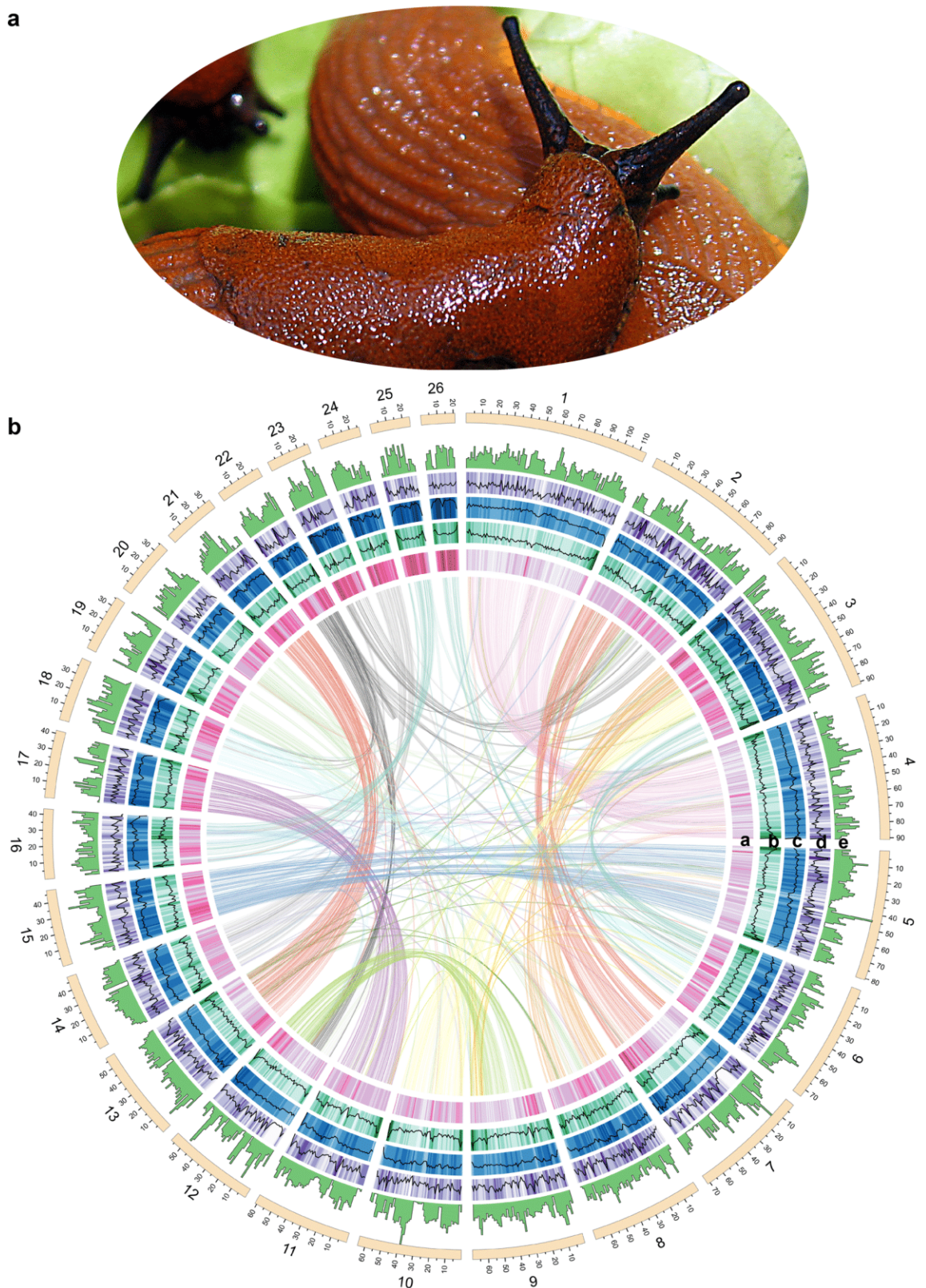
## Results

**Arion vulgaris genome assembly and annotation.** The genome size of *A. vulgaris* (Fig. 1a) is estimated to be around 1.45 Gb from *k*-mer analysis with short reads (Supplementary Fig. S1; Supplementary Table S1). We sequenced 75 Gb (52x) of long reads (mean length 19.39 kb, N50 length 25.80 kb) using Oxford Nanopore sequencing technology to produce a draft genome assembly. The draft assembly was polished using a combination of 57 Gb (40x) Illumina short reads and 138 Gb (95x) 10X Genomics linked reads. Next, the polished assembly was scaffolded using linked reads and then improved into a chromosome-level assembly with 135 Gb (93x) Hi-C data (Supplementary Table S1). Finally, we obtained an assembly with a total length of 1.54 Gb, a contig N50 of 8.6 Mb, and a scaffold N50 of 63.3 Mb, and with 93.8% of the sequences anchored onto 26 scaffolds (Supplementary Fig. S1–2; Supplementary Table S2). The number of chromosome-scale scaffolds is consistent with the species' determined chromosome number based on karyotype studies<sup>24</sup>. We assessed the quality of the genome assembly in three aspects: (1) more than 95.99% of the Illumina short reads could be mapped to the assembly; (2) a total of 886 (90.59%) conserved genes in BUSCO's metazoan (odb9) benchmark set<sup>25</sup> were present and complete in the genome (Supplementary Table S2); (3) the *k*-mer distribution showed a relatively collapsed assembly including mostly single copies of the homozygous content and a partial representation of the heterozygous content, as is expected in a haploid assembly<sup>26</sup> (Supplementary Fig. S3). These results all suggested a high-quality genomic resource of this initial genome assembly of *A. vulgaris*, which is comparable to other mollusk genomes, especially in a high level of heterozygosity and repeats content (Supplementary Fig. S4; Supplementary data 1). Gene annotation combining the evidence from transcripts, homologous proteins, and ab initio prediction revealed 32,518 predicted genes with an average length of 15,429 bp (Supplementary Table S4–5). The length distribution of transcripts, coding sequences, exons and introns, and the distribution of exon numbers per gene were comparable to that of other gastropods (Supplementary Fig. S5). Among the predicted protein-coding genes, 97.6% could be annotated through at least one of the following protein-related databases: the EggNOG<sup>27</sup> database (51.64%), the Swiss-Prot<sup>28</sup> protein database (97.57%), the Translated European Molecular Biology Laboratory (TrEMBL)<sup>28</sup> database (96.65%), the protein families (Pfam)<sup>29</sup> database (81.55%), and the Kyoto Encyclopedia of Genes and Genomes (KEGG)<sup>30</sup> database (29.65%) (Supplementary Table S6).

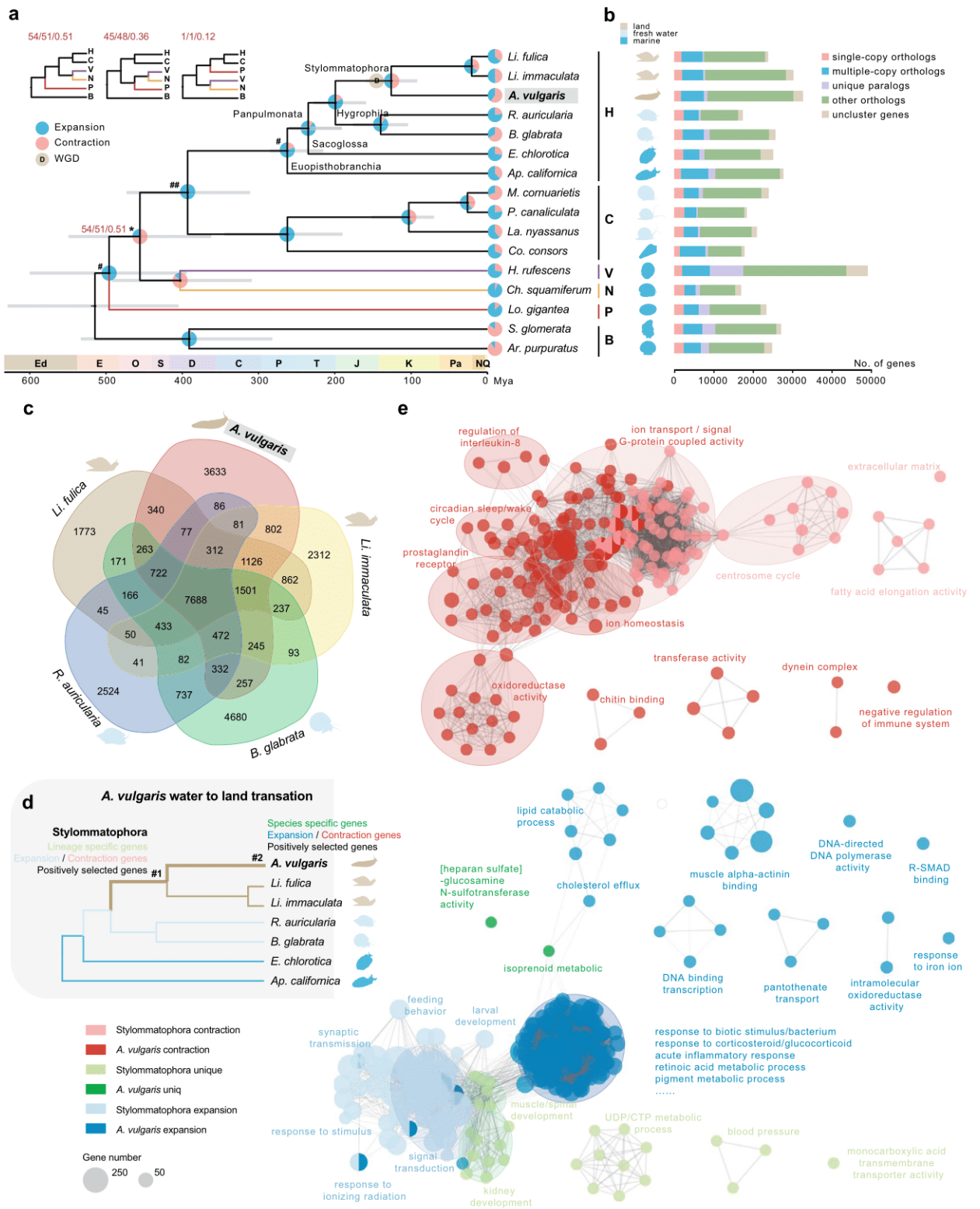
**Phylogenetic relationships within gastropod lineages.** The relationship of early gastropods has been controversial for a long time as different datasets and methodology show different topologies<sup>31–33</sup>. By means of comparing whole genomic data, a total of 223 single-copy orthologous genes (158,094 amino acid sites) were identified from 14 gastropod species that cover five main gastropod subclasses and 2 bivalve species (Supplementary Table S3). Both concatenated and coalescent-based methods produced an identical strongly supported topology (bootstrap value = 100, posterior probabilities = 1), except for the position of the Patellogastropoda and the Vetigastropoda + Neomphalina clades (Fig. 2a; Supplementary Fig. S6a). Our results show Patellogastropoda as sister to all other gastropods, and monophyletic Vetigastropoda + Neomphalina as sister to the clade Apogastropoda (Heterobranchia + Caenogastropoda) with relatively higher support compared with the other two topologies: Patellogastropoda as sister to Vetigastropoda + Neomphalina, and Patellogastropoda as sister to Heterobranchia + Caenogastropoda (Fig. 2a; Supplementary Fig. S6). The results thus favor the hypothesis of a clade Orthogastropoda (the united clade of Heterobranchia, Caenogastropoda, Vetigastropoda, and Neritimorpha), which is congruent with morphology-based and recently reported mitogenomic phylogenies<sup>32,34–36</sup>; but see Chen and Schrödl<sup>37</sup>.

Molecular dating suggests that *A. vulgaris* diverged from the most recent common ancestor with the land snails *Lissachatina (Achatina) fulica* and *Li. immaculata* about 126 million years ago (Mya, 95% confidence interval: 92–159 Mya) (Fig. 2a). The estimated divergence time is close to a previous estimate (132 Mya) based on mitochondrial genomes<sup>38</sup>. Stylommatophora split from Hygrophila around 199 Mya (95% confidence interval: 159–228 Mya), and Panpulmonata split from Sacoglossa around 235 Mya (95% confidence interval: 191–260 Mya) (Fig. 2a).

**Analysis of gene family evolution provides insights into *A. vulgaris* terrestrial adaptation.** Recently, the changes of gene families have been recognized as a primary driver of phenotypic diversity and adaptive evolution<sup>9</sup>. Hence, we investigate the genetic basis of species adaptive evolution by defining the relationship of gene families. Based on pairwise sequence similarities, we identified 26,693 putative orthologous gene families composed of 378,381 genes among *A. vulgaris*, other gastropods, and outgroup species, of which 1610 gene clusters were shared by all gastropod species, representing ancestral gastropod gene families (Fig. 2b;



**Figure 1.** Genome features of *Arion vulgaris*. (a) Adult *A. vulgaris*. (b) General characteristics of the *A. vulgaris* genome. Tracks from inside to outside correspond to (a) GC content, (b) LTRs density, (c) TEs density, (d) genes density, and (e) heterozygosity in sliding windows of 1 Mb across each of the 26 pseudo-chromosomes. Inner lines connect syntenic genes due to ancestral whole-genome duplication events.



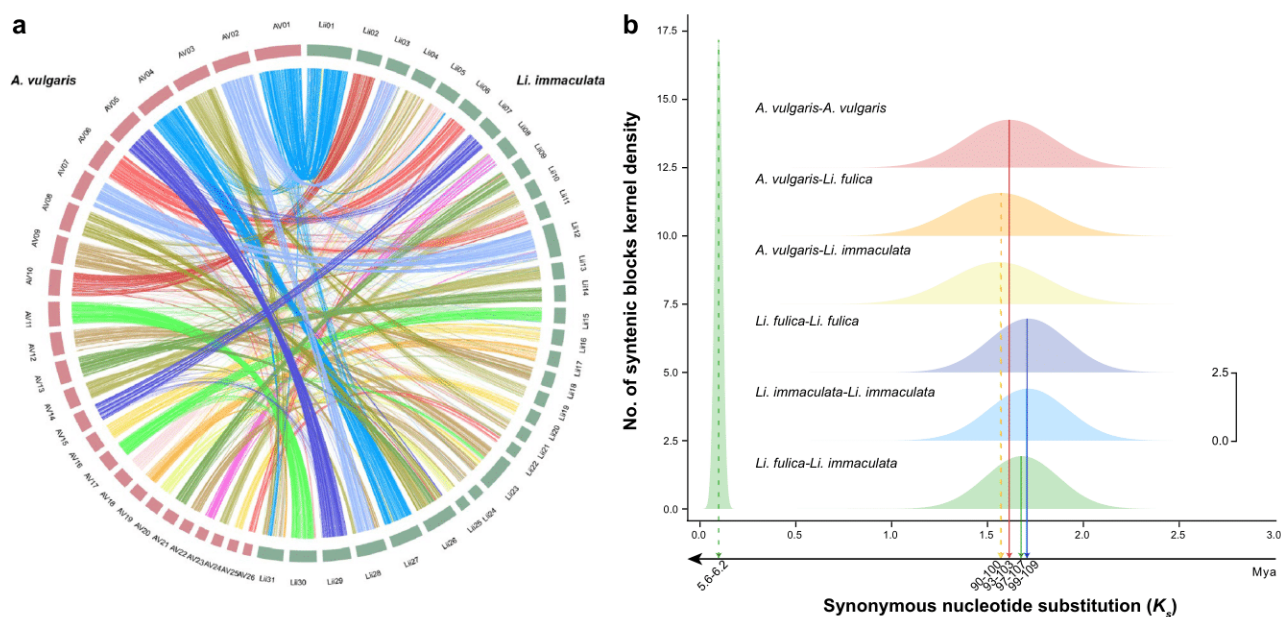
◀**Figure 2.** Phylogeny within gastropod lineages and gene family evolution. (a) Dated Maximum-likelihood phylogenetic tree among gastropod species using 223 single-copy orthologous genes. Grey lines indicate the 95% confidence intervals for the time of divergence between different clades. “#” indicates that nodes are constrained with fossil calibration and “##” indicate node is constrained with secondary calibration data. All nodes have bootstrap support values of 100 and posterior probabilities of 1 in all analyses, except the node with an asterisk. Three alternative topologies are shown on top left. Red numbers are bootstrap support percentages and posterior probabilities (from left to right, inferred by RaxML with GTR+L, IQ-TREE and ASTRAL respectively). H, Heterobranchia; C, Caenogastropoda; V, Vetigastropoda; N, Neomphalina; P, Patellogastropoda; B, Bivalve. The pie diagram on each branch of the tree represents the proportion of gene families undergoing expansion (blue) or contraction (red) events. (b) The distribution of single-copy, multiple-copy, unique, and other orthologs in the 16 mollusks. (c) Venn diagram represents the number of shared and unique gene families among five Panpulmonata species. (d) A simplified diagram showing the evolution of *A. vulgaris* water-land transition. (e) Gene Ontology (GO) enrichment map summarizing major biological networks of *Arion vulgaris* and *Stylommatophora* specific, expanded and contraction genes. Each node represents one GO term with adjusting P-value < 0.05. Node sizes indicate the number of genes within the corresponding GO term. The thickness of the edges represents the number of genes shared by two terms. Striking groups were manually circled and labeled.

Supplementary Table S7). A total of 10,311 orthologous gene families were shared by all Heterobranchia species and 7688 orthologous gene families were shared by five Panpulmonata species (Fig. 2b, c; Supplementary Table S7).

To explore the genetic basis of terrestrial adaptability shared by stylommatophoran species, we considered the properties of the 1126 gene families exclusively shared by three stylommatophoran species (Fig. 2c, d). GO enrichment analyses of these lineage-specific genes demonstrated that they were mainly assigned to kidney development, CTP/UDP metabolic processes, regulation of blood pressure, muscle growth, and spinal development (Fig. 2e; Supplementary Fig. S7; Supplementary Table S8). Molluscan kidneys are involved in the secretion of waste and the resorption of metabolites from the urinary fluid<sup>39</sup>. The enrichment of a series of genes related to kidney and ureteric bud development suggests the improvement of the efficiency of maintaining water balance and nutrients re-absorption in stylommatophoran species (Supplementary Table S8). In addition, the enriched biological process related to blood pressure regulation might be responsible for overcoming the gravity problem during landing<sup>8</sup>. Moreover, the enriched functions of muscle growth and spinal development might also improve the movement and flexibility in terrestrial life (Fig. 2e; Supplementary Fig. S7; Supplementary Table S8). There were 2140 and 1958 gene families expanded and contracted in the Stylommatophora lineage respectively (Fig. 2a). The expanded genes were functionally enriched in response to stimulus, response to radiation, signaling, larval development, and regulation of feeding/eating behavior (Fig. 2e; Supplementary Fig. S8; Supplementary Table S9). Meanwhile, genes related to transmembrane transport, fatty acid elongation, and centrosome cycle were contracted in both *A. vulgaris*, *Li. fulica* and *Li. immaculata* (Fig. 2e; Supplementary Fig. S9; Supplementary Table S10). A total of 251 genes are likely positively selected in Stylommatophora, and their function mainly refers to the regulation of myelination (Supplementary Fig. S10; Supplementary Table S11).

Considering the specific adaptations of shell-less *A. vulgaris* (Fig. 2d), we identified a total of 2763 genes unique to *A. vulgaris*, of which 2629 (95.2%) have known InterPro domains (Supplementary Table S7). We found *A. vulgaris* specific genes were significantly enriched in functional categories related to isoprenoid metabolic process and organelle cell components (Fig. 2e; Supplementary Fig. S11; Supplementary Table S12). In comparison with two *Lissachatina* land snail species, *A. vulgaris* expanded genes exhibited significant enrichment in various aspects, including immune system, response to biotic/radiation stress, excretion, etc., which are very likely beneficial for its land adaptation (Fig. 2e; Supplementary Fig. S12; Supplementary Table S13). Specifically, genes related to response to corticosteroids and glucocorticoid pathways are highly increased. Corticosteroids are involved in a wide range of physiologic systems such as stress response, immune response, and regulation of inflammation<sup>40</sup>, glucocorticoids act primarily on carbohydrate and protein metabolism, and have anti-inflammatory effects<sup>41,42</sup>. Moreover, processes in acute inflammatory response and regeneration are also enriched. All of these might highly improve the ability of *A. vulgaris* to recover from damage. We found *A. vulgaris* expanded genes were also enriched in response to molecules of bacterial origin and response to lipopolysaccharide, which might improve its ability in response to biotic stress. The enrichment of pigment metabolism processes might advance the ability of shell-less *A. vulgaris* to reduce solar radiation damage. Furthermore, genes related to excretion, uronic acid metabolism, and larval development are expanded in *A. vulgaris*. Surprisingly, we also found an enrichment of genes related to pesticides, which might be the result of interaction with human agricultural activity. Similar to genes contracted in Stylommatophora, a high proportion of contracted genes were functionally related to transmembrane transport processes. In addition, contracted genes that regulate circadian rhythm and oxidase activity are also enriched (Fig. 2e; Supplementary Fig. S13; Supplementary Table S14). Strikingly, we found that genes involved in the positive regulation of interleukin-8 production were enriched in *A. vulgaris* contracted genes, and genes related to Interleukin-3,4,9,10,12,21,23,27,35 were likely positively selected in *A. vulgaris* (Supplementary Fig. S14; Supplementary Table S15). This adaptive immune response again might highly increase the ability of *A. vulgaris* in response to stress and stimuli.

**Whole-genome duplication events shared by Stylommatophora species.** Whole-genome duplication (WGD) events are proposed to be a key evolutionary event driving phenotypic complexity, functional novelty, and ecological adaptation<sup>43</sup>. An earlier study suspected a WGD event somewhere at the base of Stylommatophora by comparison of chromosome numbers among closely related mollusks<sup>44</sup>, and a recent genomic study of *Li. immaculata* and *Li. fulica* proved the WGD event using genomic analysis and deduced the WGD



**Figure 3.** Whole-genome duplication shared by *Arion vulgaris*, *Lissachatina (Achatina) fulica*, and *Li. immaculata*. **(a)** An approximately one to one corresponding relationship in the comparison of *A. vulgaris* and *Li. immaculata* genomes. **(b)** Frequency distributions are shown of values of synonymous substitutions ( $K_s$ ) for homologous gene pairs in comparisons of *A. vulgaris*, *Li. fulica* and *Li. immaculata*.

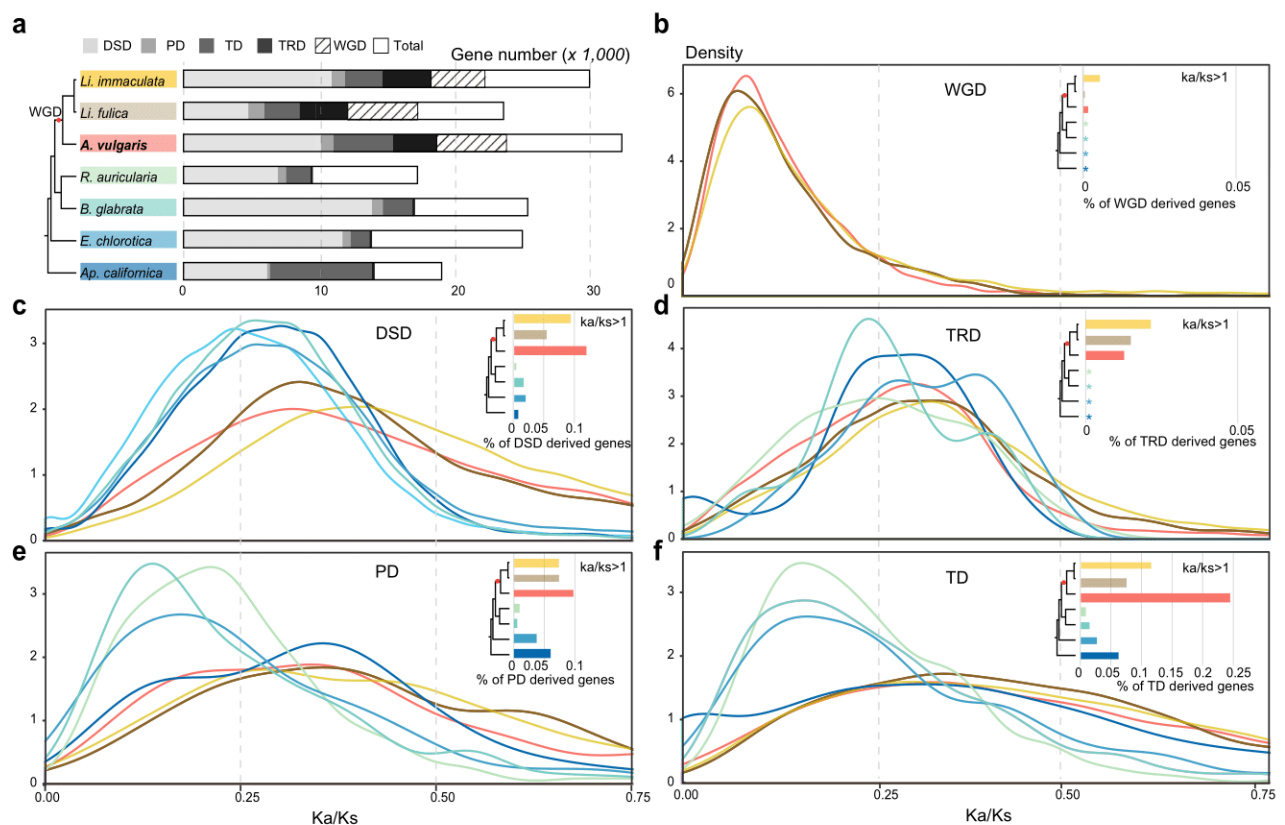
event occurred around 70 Mya<sup>11</sup>. However, 70 Mya is much later than the divergence time that we estimated between *A. vulgaris* and *Lissachatina* (126 Mya, Fig. 2a). Therefore, we raised two questions: (a) whether *A. vulgaris* also experienced a WGD event, and (b) if it has happened, whether it happened independently after divergence from *Lissachatina* or it was shared by their common ancestor.

Our results of chromosome macrosynteny show that most of the chromosomes found a corresponding one in the *A. vulgaris* genome (Fig. 1b). In addition, we detected an approximately one-to-one corresponding relationship in the comparison of *A. vulgaris* ( $n=26$ ) and *Li. immaculata* ( $n=31$ ) chromosomes (Fig. 3a) and a one-to-two corresponding relationship in the comparison of *A. vulgaris* and *Aplysia californica* ( $n=17$ ) chromosomes (Supplementary Fig. S15). In both *Lissachatina* and *A. vulgaris*, the distribution of synonymous substitutions ( $K_s$ ) shows a clear peak, which represents WGD events. There is also an overall slower synonymous substitution rate in *A. vulgaris* (max  $K_s$ : 1.61) than in *Li. fulica* (max  $K_s$ : 1.71) and *Li. immaculata* (max  $K_s$ : 1.71) (Fig. 3b; Supplementary Fig. S16). Based on our results and the previous karyotype research, we conclude that a WGD event did occur in the ancestry of *A. vulgaris*.

To figure out when the WGD event happened, we further compared the syntenic gene pairs between *A. vulgaris* and two *Lissachatina* snails. First, the results showed the best BLASTP hits of homologous gene pairs are from interspecies comparisons instead of intraspecies comparisons (Supplementary Figs. S17a–19a), which implies that the WGD event seems to have occurred before the divergence of *A. vulgaris* and *Lissachatina*. Moreover, the distribution of  $K_s$  of *A. vulgaris*-*Li. immaculata* gene pairs and *A. vulgaris*-*Li. fulica* gene pairs show only one peak each, respectively. The  $K_s$  values corresponding to the peak are smaller between species (*A. vulgaris*-*Li. fulica*: 1.56; *A. vulgaris*-*Li. immaculata*: 1.57) than within *A. vulgaris* (*A. vulgaris*-*A. vulgaris*: 1.61) (Fig. 3b). This result could be explained by the species differentiation event occurring shortly after the WGD event. Such a short time is reflected in our results as the peak of species differentiation coinciding with the peak of the WGD event in the  $K_s$  distribution of *A. vulgaris*-*Li. fulica* and *A. vulgaris*-*Li. immaculata*, and the overall distribution has moved towards small  $K_s$  (Fig. 3b). Assuming that the mutation rate of Mollusca is  $1.645 \times 10^{-9}$  per site per year<sup>45</sup>, we estimated the WGD event happened at approximately 93–109 Mya, and the species differentiation of *A. vulgaris* and *Lissachatina* occurred a very short time after the WGD, with molecular dating estimates for this at approximately 90–103 Mya.

After WGD, the two sets of chromosomes evolved differently with one set of chromosomes being more structurally stable and conserved compared to the other (Supplementary Figs. S17–19b), and this imbalance might provide a rich genomic resource for rapid evolution and adaptation<sup>46</sup>. Since the differentiation of Arionoidea and Achatinoidea is almost at the base of Stylommatophora differentiation<sup>47</sup>, we further speculate that all Stylommatophora species shared the common WGD event. The newly generated chromosome set provided abundant evolutionary resources in functional novelty and ecological adaptation, which may have led to the successful territorialization and diversity of Stylommatophora species.

**Evolution of gene duplication and adaptability.** Gene duplication is another important evolutionary mechanism to provide new genetical material and opportunities to acquire new gene functions for an organism<sup>48</sup>. We found that Heterobranchia species have an abundance of duplicate genes. In our analysis, between 55% (*Ely-*



**Figure 4.** Gene duplication and evolution in Heterobranchia species. **(a)** The number of gene pairs derived from different modes of duplication. **(b–f)** The  $Ka/Ks$  ratio distributions of gene pairs derived from different modes of duplication in different species. Bar graphs show the proportion of positively selected genes ( $Ka/Ks > 1$ ) that were duplicated by different mechanisms. Colors represent different species. WGD, whole-genome duplication; TD, tandem duplication; PD, proximal duplication; TRD, transposed duplication; DSD, dispersed duplication; Total, total annotated genes.

*sia chlorotica*) and 75% (*Ap. californica*, *A. vulgaris*) of genes were identified as paralogous (Fig. 4a). Three Stylommatophora species (*A. vulgaris*, *Li. fulica*, and *Li. immaculata*) have an average of 16.8% more duplicate genes than other species. Among them, WGD events contribute 13% (*Li. immaculata*) to 22% (*Li. fulica*) to existing duplicates, and WGD-derived gene pairs are the most conserved among all types of duplicated genes (Fig. 4a, b; Supplementary Table S16). Another type of duplicate gene that has increased significantly in both Stylommatophora species is transposed duplication (TRD) gene, which is on average 48 times more frequent than in other species (Fig. 4a; Supplementary Table S16). Dispersed duplication (DSD) accounts for a high proportion (mean 56% of all duplicated genes,  $SD = 24\%$ ), while proximal duplication (PD) generates a small proportion (mean 5% of all duplicated genes,  $SD = 1.8\%$ ) of gene copies in all Heterobranchia species. Strikingly, tandem duplication (TD) gene pairs account for the highest proportion in *Ap. californica*, which is about 1.7–5.3 times that of other species (Fig. 4a; Supplementary Table S16).

The evolutionary pattern of duplicated genes is similar within three Stylommatophora species. The overall age of duplicated genes is young and under a weak purifying selection ( $Ka/Ks < 1$ ) in both duplicated modes compared to other species (Fig. 4b–f; Supplementary Fig. S20). For *Ap. californica*, the TD- and PD-derived gene pairs have similar distribution of  $Ka/Ks$  (mode: TD-0.30; PD-0.35) between Stylommatophoran species (mean of the modes: TD-0.32; PD-0.36) when compared to other more closely related species (mean of the modes: TD-0.15; PD-0.17) suggesting that tandem and proximal duplicates happened recently and experienced relatively relaxed purifying selection (Fig. 4e, f). However, DSD- and TRD- derived gene pairs are more conserved in *Ap. californica*, which is more similar to *Radix auricularia*, *Biomphalaria glabrata*, and *E. chlorotica* (Fig. 4c, d; Supplementary Fig. S20).

We further explored the roles of positive selection ( $Ka/Ks > 1$ ) in the evolution of duplicated genes in seven Heterobranchia species. As expected, *A. vulgaris*, *Li. fulica* and *Li. immaculata* experienced stronger positive selection than other species, reflected by the high percentages of gene pairs showing  $Ka/Ks > 1$  in all kinds of duplicated gene pairs (Fig. 4c, d; Supplementary Table S17). Among all duplicate genes, TD-, PD-, DSD- derived gene pairs have experienced stronger positive selection compared with genes generated by other duplication mechanisms (Fig. 4b–f; Supplementary Table S17). In *A. vulgaris*, 24% TD- derived genes were likely positively selected, which is 2–3 times that of *Li. immaculata* and *Li. fulica*, and 4–29 times that of other species (Fig. 4c, e, f; Supplementary Table S17). Interestingly, we found that the functional enrichment of genes caused by TD in *A. vulgaris* concerns response to external stress, pigment catabolism, and acute inflammatory process, which

echoes the previous enrichment results of *A. vulgaris* unique and expanded genes and is related to its unique adaptation (Supplementary Fig. S21, 22). On the other hand, only 0.3% WGD—which only exists in *A. vulgaris*, *Li. fulica*, and *Li. immaculata*, and 1.6% TRD—which are highly expanded in *A. vulgaris*, *Li. fulica*, and *Li. immaculata*, Fig. 4a) derived gene pairs were likely positively selected in *A. vulgaris* (Fig. 4b, d; Supplementary Table S17). The TRD-derived gene pairs which were functionally enriched mostly refer to cell components (Supplementary Fig. S23), and WGD-derived genes were prone to be enriched in basic biological functions such as signal transduction, ion transport, muscle development (Supplementary Fig. S24).

**Massive expansion of transposable elements in *A. vulgaris* genome.** Repeat content analysis showed that the repeat sequences occupy approximately 75.09% (1.15 Gb) of the *A. vulgaris* assembly (Supplementary Table S18), which is the highest value among all studied gastropod species<sup>23</sup>. We also found that species in Heterobranchia have a higher repeat content than other gastropod groups (i.e., Caenogastropoda, Vetigastropoda, Neomphalina, and Patellogastropoda) (Supplementary Fig. S25). In all types of repetitive sequences, transposable elements (TEs) account for 61.08% of the *A. vulgaris* assembly, and among them, long interspersed elements (LINEs), DNA transposons (DNAs), and short interspersed elements (SINEs) account for 36.39%, 5.44%, 1.78% of the assembly, respectively.

A high proportion of unclassified TEs (17.76%) was also detected in the *A. vulgaris* genome (Fig. 5a; Supplementary Table S18). Overall, the composition of TEs of *A. vulgaris* is similar to *Li. fulica* and *Li. immaculata*, in which LINEs are dominant, whereas in other Heterobranchia species DNA transposons are most abundant (except *B. glabrata*, see below). Most of the LINEs in *A. vulgaris* showed a low divergence rate, indicating a recent explosion of LINEs in the *A. vulgaris* genome (peak % divergence to consensus = 3). However, *Li. fulica* and *Li. immaculata* LINEs were not recent invaders since they exhibit a large divergence from the consensus (the distributions peak at 31% divergence for *Li. fulica* and 33% for *Li. immaculata*) (Fig. 5a). Two freshwater snails (*R. auricularia*, *B. glabrata*) and *Ap. californica* also showed recent expansion of LINEs, which even resulted in LINEs that replaced DNAs and became the dominant TE type in the *B. glabrata* genome (Fig. 5a). We found that although the total TE number of *A. vulgaris* is 1.35–6.09 times greater than in the other species considered, the insertion of TEs was very conservative. Specifically, genes with TEs distributed in putatively functional regions, i.e., 2 kb upstream, 1 kb downstream, or intron, exon regions in *A. vulgaris* were 1.21–1.87 times that of all other species. However, the number of TEs inserted into exons in *A. vulgaris* only accounts for 51% and 66% of that of *Li. fulica* and *B. glabrata*, respectively (Fig. 5b; Supplementary Table S19). Among all species, TEs were mainly inserted into introns in different degrees of divergence from consensus (Fig. 5b; Supplementary Fig. S26). The insertion of *A. vulgaris* TEs in intron regions greatly increased compared to other species (1.29–6.74 times), especially young TEs with a low divergence rate (%divergence to consensus < 16, Supplementary Fig. S26). The insertion into upstream and downstream is also increased, by 1.81–3.72 and 2.06–5.39 times that of other species, respectively. Previous reports have shown that TEs are powerful facilitators of rapid adaptation to novel environments<sup>49–51</sup>. The recent expansion of LINEs in *A. vulgaris* may also have played an important role in promoting potential plasticity and stress resistance correlated with its invasiveness and competitiveness.

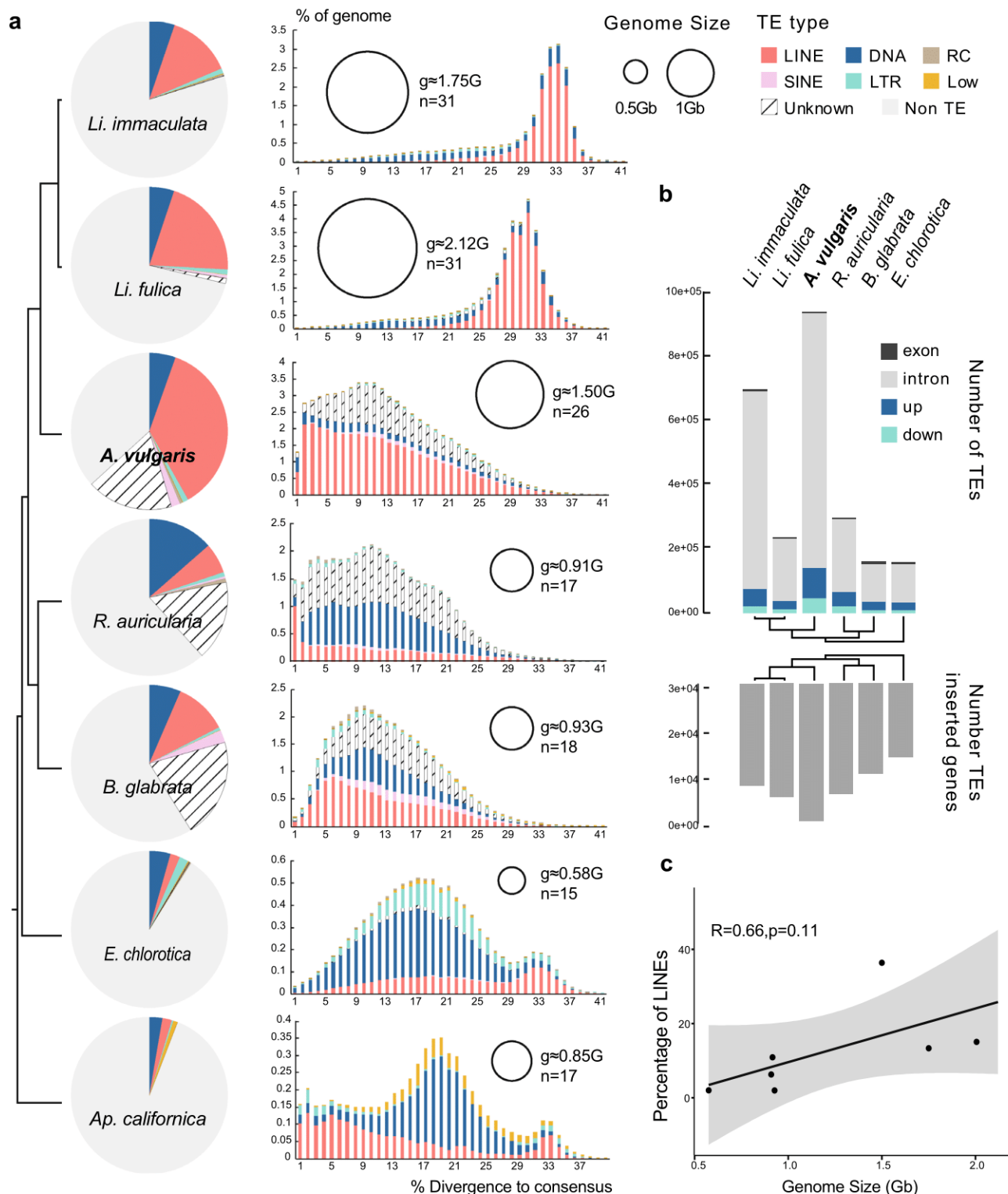
Recent studies showed that TEs have driven massive changes in genome size<sup>52–54</sup>. In our results, although we found TE coverage is slightly positively correlated with genome size, the correlation is not significant (Supplementary Fig. S27a). In further analyses, we determined that these positive contributions all come from the LINEs (Fig. 5c, Supplementary Fig. S27b–d), but are still not significantly related to genome size. However, *A. vulgaris* and *Lissachatina* have larger genome sizes compared to other species, thus we assume that the changes in the Heterobranchia species genome size might be the result of the expansion of LINEs and the WGD event.

**Population dynamics of *A. vulgaris*.** We observed an average genome-wide heterozygosity rate of 1.55 per hundred base pairs in *A. vulgaris*, which is about three times of the invasive land snail, *Li. fulica* (0.47 per hundred base pair)<sup>55</sup>, but comparable to freshwater snails *Pomacea canaliculata* (Caenogastropoda, 1.41%) and *P. maculate* (1.22%) which are also notable invaders<sup>56</sup>. We further compared the population dynamic history of *A. vulgaris* with relatively closely related invasive species *Li. fulica* and *B. glabrata*. We found that *A. vulgaris* and *B. glabrata* populations exhibited similar demographic histories, with a high  $N_e$  ( $4 \times 10^5$ ) ~ 1.2 Mya and both increased between 1.2 and 0.8 Mya (Fig. 6). The *A. vulgaris* population continuously declined after the Pre-Pastonian glaciation and dramatically decreased ~ 40,000 years ago, which is consistent with the sharp temperature drop. *Li. fulica* population shows a relatively small  $N_e$  ( $3 \times 10^5$ ) ~ 1.2 Mya and a continuously prolonged decline until ~ 10,000 years ago, and then the population increased significantly to almost the initial level (Fig. 6). The very recent expansion of *Li. fulica* from a relatively small effective population size can also explain that the heterozygosity of *Li. fulica* is much smaller than that of *A. vulgaris*. On the other hand, the relatively long-term large effective population size of *A. vulgaris* may cause the complexity in its population structure, thereby increasing the difficulty of research on population expansion/invasion studies<sup>22,57,58</sup>.

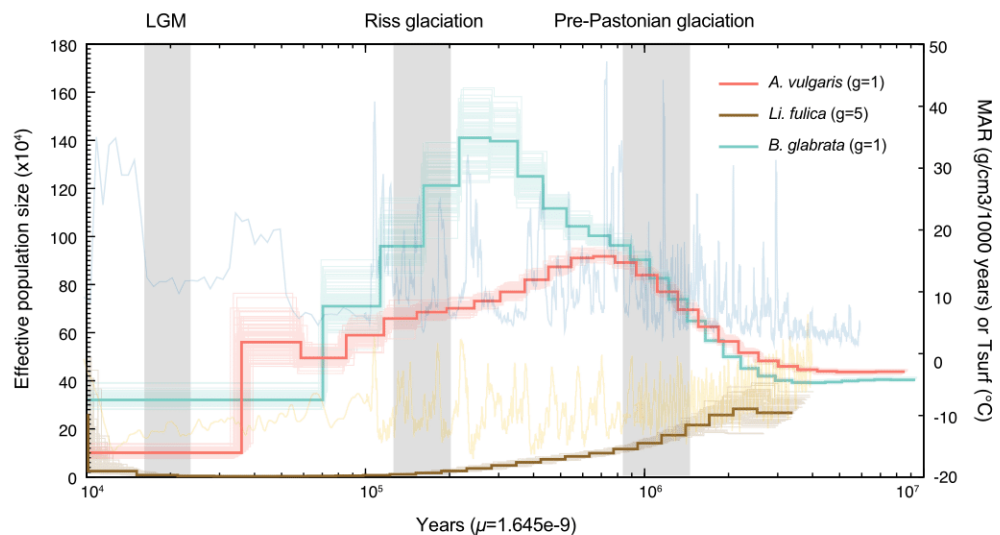
## Discussion

Whole-genome duplication (WGD) is a common phenomenon in plants and has been shown in invertebrate species<sup>59,60</sup>. It plays an important role in providing evolutionary novelties and promoting speciation<sup>43,61</sup>. Based on chromosome-level genomic analysis of two land snails, Liu et al.<sup>11</sup> first reported the WGD on the Sigmurethra-Orthurethra branch within Stylommatophora at ~ 70 Mya. However, our results indicate that the WGD is most likely an event shared by all Stylommatophora species, which we have dated back to 90–103 Mya (Fig. 3). The inconsistency in timing inference may be caused by the identification of paralogous gene pairs derived by the WGD event. In the study of Liu et al., MCScanX<sup>62</sup> was used with default parameters to identify the collinearity blocks in *Li. immaculata* and *Li. fulica* and the  $K_s$  distribution was calculated using the gene pairs in the





**Figure 5.** Massive expansion of transposable elements (TEs) in *Arion vulgaris* genome. **(a)** Left: TE composition by class among Heterobranchia species. Right: TE accumulation history corresponding to each species. The size of the circle indicates the estimated genome size of each species. n represents the number of haploid chromosomes; g represents the estimated genome size. **(b)** Bar graph shows the number of TEs in different genic regions and the number of TEs-inserted genes in each species. **(c)** Relationship between LINES coverage and genome size in Heterobranchia species. Numbers on linear regressions correspond to adjusted  $r^2$  coefficients (Pearson's test).



**Figure 6.** Demographic histories of *Arion vulgaris*, *Lissachatina (Achatina) fulica* and *Biomphalaria glabrata*. Solid bold lines indicate inferred ancestral effective population sizes for three species, while background blue and yellow lines represent the mass accumulation rate (MAR) and the atmospheric surface air temperature (Tsurf) relative to the present, respectively.

collinearity blocks. In our initial analysis, we used the same method as Liu et al., described. We did observe  $K_s$  peaks shared by *A. vulgaris* and *Li. fulica* which represents the WGD event, however, the  $K_s$  distribution of *Li. immaculata* has a relatively large deviation (Supplementary Fig. S28). From our results of the syntenic dot plots, *Li. immaculata* exhibits lower synteny in comparisons to *A. vulgaris* than the *Li. fulica*-*A. vulgaris* comparison (Supplementary Figs. S17, 18), implying that *Li. immaculata* has experienced more genome reconfiguration and chromosome rearrangement, and this may increase errors and difficulty in the identification of collinearity gene pairs within *Li. immaculata*. We addressed this problem by implementing WGDI<sup>63</sup>, a new tool which can identify collinearity more accurately and comprehensively. In the  $K_s$  distribution obtained by WGDI, the three species have relatively consistent  $K_s$  peaks (Fig. 3b). Therefore, we suppose that the estimation based on  $K_s$  distribution derived from WGDI can more accurately represent the older than expected time of the WGD event.

In our results, after WGD, the extra chromosome copy shows a release of selective pressure with large structural variations and increased synonymous mutation rate (Supplementary Figs. S17–19), which might serve as an abundant resource for mutations and novo functions, and may have facilitated the stylommatophoran transition from water to land. For example, expansion genes derived by WGD duplications are enriched in nerve and muscle development, which might enhance the locomotion and movement ability in terrestrial environments (Supplementary Fig. S24). Moreover, we also detected genes related to kidney development, response to stimulation, radiation, larval development, and dietary habits that were expanded in both *A. vulgaris* and shelled land pulmonates (Fig. 2d). These genes might have contributed to the stylommatophoran ancestor's ability to overcome challenges such as gravitational pressure and water loss brought by the terrestrial environment.

The split of *A. vulgaris* and *Lissachatina* land snail lineages happened in a very short time after WGD (Fig. 3b). However, the slug *A. vulgaris* has evolved its unique adaptability in further improving water re-absorption and resistance to external stimuli (Fig. 2d). For example, a series of interleukin genes were positively selected in *A. vulgaris* genome, which might enhance the immune response; genes related to acute inflammatory processes were expanded, which might improve the innate defense; the expansion of genes related to regeneration might help to quickly recover from body/organ damage and increase the survival rate, and the expansion of genes and pathways in pigment biosynthesis might protect *A. vulgaris* from solar radiation.

In addition to WGD, small scale gene duplication also plays an important role in providing new genetic material for mutation, drift, and selection<sup>61</sup>. We found Heterobranchia species have an abundance of duplicated genes. Among all types of duplicated genes, the largest category is dispersed duplication (DSD) genes (Fig. 4). The proportion of tandem duplication (TD) genes has greatly increased in *A. vulgaris* as compared to other Heterobranchia species, and ~24% of them were positively selected ( $K_a/K_s > 1$ ). Enrichment analysis showed the functions of TD derived genes largely overlapped with *A. vulgaris* expansion gene functions, e.g., response to external stress, pigment catabolism, acute inflammation, which thus implies that tandem duplication of genes might be one of the forces driving evolution, adaptation, and potential invasiveness of *A. vulgaris*.

Previous studies have shown that transposable element (TE) insertions play a critical role in rapid phenotypic variation and might help invasive species to successfully adapt to a novel environment<sup>49</sup>. In our results, the recent massive expansion of TEs (more precisely, LINES) in *A. vulgaris* might act as potent insertional mutagens, greatly enhancing the adaptive success, invasiveness, and the ability to outcompete other land slugs.

All in all, our genomic analysis reveals the powerful potential of *A. vulgaris* for adaptation and evolution, which may explain why *A. vulgaris* is considered as an invasive species in central Europe. However, there is ongoing controversy about its native range and invasiveness. According to the record of first discovery in many

European countries, it was believed that the slug originated on the Iberian Peninsula and expanded its range into central and eastern Europe over the last five decades<sup>12</sup>. However, the very similar external appearance with other closely related native arionids as well as (potential) hybrid species between *A. vulgaris*, *A. ater*, and *A. rufus*<sup>64–66</sup>, might have caused the misidentification of *A. vulgaris*, obscured the specimen records, and made it difficult to trace its origin and monitoring the spread only by morphological identification<sup>66,67</sup>. Recent studies based on the genetic diversity patterns of mitochondrial and nuclear loci suggested that *A. vulgaris* is native in central Europe rather than alien/invasive while probably invasive in other parts of Europe<sup>20–22</sup>. Our *A. vulgaris* individual was collected in Munich, Southern Germany and has a relatively rich genetic diversity, which implies a large effective population size. This result seems to support the point of view that *A. vulgaris* is more likely native rather than invasive at least in south Germany (Fig. 6). However, more robust conclusions still require extensive sampling and more population data. Our high-quality *A. vulgaris* genome will promote future population studies from the use of single/multiple molecular markers to the use of whole genome-wide polymorphism and will help us to understand its origin, expansion, and potential invasiveness more comprehensively.

## Materials and methods

**Sample collection and sequencing.** An adult *A. vulgaris* was collected in the garden of the Zoologische Staatssammlung München, Germany. Genomic DNA was extracted from the foot muscle tissue with MagAttract HMW DNA Kit and CTAB method<sup>69</sup>. Quality was checked using agarose gel electrophoresis. Four different sequencing technologies were used to obtain the genome sequence (Supplementary Table S1). First, one Illumina paired-end sequencing library was generated following the manufacturer's standard protocol (Illumina) with an insert size of 350 bp. Also, high molecular weight DNA was separated and loaded onto the 10X Genomics Chromium microfluidics controller for barcoding and generated two 10X Genomics linked-read libraries with an insert size of 350 bp. Those reads not only provided the long-range positional information to assemble contigs into scaffolds but were also used for the genome survey analysis and final base-level genome sequence correction<sup>70</sup>. One Hi-C library digested with MboI and with an insert size of 350 bp was constructed for providing long-range information on the grouping and linear organization of sequences along entire chromosomes to assemble the scaffolds into chromosome-level scaffolds<sup>71</sup>. The Illumina paired-end sequencing library, 10X Genomics linked-read libraries, and Hi-C library were sequenced on an Illumina HiSeqX Ten platform (Illumina, San Diego, CA, USA) with 150 bp paired-end reads. The raw reads generated by Illumina HiSeqX Ten platform were all filtered with the following criteria: reads with adapters, reads with N bases more than 5%, and reads with more than 65% of low-quality bases ( $\leq$  seven) using Fastp v0.20.0<sup>72</sup>. Meanwhile, Nanopore libraries were prepared using SQK-LSK109 kit and sequenced in the platform Nanopore PromethION (Oxford Nanopore Technologies). We performed a base calling of the raw Nanopore data with Guppy v2.2.3<sup>73</sup>.

Total RNA was extracted from the 'head' part of the sample which includes tentacles, mantle, inner head and anterior visceral organs, and foot and sequenced on an Illumina NovaSeq platform with paired-end 150 bp.

**Genome feature estimation and assembly.** The genome size and heterozygosity were estimated by GenomeScope v1.0.0<sup>74</sup> using the quality-controlled paired-end Illumina sequence data and linked reads. We combined reads generated using different sequencing platforms to generate a high-quality de novo genome assembly (Supplementary Table S2). Specifically, long reads, generated with the Nanopore PromethION platform, were assembled into contigs using the wtdbg2 v2.2 assembler<sup>75</sup>. The contigs were subsequently polished by ntEdit v1.3.1<sup>76</sup> using Illumina short reads and linked reads. The resulting contigs were then connected into scaffolds by 10X Genomics linked-read data using Scaff10X v4.2<sup>77</sup>. Hi-C reads were mapped to the draft assembly and processed using the hicstuff v2.2.2 pipeline<sup>78</sup> with the parameters --aligner bowtie2 --enzyme MboI --iterative --matfmt graal --quality-min 30 --size 0. We ran instaGRAAL v0.1.2<sup>79</sup> on the resulting matrix and the draft assembly with parameters --level 5 --cycles 100 --coverage-std 1 --neighborhood 5 and the module instagraal-polish for refinement. After building the interaction map of the final scaffolds with hicstuff, we noticed an intra-chromosomal translocation on chromosome 9 which could have been due to a misassembly. In the subsequent analysis, we mapped all reads to the assembly 'chromosome 9' and identified two breakpoints (at site 21,000,000 and 27,600,000 respectively) based on the read' depth and gene distribution. We corrected the orders manually and reconnected sequences with 10 N's at the new junction sites.

**Gene prediction.** Protein-coding genes were predicted using the following approaches: ab initio prediction, homology-based prediction, and transcriptome-based prediction. For ab initio prediction, RNA-seq reads were first aligned to the *A. vulgaris* genome sequence using STAR v2.7.2b<sup>80</sup>, then the read alignment information was merged and used for Braker2 v2.1.5<sup>81</sup> gene prediction pipeline. For homology-based prediction, we selected six gastropods from closely to distantly related to *A. vulgaris*, namely *Li. fulica*, *B. glabrata*, *Ap. californica*, *E. chlorotica*, *P. canaliculata*, *Haliothis rufescens* (Supplementary Table S3). The protein sequences of the six species were downloaded from NCBI and aligned against the assembled genome with MMseqs v11.e1a1c<sup>82</sup>. These results were then combined into gene models separately with GeMoMa v1.3.1<sup>83</sup> using mapped RNA-seq data for splice site identification. The resulting gene annotation sets were further filtered using the GeMoMa module GAF with default parameters. For the transcriptome-based prediction, RNA-seq data had been assembled using both de novo and genome-guided approaches with Trinity vr20140413p1<sup>84</sup>, and the gene predictions were carried out with PASA v2.0.2<sup>85</sup>. All gene annotations were combined with EVM v1.1.1<sup>86</sup> (Supplementary Table S4). Partial genes and genes with a coding length of less than 150 bp were removed from further analysis.

The predicted genes were functionally annotated by aligning them to the eggNOG<sup>27</sup>, SWISS-PROT<sup>28</sup>, TrEMBL<sup>28</sup>, KEGG<sup>30</sup>, and InterPro<sup>29</sup> databases using BLAST v2.2.31<sup>87</sup> with a maximal e-value of 1e–5 and by

aligning to the Pfam database using HMMer v3.0<sup>88</sup>. Gene Ontology (GO) terms (Gene Ontology, RRID:SCR002811) were assigned to the genes using the BLAST2GO v2.5 pipeline<sup>89</sup>.

**Gene family cluster and terrestrial adaptation analysis.** To resolve the early phylogeny of gastropods, we selected the species according to the following rules: (1) coverage of as many subclasses as possible; (2) the lineage diversity within each subclass should be covered; (3) in case of closely related species, those with high-quality genomes or better gene annotations were preferred. As a result, fourteen Gastropoda species were selected, including six Heterobranchia: *Li. fulica*, *Li. immaculata*, *B. glabrata*, *R. auricularia*, *Ap. californica*, *E. chlorotica*; four Caenogastropoda: *P. canaliculata*, *Marisa cornuarietis*, *Lanistes nyassanus*, *Conus consors*; one Vetigastropoda: *H. rufescens*; one Neomphalina: *Chrysomallon squamiferum*, and one Patellogastropoda: *Lottia gigantea*. Two bivalve species: *Argopecten purpuratus* and *Saccostrea glomerata* were selected as outgroups (Supplementary Table S3). Protein sequences were extracted from each species and an all-against-all comparison was performed using BLASTP v2.9.0<sup>90</sup> with an e-value cut-off of  $1e-5$ . OrthoFinder v2.4.0<sup>91</sup> was used to cluster gene families.

Based on the clustered gene families, we explored the terrestrial adaptation of *A. vulgaris* from two aspects. One is the genetic basis of adaptability shared by stylommatophoran species relative to other aquatic or marine Heterobranchia species, another is the specific adaptations of shell-less *A. vulgaris* compared to two land snails, *Li. fulica* and *Li. immaculata*. For both cases, we tested lineage/species specific genes, expansion/contraction genes, and positively selected genes (PSGs) and performed Gene Ontology (GO) enrichment analysis.

We applied the CAFÉ v4.2.1<sup>92</sup> program to examine gene family expansion and contraction across entire genomes with default parameters. To identify PSGs, OrthoFinder v2.4.0<sup>91</sup> was used to cluster gene families from five Heterobranchia species: *Li. fulica*, *Li. immaculata*, *B. glabrata*, *R. auricularia*, *Ap. californica* (Supplementary Table S3). Single-copy orthologous genes were extracted based on the results of clustered gene families. MAFFT v7.455<sup>93</sup> was used for multiple sequence alignments and converted to codon sequences by PAL2NAL v14<sup>94</sup>. Poorly aligned positions were removed with Gblocks v0.91b<sup>95</sup> with parameters “-b2 = 85% alignment length -b3 = 6 -b4 = 10 -b5 = h -t = c”. The PSGs were identified by comparing the null model (fix\_omega = 1) to the alternative model (fix\_omega = 0) using codeml branch-site model in the PAML package<sup>96</sup>. The foreground branch was set to (1) the node of the most common ancestor of *A. vulgaris*, *A. vulgaris* and *Li. immaculata*, to identify putative PSGs shared by stylommatophoran species, and (2) *A. vulgaris*, for the detection of potential PSGs of *A. vulgaris*. Chi-square tests were performed for each pair and genes with a 5% significance level were selected as putative PSGs<sup>96</sup>. Cytoscape v3.8.2<sup>97</sup> was used for visualizing molecular interaction networks and biological pathways.

**Phylogenetic analysis.** Gene families with only one copy from each of 16 species were selected as single-copy genes and were concatenated and aligned by MUSCLE v3.8.1551<sup>98</sup> with default parameters. The maximum likelihood (ML) trees were inferred using both RAxML v8.2.8<sup>99</sup> with the GTR+Γ model and IQ-TREE v1.6.9<sup>100</sup>, which automatically selected the best-fit substitution model using ModelFinder<sup>101</sup>. For coalescent-based analysis, gene trees were first estimated using RAxML v8.2.8<sup>99</sup> with 100 replicates from each single copy gene. The best tree was then selected as input to ASTRAL v5.6.1<sup>102</sup> to infer the species tree with default parameters. Gene trees were visualized using DensiTree v2.01<sup>103</sup>.

Divergence time was computed using the MCMCTREE program implemented in the PAML v4.8<sup>96</sup> package. For calibration, we used the soft bounds of Euopisthobranchia—Panpulmonata (divergence time between 190 and 270 MY)<sup>33</sup>, the fossil of Sublitoidea (418 MY) constraints on the node of Heterobranchia and Caenogastropoda<sup>104</sup>, and the fossil of *Fordilla troyensis* (530 MY) for the root<sup>105,106</sup>.

**Identification of whole-genome duplication event.** For macrosynteny analysis, LASTZ v1061<sup>107</sup> was used to perform whole-genome alignments between chromosome-level assemblies of *A. vulgaris*, *Li. immaculata*<sup>11</sup> and *Ap. californica*<sup>108</sup>. The alignments were visualized by Circos v0.69-6<sup>109</sup>. For synteny analysis of homologous gene pairs, the protein sequences of *A. vulgaris*, *Li. fulica* and *Li. immaculata* were first searched against themselves and also between species using BLASTP v2.9.0<sup>90</sup>, then subjected to WGDI v0.5.1<sup>63</sup> to determine syntenic blocks, estimate  $K_s$  values for each block and calculate  $K_s$  distributions of gene pairs in collinearity blocks. Curves were fitted using the Gaussian approximation function in the WGDI package.

**Identifying gene duplications.** The different modes of gene duplication were identified using the Dup-Gen\_finder v1.07 pipeline<sup>110</sup> using *P. canaliculata* as a reference<sup>56</sup>. Gene pairs were further filtered to remove overlaps between different duplicate modes. For each duplicated gene pair, we aligned their protein sequences using MAFFT v7.455<sup>93</sup> and converted the protein alignment into a codon alignment using PAL2NAL v14<sup>94</sup>. Then, the resulting codon alignment was formatted into an AXT format and  $K_a$  (number of substitutions per nonsynonymous site) and  $K_s$  (number of substitutions per synonymous site) values were calculated by KaKs\_Calculator v2.0<sup>111</sup>.

**Repeat prediction and expansions of transposable elements.** TRF v4.09<sup>112</sup> was used for tandem repeats identification with default parameters. Transposable elements (TEs) were annotated using a combination of ab initio and homology-based approaches. First, repeat elements were identified de novo using RepeatModeler v2.0.1<sup>113</sup>. The database predicted by RepeatModeler, the RepBase<sup>114</sup> (RepBase-20170127) and the Dfam<sup>115</sup> (Dfam\_Consensus-20170127) libraries were then merged together and used as a custom library for RepeatMasker v4.0.7<sup>113</sup> to identify repeats comprehensively. The repeat divergence rate was measured by the percentage of substitutions in the corresponding regions between annotated repeats and consensus sequences in the RepBase database. For species with incomplete TE annotations (e.g., *Li. fulica*, *R. auricularia*), we predicted their TEs

using the same approaches as just described. We regarded genes with TEs inserted in introns, exons or with 2-kb upstream or 1 kb downstream of the terminal exons as likely to be affected by these insertions and compared the number of genes affected by TEs in different insertion regions.

**Genome heterozygosity and reconstruction of effective population size ( $N_e$ ).** Heterozygosity was estimated by the following steps. First, the clean Illumina reads and linked reads were merged and mapped onto the *A. vulgaris* assembly by BWA-MEM v0.7.17-r1188<sup>116</sup> with default parameters. The sequence alignment/map (SAM) file format was processed using SAMtools v1.9<sup>117</sup>, and Picard v2.23.3<sup>118</sup> was used to mark duplicates. Finally, single nucleotide polymorphisms (SNPs) calling was implemented in GATK v4.1.6.0<sup>119</sup> using default parameters, and several filtering steps were performed to reduce false positives, including (1) remove SNPs with more than two alleles; (2) remove SNPs with a quality score less than 30; (3) remove SNPs at or within 5 bp from any InDels; (4) remove sites with extremely low (less than one-third average depth) or extremely high (more than three-fold average depth) coverage.

We inferred the demographic history by applying the Pairwise Sequentially Markovian Coalescence model using PSMC v0.6.5-r67<sup>120</sup> with the following parameters: -N25 -t15 -r5 -p '4 + 25 × 2 + 4 + 6'. This method reconstructs the history of changes in population size over time using the distribution of the most recent common ancestor (tMRCA) between two alleles in an individual. The generation time of *A. vulgaris* and *B. glabrata* was assumed to be 1 year<sup>121,122</sup> and *Li. fulica* was assumed to be 5 years<sup>123</sup>.

**Ethics declarations.** No specific permits were required for the described field studies, no specific permissions were required for these locations/activities, and the field studies did not involve endangered or protected species.

### Data availability

The *A. vulgaris* genome project of this study was deposited at the National Center for Biotechnology Information (NCBI) under BioProject number PRJNA680311. Genomic and transcriptome sequence reads was deposited in the SRA database with BioSample: SAMN16874494. The assembled genome had been deposited at GenBank with accession number: GCA\_020796225.1. In addition, the genome annotation files had been submitted at the Figshare: <https://doi.org/10.6084/m9.figshare.15022212.v1>; <https://doi.org/10.6084/m9.figshare.15022203.v1>.

Received: 19 February 2022; Accepted: 5 August 2022

Published online: 20 August 2022

### References

- Barker, G. M. In *The Biology of Terrestrial Molluscs* (ed. Braker, G.M.) 20 (CABI Publishing, England, 2001).
- Kano, Y., Brenzinger, B., Nützel, A., Wilson, N. G. & Schrödl, M. Ringiculid bubble snails recovered as the sister group to sea slugs (Nudipleura). *Sci. Rep.* **6**, 30908 (2016).
- Klussmann-Kolb, A., Dinapoli, A., Kuhn, K., Streit, B. & Albrecht, C. From sea to land and beyond—new insights into the evolution of euthyneuran Gastropoda (Mollusca). *BMC Evol. Biol.* **8**, 57 (2008).
- Romero, P. E., Weigand, A. M. & Pfenninger, M. Positive selection on panpulmonate mitogenomes provide new clues on adaptations to terrestrial life. *BMC Evol. Biol.* **16**, 164 (2016).
- Romero, P. E., Pfenninger, M., Kano, Y. & Klussmann-Kolb, A. Molecular phylogeny of the Ellobiidae (Gastropoda: Panpulmonata) supports independent terrestrial invasions. *Mol. Phylogenet. Evol.* **97**, 43–54 (2016).
- Baur, B. Parental care in terrestrial gastropods. *Experientia* **50**, 5–14 (1994).
- Takei, Y. From aquatic to terrestrial life: evolution of the mechanisms for water acquisition. *Zool. Sci.* **32**, 1–7 (2015).
- Mordan, P. & Wade, C. In *Phylogeny and Evolution of the Mollusca* (eds. Ponder, W. & Lindberg, D.R.) 409–426 (University of California Press, Cambridge, 2008).
- Ohta, T. Gene conversion and evolution of gene families: an overview. *Genes* **1**, 349–356 (2010).
- Wang, K. *et al.* African lungfish genome sheds light on the vertebrate water-to-land transition. *Cell* **184**, 1362–1376 (2021).
- Liu, C. *et al.* Giant African snail genomes provide insights into molluscan whole-genome duplication and aquatic-terrestrial transition. *Mol. Ecol. Resour.* **21**, 478–494 (2020).
- Zajac, K. S., Gaweł, M., Filipiak, A. & Kramarz, P. *Arion vulgaris* Moquin-Tandon, 1855—the aetiology of an invasive species. *Folia Malacol.* **25**, 81–93 (2017).
- Frank, T. Slug damage and number of slugs (Gastropoda: Pulmonata) in winter wheat in fields with sown wildflower strips. *J. Molluscan Stud.* **64**, 319–328 (1998).
- Gren, I. M., Isacs, L. & Carlsson, M. Costs of alien invasive species in Sweden. *Ambio* **38**, 135–140 (2009).
- Kozłowski, J. The significance of alien and invasive slug species for plant communities in Agrocenoses. *J. Plant Prot. Res.* **52**, 67–76 (2012).
- Gismervik, K. *et al.* Invading slugs (*Arion vulgaris*) can be vectors for *Listeria monocytogenes*. *J. Appl. Microbiol.* **118**, 809–816 (2015).
- Slotsbo, S. *et al.* Cold tolerance and freeze-induced glucose accumulation in three terrestrial slugs. *Comp. Biochem. Physiol. Part A Mol. Integr. Physiol.* **161**, 443–449 (2012).
- Frank, T. Influence of slug herbivory on the vegetation development in an experimental wildflower strip. *Basic. Appl. Ecol.* **4**, 139–147 (2003).
- Rabitsch, W. DAISIE—Delivering Alien Invasive Species Inventories for Europe. <http://www.europe-aliens.org> (2006).
- Pfenninger, M., Weigand, A., Balint, M. & Klussmann-Kolb, A. Misperceived invasion: The Lusitanian slug (*Arion lusitanicus* auct. non-Mabille or *Arion vulgaris* Moquin-Tandon 1855) is native to Central Europe. *Evol. Appl.* **7**, 702–713 (2014).
- Zajac, K. S. *et al.* A comprehensive phylogeographic study of *Arion vulgaris* Moquin-Tandon, 1855 (Gastropoda: Pulmonata: Arionidae) in Europe. *Org. Divers. Evol.* **20**, 37–50 (2019).
- Zemanova, M. A., Knop, E. & Heckel, G. Phylogeographic past and invasive presence of *Arion* pest slugs in Europe. *Mol. Ecol.* **25**, 5747–5764 (2016).
- Gomes-dos-Santos, A., Lopes-Lima, M., Castro, L. F. C. & Froufe, E. Molluscan genomics: The road so far and the way forward. *Hydrobiologia* **847**, 1705–1726 (2019).
- Beeson, G. E. Chromosome numbers of slugs. *Nature* **186**, 257–258 (1960).

25. Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
26. Mapleson, D., Garcia-Accinelli, G., Kettleborough, G., Wright, J. & Clavijo, B. J. KAT: A K-mer analysis toolkit to quality control NGS datasets and genome assemblies. *Bioinformatics* **33**, 574–576 (2017).
27. Huerta-Cepas, J. *et al.* eggNOG 50: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **47**, D309–D314 (2019).
28. Bairoch, A. & Apweiler, R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* **28**, 45–48 (2000).
29. Hunter, S. *et al.* InterPro: The integrative protein signature database. *Nucleic Acids Res.* **37**, D211–215 (2009).
30. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
31. Cunha, T. J. & Giribet, G. A congruent topology for deep gastropod relationships. *Proc. R. Soc. B* **286**, 20182776 (2019).
32. Uribe, J. E., Irisarri, L., Templado, J. & Zardoya, R. New patellogastropod mitogenomes help counteracting long-branch attraction in the deep phylogeny of gastropod mollusks. *Mol. Phylogenet. Evol.* **133**, 12–23 (2019).
33. Zapata, F. *et al.* Phylogenomic analyses of deep gastropod relationships reject Orthogastropoda. *Proc. R. Soc. B* **281**, 20141739 (2014).
34. Haszprunar, G. On the origin and evolution of major gastropod groups, with special reference to the Streptoneura. *J. Molluscan Stud.* **54**, 367–441 (1988).
35. Ponder, W. F. & Lindberg, D. R. Gastropod phylogeny—challenges for the 90s. In *Origin and evolutionary radiation of the mollusca*. (ed. Taylor, J.D.) 135–154 (Oxford University Press, Oxford, 1996).
36. Ponder, W. F. & Lindberg, D. R. Towards a phylogeny of gastropod molluscs: An analysis using morphological characters. *Zool. J. Linn. Soc.* **119**, 83–265 (1997).
37. Chen, Z. & Schrödl, M. How many single-copy orthologous genes from whole genomes reveal deep gastropod relationships?. *PeerJ* **10**, e13285 (2022).
38. Doğan, O., Schrödl, M. & Chen, Z. The complete mitogenome of *Arion vulgaris* Moquin-Tandon, 1855 (Gastropoda: Stylommatophora): Mitochondrial genome architecture, evolution and phylogenetic considerations within Stylommatophora. *PeerJ* **8**, e8603 (2020).
39. Klobucar, G., Lajtner, J. & Erben, R. Increase in number and size of kidney concretions as a result of PCP exposure in the freshwater snail *Planorbis corneus* (Gastropoda, Pulmonata). *Dis. Aquat. Organ* **44**, 149–154 (2001).
40. Barnes, P. J. How corticosteroids control inflammation: Quintiles prize lecture 2005. *Br. J. Pharmacol.* **148**, 245–254 (2006).
41. Cain, D. W. & Cidlowski, J. A. Immune regulation by glucocorticoids. *Nat. Rev. Immunol.* **17**, 233–247 (2017).
42. Munck, A., Guyre, P. M. & Holbrook, N. J. Physiological functions of glucocorticoids in stress and their relation to pharmacological actions. *Endocr. Rev.* **5**, 25–44 (1984).
43. Van de Peer, Y., Mizrachi, E. & Marchal, K. The evolutionary significance of polyploidy. *Nat. Rev. Genet.* **18**, 411–424 (2017).
44. Hallinan, N. M. & Lindberg, D. R. Comparative analysis of chromosome counts infers three paleopolyploidies in the mollusca. *Genome Biol. Evol.* **3**, 1150–1163 (2011).
45. Allio, R., Donega, S., Galtier, N. & Nabholz, B. Large variation in the ratio of mitochondrial to nuclear mutation rate across animals: Implications for genetic diversity and the use of mitochondrial DNA as a molecular marker. *Mol. Biol. Evol.* **34**, 2762–2772 (2017).
46. Gillard, G. B. *et al.* Comparative regulomics supports pervasive selection on gene dosage following whole genome duplication. *Genome Biol.* **22**, 103 (2021).
47. Wade, C. M., Mordan, P. & Naggs, F. Evolutionary relationships among the Pulmonate land snails and slugs (Pulmonata, Stylommatophora). *Biol. J. Linn. Soc.* **87**, 593–610 (2006).
48. Lallemant, T., Leduc, M., Landes, C., Rizzon, C. & Lerat, E. An overview of duplicated gene detection methods: Why the duplication mechanism has to be accounted for in their choice. *Genes* **11**, 5 (2020).
49. Niu, X. M. *et al.* Transposable elements drive rapid phenotypic variation in *Capsella rubella*. *Proc. Natl. Acad. Sci. USA* **116**, 6908–6913 (2019).
50. Schrader, L. *et al.* Transposable element islands facilitate adaptation to novel environments in an invasive species. *Nat. Commun.* **5**, 5495 (2014).
51. Schrader, L. & Schmitz, J. The impact of transposable elements in adaptive evolution. *Mol. Ecol.* **28**, 1537–1549 (2019).
52. Canapa, A., Barucca, M., Biscotti, M. A., Forconi, M. & Olmo, E. Transposons, genome size, and evolutionary insights in animals. *Cytogenet. Genome Res.* **147**, 217–239 (2015).
53. Naville, M. *et al.* Massive changes of genome size driven by expansions of non-autonomous transposable elements. *Curr. Biol.* **29**, 1161–1168 (2019).
54. Zhou, W., Liang, G., Molloy, P. L. & Jones, P. A. DNA methylation enables transposable element-driven genome expansion. *Proc. Natl. Acad. Sci. USA* **117**, 19359–19366 (2020).
55. Guo, Y. *et al.* A chromosomal-level genome assembly for the giant African snail *Achatina fulica*. *Gigascience* **8**, giz124 (2019).
56. Sun, J. *et al.* Signatures of divergence, invasiveness, and terrestrialization revealed by four apple snail genomes. *Mol. Biol. Evol.* **36**, 1507–1520 (2019).
57. Zemanova, M. A. *et al.* Slimy invasion: Climatic niche and current and future biogeography of *Arion* slug invaders. *Divers. Distrib.* **24**, 1627–1640 (2018).
58. Zemanova, M. A., Knop, E. & Heckel, G. Introgressive replacement of natives by invading *Arion* pest slugs. *Sci. Rep.* **7**, 14908 (2017).
59. Alix, K., Gerard, P. R., Schwarzacher, T. & Heslop-Harrison, J. S. P. Polyploidy and interspecific hybridization: Partners for adaptation, speciation and evolution in plants. *Ann. Bot.* **120**, 183–194 (2017).
60. Sacerdot, C., Louis, A., Bon, C., Berthelot, C. & Roest Crollius, H. Chromosome evolution at the origin of the ancestral vertebrate genome. *Genome Biol.* **19**, 166 (2018).
61. Crow, K. D. & Wagner, G. P. What is the role of genome duplication in the evolution of complexity and diversity?. *Mol. Biol. Evol.* **23**, 887–892 (2006).
62. Wang, Y. *et al.* MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**, e49 (2012).
63. Sun, P. *et al.* WGDI: A user-friendly toolkit for evolutionary analyses of whole2 genome duplications and ancestral karyotypes. <https://www.biorxiv.org/content/https://doi.org/10.1101/2021.04.29.441969v1> (2021).
64. Dreijers, E., Reise, H. & Hutchinson, J. M. C. Mating of the slugs *Arion* *isutanicus* auct. non *Mabille* and *A. rufus* (L.): Different genitalia and mating behaviours are incomplete barriers to interspecific sperm exchange. *J. Molluscan Stud.* **79**, 51–63 (2013).
65. Hatteland, B. A. *et al.* Introgression and differentiation of the invasive slug *Arion vulgaris* from native *A. ater*. *Malacologia* **58**, 303–321 (2015).
66. Roth, S., Hatteland, B. A. & Solhoy, T. Some notes on reproductive biology and mating behaviour of *Arion vulgaris* Moquin-Tandon 1855 in Norway including a mating experiment with a hybrid of *Arion rufus* (Linnaeus 1758) x *ater* (Linnaeus 1758). *J. Conchol.* **41**, 249–258 (2012).
67. Barr, N. B. *et al.* Application of a DNA barcode using the 16S rRNA gene to diagnose pest *Arion* species in the USA. *J. Molluscan Stud.* **75**, 187–191 (2009).

68. Quinteiro, J., Rodríguez-Castro, J., Castillejo, J., Iglesias-Piñero, J. & Rey-Méndez, M. Phylogeny of slug species of the genus *Arion*: Evidence of monophyly of Iberian endemics and of the existence of relict species in Pyrenean refuges. *J. Zool. Syst.* **43**, 139–148 (2005).
69. Arseneau, J. R., Steeves, R. & Laflamme, M. Modified low-salt CTAB extraction of high-quality DNA from contaminant-rich tissues. *Mol. Ecol. Resour.* **17**, 686–693 (2017).
70. Mostovoy, Y. *et al.* A hybrid approach for de novo human genome sequence assembly and phasing. *Nat. Methods* **13**, 587–590 (2016).
71. Burton, J. N. *et al.* Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat. Biotechnol.* **31**, 1119–1125 (2013).
72. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
73. Wick, R. R., Judd, L. M. & Holt, K. E. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol.* **20**, 129 (2019).
74. Vurture, G. W. *et al.* GenomeScope: Fast reference-free genome profiling from short reads. *Bioinformatics* **33**, 2202–2204 (2017).
75. Ruan, J. & Li, H. Fast and accurate long-read assembly with wtdbg2. *Nat. Methods* **17**, 155–158 (2020).
76. Warren, R. L. *et al.* ntEdit: scalable genome sequence polishing. *Bioinformatics* **35**, 4430–4432 (2019).
77. Ning, Z. H. In *Scaff10X v4.2: Pipeline for Scaffolding and Breaking a Genome Assembly Using 10x Genomics Linked-Reads*. <https://github.com/wtsi-hpag/Scaff10X> (2020).
78. Matthey-Doret, C. *et al.* hicstuff: Simple library/pipeline to generate and handle Hi-C data. Zenodo. 10.5281/zenodo.4066363 (2020)
79. Baudry, L. *et al.* instaGRAAL: Chromosome-level quality scaffolding of genomes using a proximity ligation-based scaffold. *Genome Biol.* **21**, 148 (2020).
80. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
81. Brůna, T., Hoff, K. J., Lomsadze, A., Stanke, M. & Borodovsky, M. BRAKER2: Automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom. Bioinform.* **3**, 108 (2021).
82. Hauser, M., Steinegger, M. & Soding, J. MMseqs software suite for fast and deep clustering and searching of large protein sequence sets. *Bioinformatics* **32**, 1323–1330 (2016).
83. Keilwagen, J., Hartung, F. & Grau, J. GeMoMa: Homology-Based gene prediction utilizing intron position conservation and RNA-seq data. *Methods Mol. Biol.* **1962**, 161–177 (2019).
84. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
85. Haas, B. J. *et al.* Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* **31**, 5654–5666 (2003).
86. Haas, B. J. *et al.* Automated eukaryotic gene structure annotation using EVIDENCEModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **9**, R7 (2008).
87. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
88. Wheeler, T. J. & Eddy, S. R. nhmmer: DNA homology search with profile HMMs. *Bioinformatics* **29**, 2487–2489 (2013).
89. Conesa, A. *et al.* Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674–3676 (2005).
90. Camacho, C. *et al.* BLAST+: Architecture and applications. *BMC Bioinform.* **10**, 421 (2009).
91. Emms, D. M. & Kelly, S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
92. De Bie, T., Cristianini, N., Demuth, J. P. & Hahn, M. W. CAFE: A computational tool for the study of gene family evolution. *Bioinformatics* **22**, 1269–1271 (2006).
93. Nakamura, T., Yamada, K. D., Tomii, K. & Katoh, K. Parallelization of MAFFT for large-scale multiple sequence alignments. *Bioinformatics* **34**, 2490–2492 (2018).
94. Suyama, M., Torrents, D. & Bork, P. PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34**, W609–612 (2006).
95. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540–552 (2000).
96. Yang, Z. PAML: A program package for phylogenetic analysis by maximum likelihood. *Bioinformatics* **13**, 555–556 (1997).
97. Shannon, P. *et al.* Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
98. Edgar, R. C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
99. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
100. Minh, B. Q. *et al.* IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
101. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermini, L. S. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
102. Zhang, C., Scornavacca, C., Molloy, E. K. & Mirarab, S. ASTRAL-Pro: Quartet-based species-tree inference despite paralogy. *Mol. Biol. Evol.* **37**, 3292–3307 (2020).
103. Bouckaert, R. R. & Heled, J. DensiTree 2: Seeing trees through the forest. <https://www.biorxiv.org/content/https://doi.org/10.1101/012401v1> (2014).
104. Frýda, J., Nützel, A. & Wagner, P. J. In *Phylogeny and Evolution of the Mollusca*. (eds. Ponder W. & Lindberg D. R.) 238–270 (University of California Press, Berkeley, 2008).
105. Nützel, A., Erwin, D. H. & Mapes, R. H. Identify and phylogeny of the Late Paleozoic Sublitoidea (Gastropoda). *J. Paleont.* **74**, 575–598 (2000).
106. Pojeta, J. & Runnegar, B. *Fordilla troyensis* and the early history of pelecypod mollusks: Early cambrian fossils from New York State provide important clues to the evolution of the class. *Am. Sci.* **62**, 706–711 (1974).
107. Harris, R. S. In *Improved pairwise alignment of genomic DNA*, Pennsylvania State University (2007).
108. Di Palma, F. *et al.* In *The Draft Genome of Aplysia californica*. <https://www.ncbi.nlm.nih.gov/nuccore/AASC00000000.3> (2014).
109. Krzywinski, M. *et al.* Circos: An information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).
110. Qiao, X. *et al.* Gene duplication and evolution in recurring polyploidization-diploidization cycles in plants. *Genome Biol.* **20**, 38 (2019).
111. Wang, D., Zhang, Y., Zhang, Z., Zhu, J. & Yu, J. KaKs\_Calculator 2.0: A toolkit incorporating gamma-series methods and sliding window strategies. *Genom. Proteom. Bioinform.* **8**, 77–80 (2010).
112. Benson, G. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).
113. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinform.* **25**, 41011–141014 (2009).

114. Bao, W., Kojima, K. K. & Kohany, O. Repbase update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* **6**, 11 (2015).
115. Wheeler, T. J. *et al.* Dfam: a database of repetitive DNA based on profile hidden Markov models. *Nucleic Acids Res.* **41**, D70–82 (2013).
116. Li, H. In *Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM*. <https://www.arxiv-vanity.com/papers/1303.3997> (2013).
117. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
118. *Picard toolkit*. In *Broad Institute, GitHub Repository*. <http://broadinstitute.github.io/picard> (2019).
119. do-Valle, I. F. *et al.* Optimized pipeline of MuTect and GATK tools to improve the detection of somatic single nucleotide polymorphisms in whole-exome sequencing data. *BMC Bioinform.* **17**, 341 (2016).
120. Nadachowska-Brzyska, K., Burri, R., Smeds, L. & Ellegren, H. PSMC analysis of effective population sizes in molecular ecology and its application to black-and-white *Ficedula* flycatchers. *Mol. Ecol.* **25**, 1058–1072 (2016).
121. Davies, S. M. *Arion flagellus* Collinge and *A. lusitanicus* Mabile in the British Isles: A morphological, biological and taxonomic investigation. *J. Conchol.* **32**, 339–354 (1987).
122. Pimentel, D. Life history of *Australorbis glabratus*, the intermediate snail host of *Schistosoma mansoni* in Puerto rico. *Ecology* **38**, 576–580 (1957).
123. Pawson, P. A. & Chase, R. The life-cycle and reproductive of *Achatina fulica* (bowdich) in laboratory culture. *J. Molluscan Stud.* **50**, 85–91 (1984).

## Acknowledgements

We thank Gert Wörheide and Michael Eitel (Department of Earth and Environmental Sciences, Ludwig-Maximilians-Universität München, LMU) for organizing the IGNITE Comparative Genomics of Non-Model Invertebrates program; Denis Tagu (INRAE), Fabrice Legeai (INRAE, INRIA), Claire Lemaitre (INRIA), Grace McCormack and Kenneth Sandoval (National University of Ireland) for the guidance and advice on genome assembly during secondments; Ramon Rivera (LMU) for software installation and maintenance; Juan Moles (ZSM) for polishing an earlier version of the article; Isabella Stöger (ZSM) for laboratory work and other ZSM mollusk teammates as well as colleagues from IGNITE for valuable feedback and comments. The bioinformatics analyses were performed at Leibniz-Rechenzentrum der Bayerischen Akademie der Wissenschaften, TUBITAK ULAKBIM, High Performance and Grid Computing Center (TRUBA resources) and GenOuest IRISA-INRIA, France. This study has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 764840.

## Author contributions

Z.C. and M.S. conceived the study. Z.C., Ö.D., N.G., A.G. performed genome assembly. Z.C. performed genome annotation and evolution analysis. Z.C. and M.S. wrote the manuscript.

## Funding

Open Access funding enabled and organized by Projekt DEAL.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-18099-7>.

**Correspondence** and requests for materials should be addressed to Z.C.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022



Supplemental Tables and Figures for Publication III

**Pulmonate slug evolution is reflected in the de novo genome of  
*Arion vulgaris* Moquin-Tandon, 1855.**

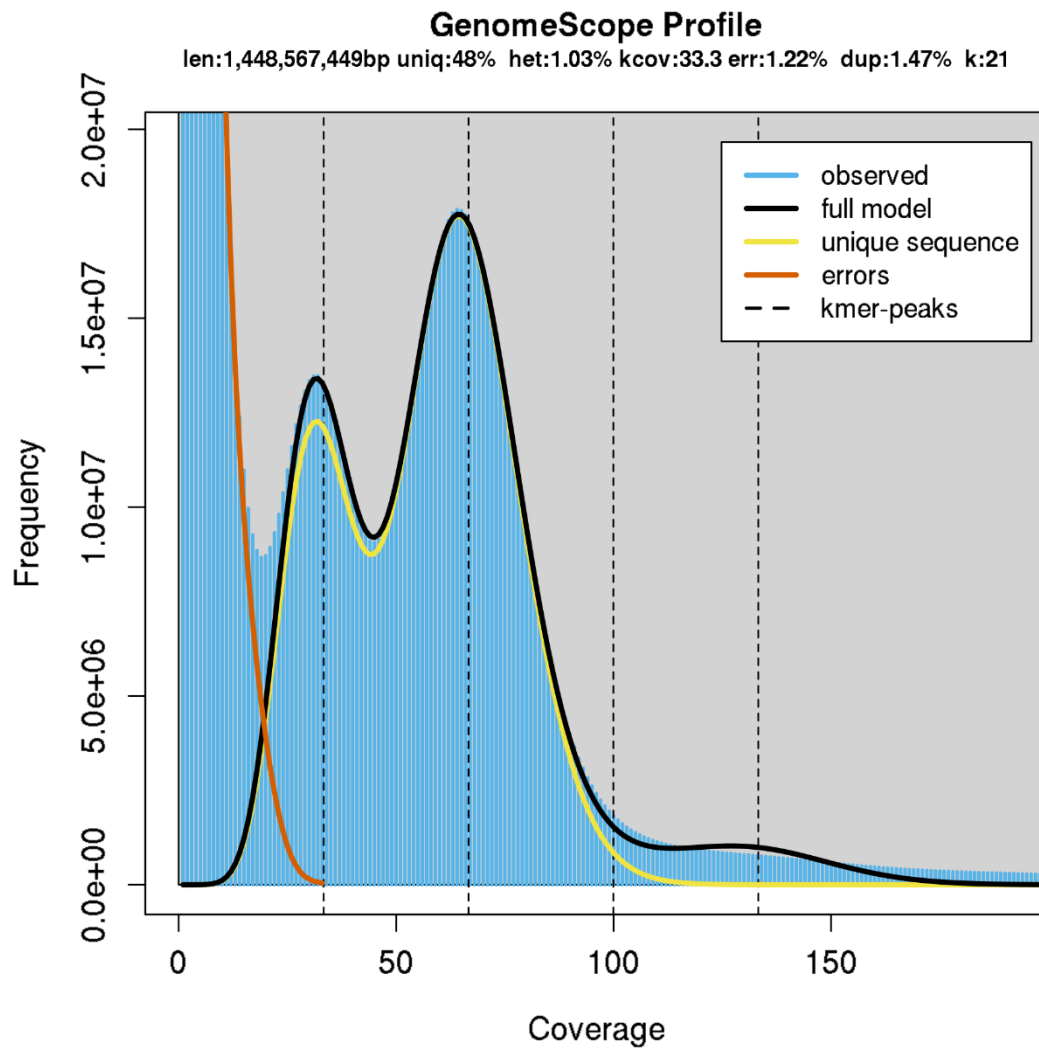


Fig S1. Estimation of genome size of *Arion vulgaris* based on the distribution of 21-mer frequency in the combination of short reads and linked reads. Two peaks indicate high heterozygosity.

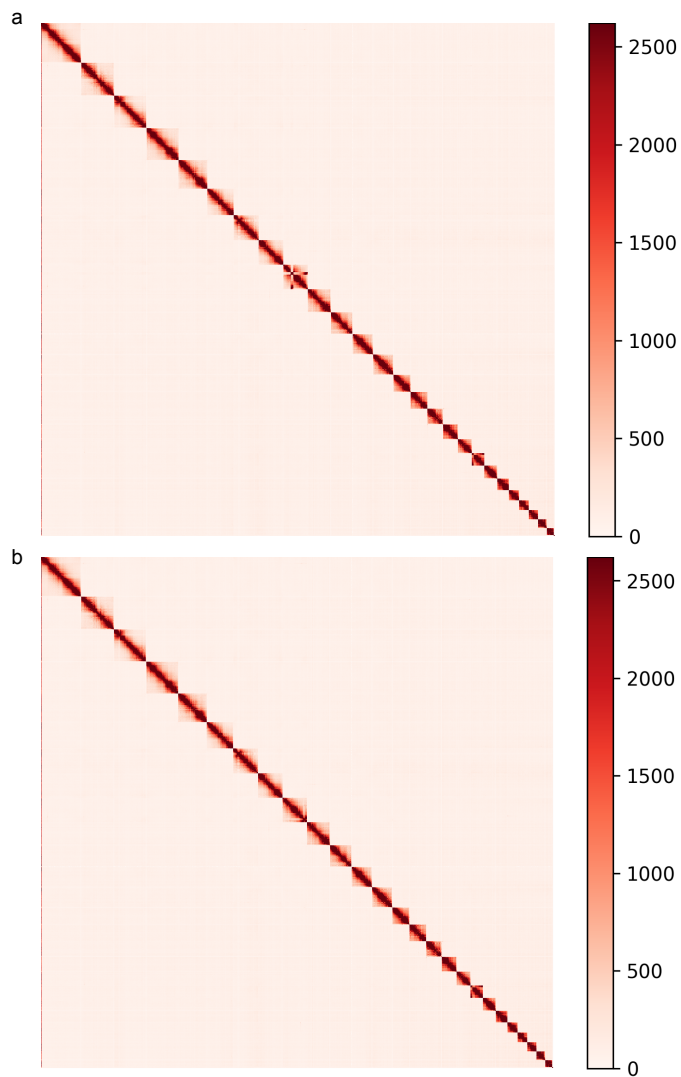


Fig S2. Hi-C chromosome contact maps. Each block represents a Hi-C contact between two genomic loci within a 1 Mb window. The darker the color of a block, the higher the contact intensity. a) An initial Hi-C contact map shows incongruous signals in chromosome 9, which might reflect a misassembly. b) Hi-C contact map after manual correction of the chromosome 9 error in the assembly.

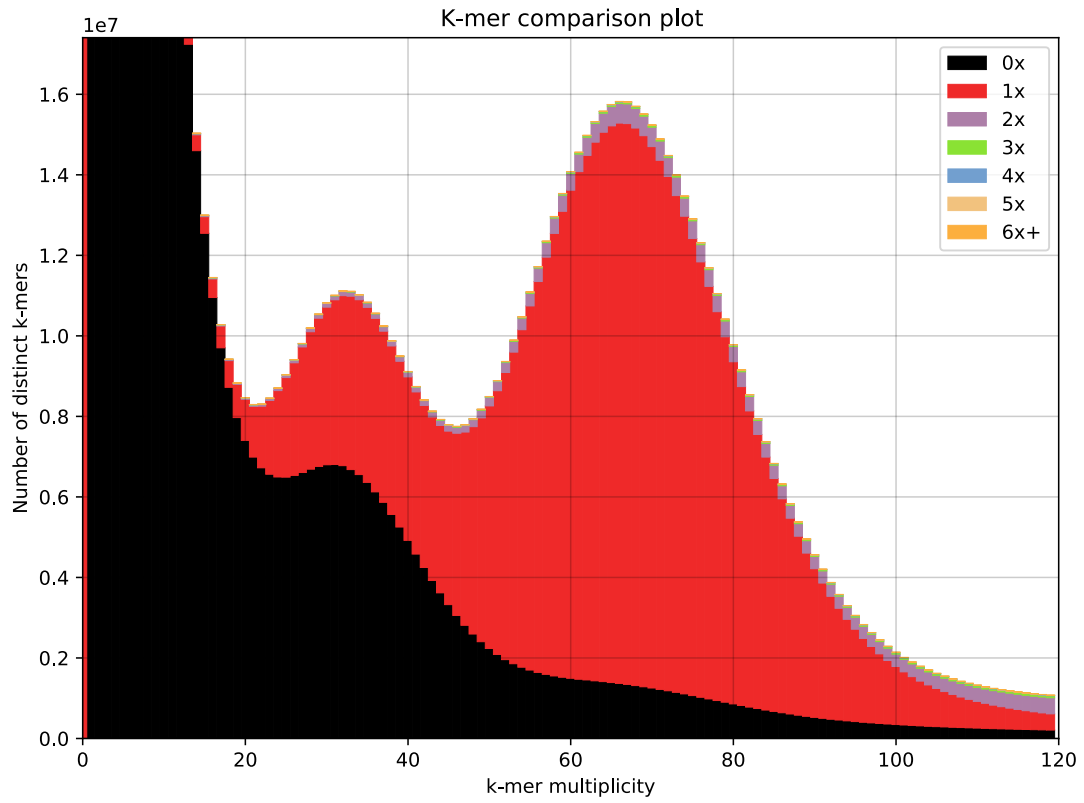


Fig S3. Read *k-mer* frequency versus assembly copy number stacked histograms for final *Arion vulgaris* assembly. Read content in black is absent from the assembly, red occurs once, purple twice, etc.

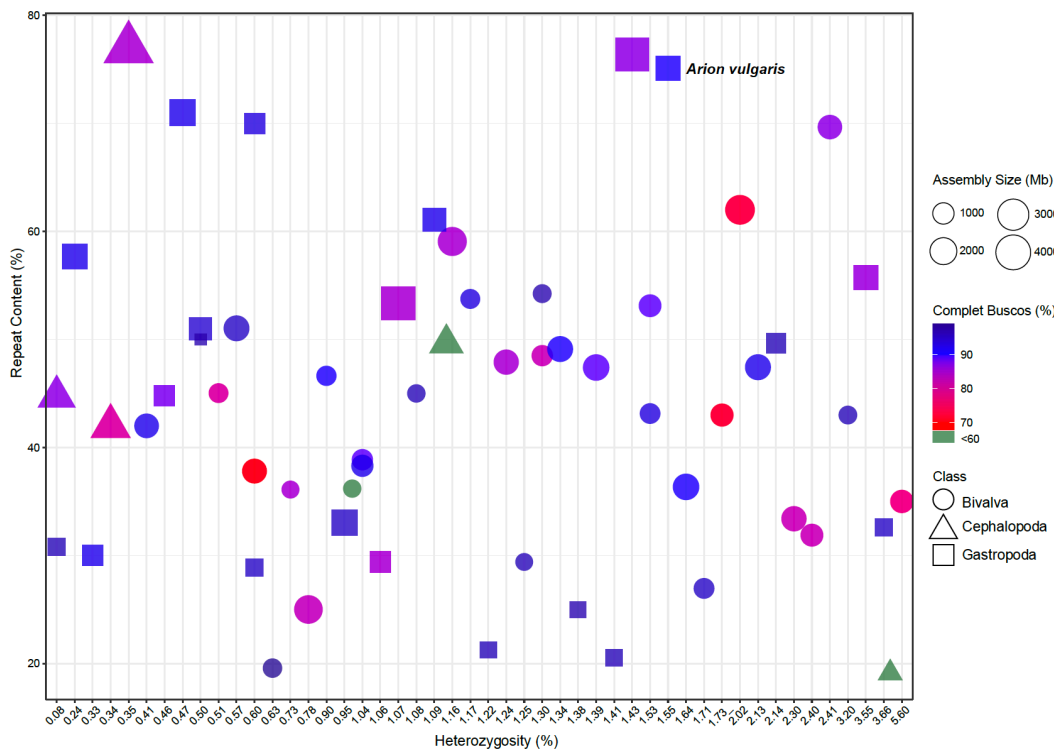


Fig S4. Genome features and completeness of Mollusca genomes. The figure includes all published mollusk genomes (statistics date until December 2021) with reported heterozygosity, repeat contents, assembled genome size and BUSCOs assessments. Each icon represents one genome, and the area of the icon represents the size of the assembled genome. Color shows the percentage of complete BUSCOs. Different classes are represented by different shapes. The *Arion vulgaris* genome is annotated in the Figure.

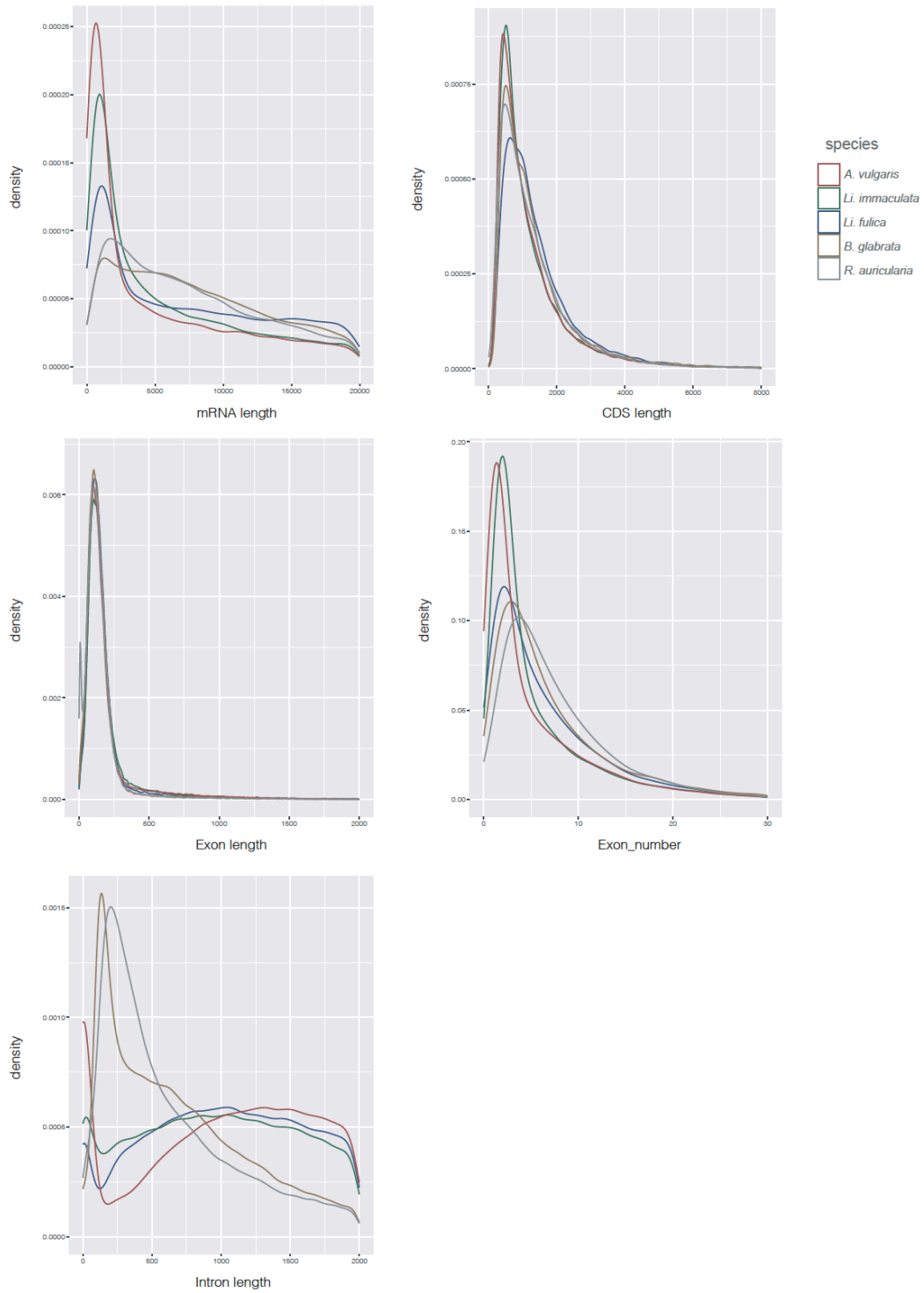


Fig S5. Comparison of mRNA length (a), CDS length (b), exon length (c), exon number per gene (d), and intron length (e), between *Arion vulgaris*, *Lissachatina (Achatina) fulica*, *Li. immaculata*, *Biomphalaria glabrata*, and *Radix auricularia*.

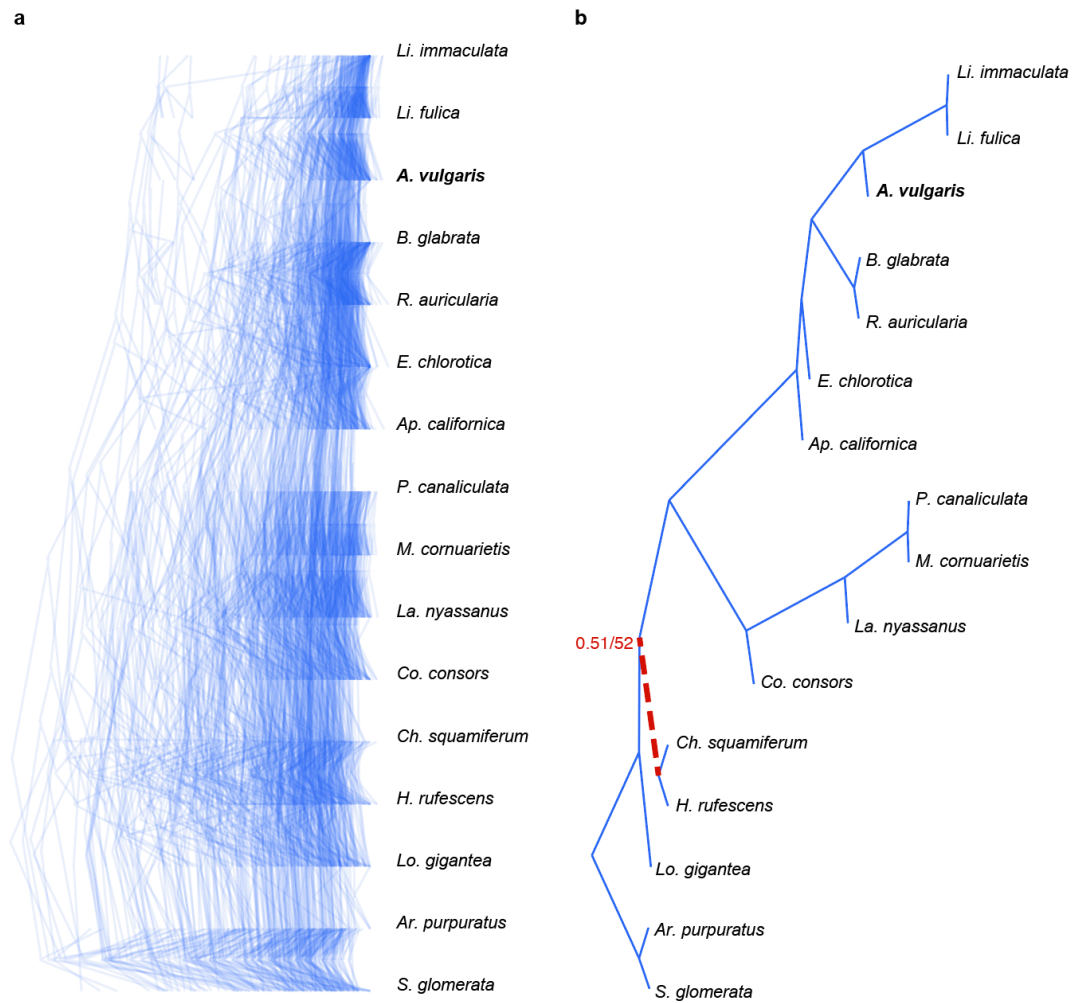


Fig S6. Gene trees estimated using 223 orthologous genes. (a) Superimposed ultrametricity gene trees in a consensus DensiTree plot. (b) Final ASTRAL tree inferred by gene trees. The sole branch without 100% bootstrap support is highlighted by a red dashed line and posterior probabilities and bootstrap support are indicated respectively. *Arion vulgaris* was highlighted in bold.

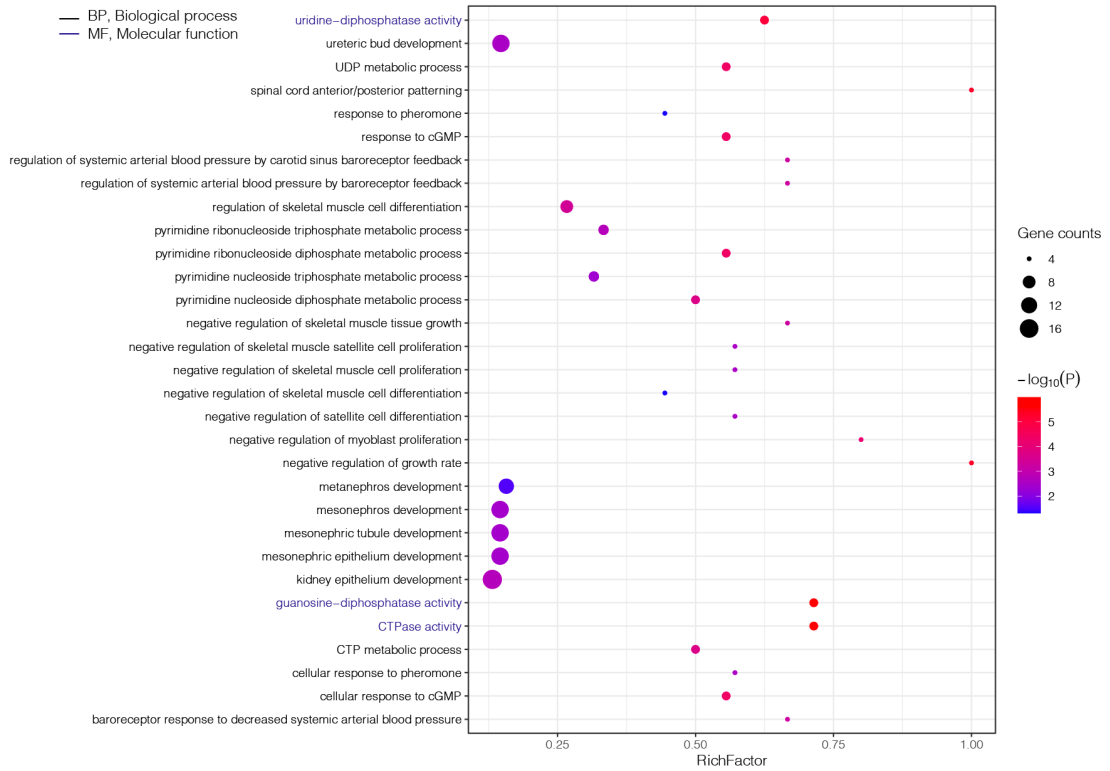


Fig S7. Gene Ontology (GO) enrichment analysis of Stylommatophora specific genes. Only significantly enriched terms with corrected  $P < 0.05$  were indicated. The color and size of each point represented the  $-\log_{10}$  (FDR) values and Gene counts. A higher  $-\log_{10}$  (FDR) value and enrichment score indicated a greater degree of enrichment. Enrichment factor refers to the ratio of the number of Stylommatophora-specific genes in the pathway and the number of all annotated genes in the pathway.



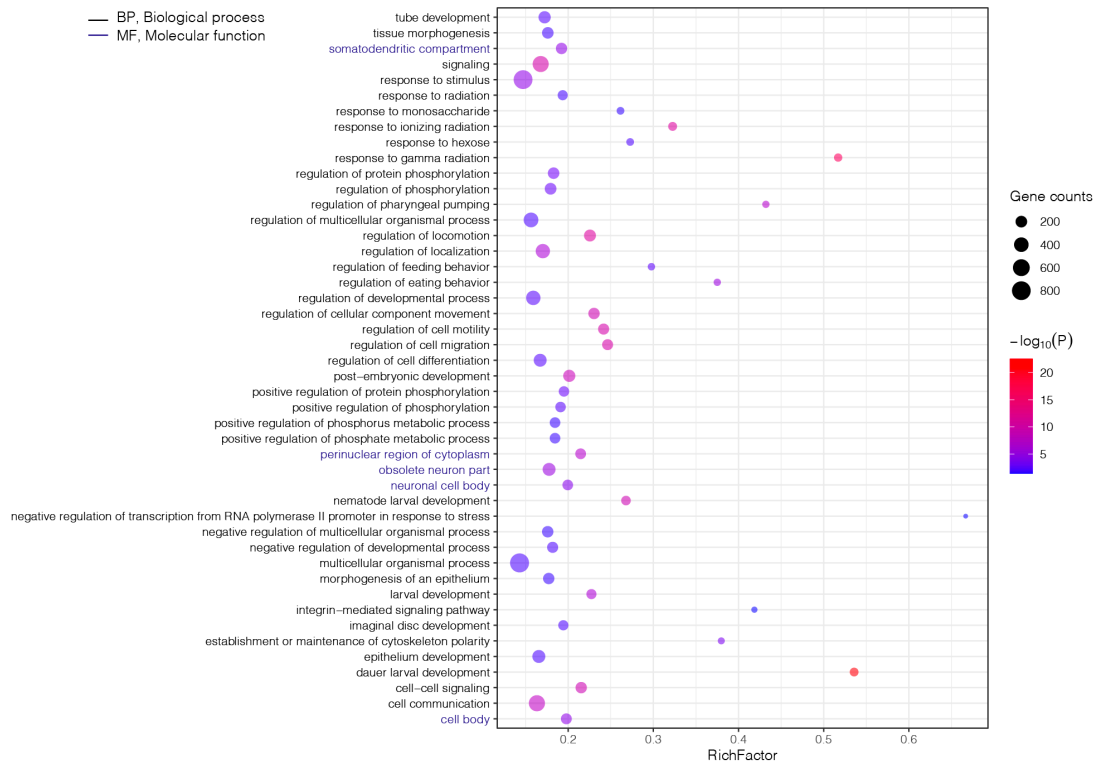


Fig S8. Gene Ontology (GO) enrichment analysis of Stylommatophora expansion genes. Only significantly enriched terms with corrected  $P < 0.05$  were indicated. The color and size of each point represented the  $-\log_{10}$  (FDR) values and gene counts. A higher  $-\log_{10}$  (FDR) value and enrichment score indicated a greater degree of enrichment. Rich factor refers to the ratio of the number of Stylommatophora-expanded genes in the pathway and the number of all annotated genes in the pathway.

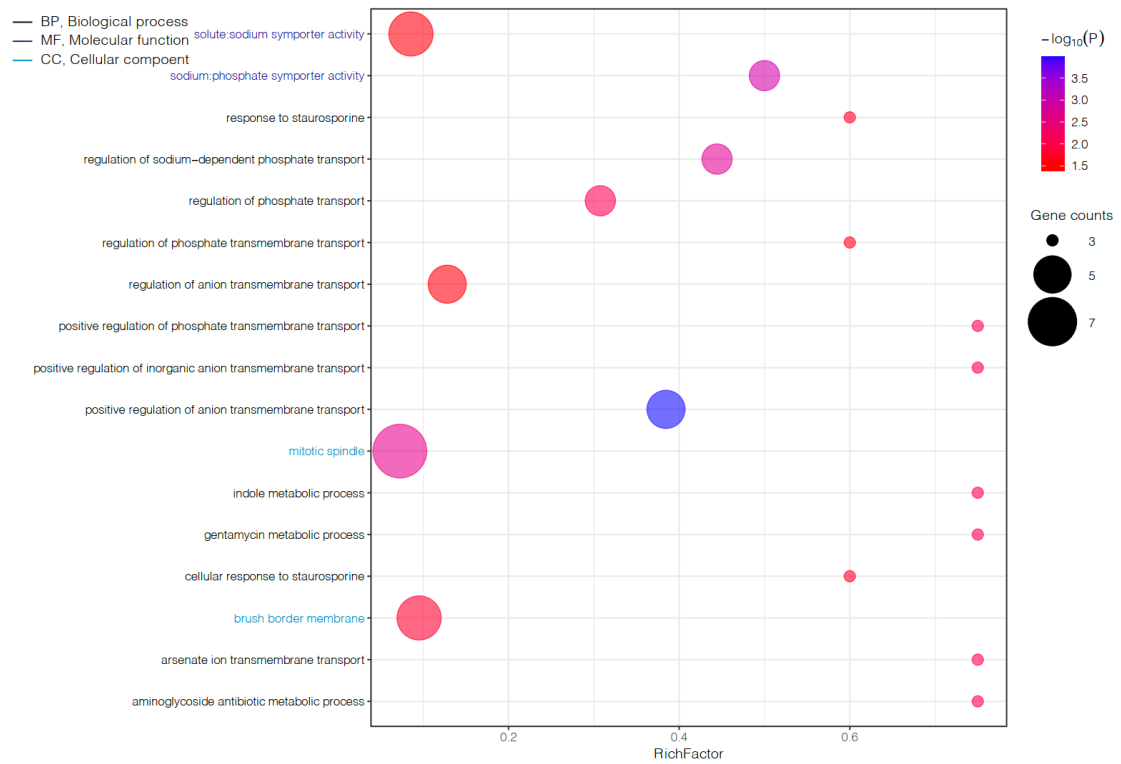


Fig S9. Gene Ontology (GO) enrichment analysis of Stylommatophora contraction genes. Only significantly enriched terms with corrected  $P < 0.05$  were indicated. The color and size of each point represented the  $-\log_{10}$  (FDR) values and Gene counts. A higher  $-\log_{10}$  (FDR) value and enrichment score indicated a greater degree of enrichment. Rich factor refers to the ratio of the number of Stylommatophora-contracted genes in the pathway and the number of all annotated genes in the pathway.

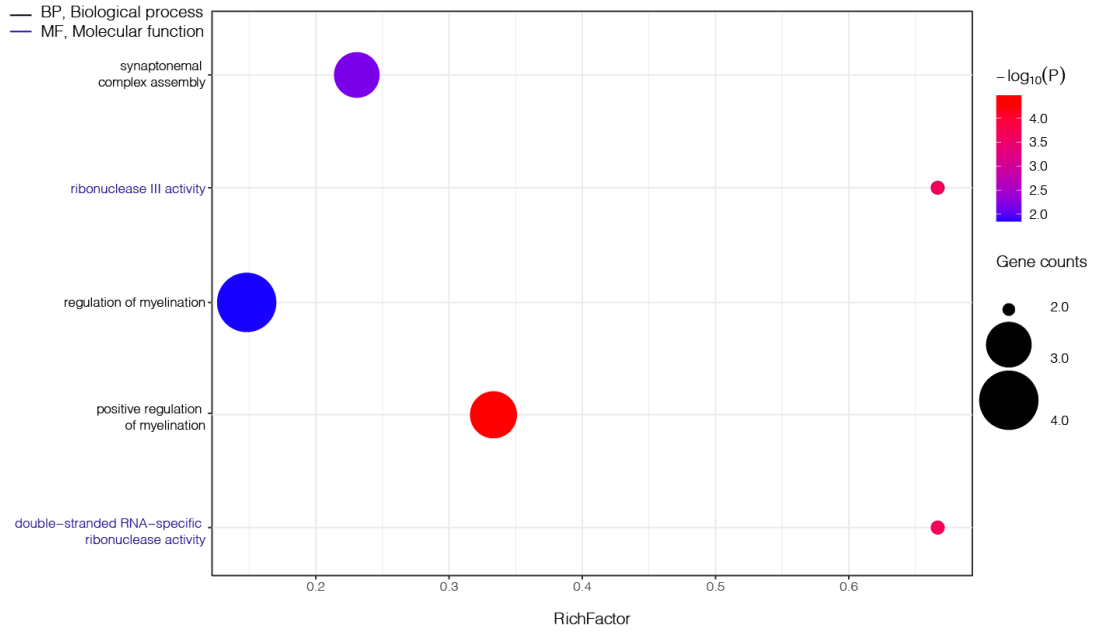


Fig S10. Gene ontology (GO) Enrichment analysis of putatively positively selected genes in *Styломmatophora*. Only significantly enriched terms with corrected  $P < 0.05$  were indicated. The color and size of each point represented the  $-\log_{10}(\text{FDR})$  values and Gene counts. A higher  $-\log_{10}(\text{FDR})$  value and enrichment score indicated a greater degree of enrichment. Rich factor refers to the ratio of the number of *Styломmatophora*-positive selected genes in the pathway and the number of all genes annotated in the pathway.

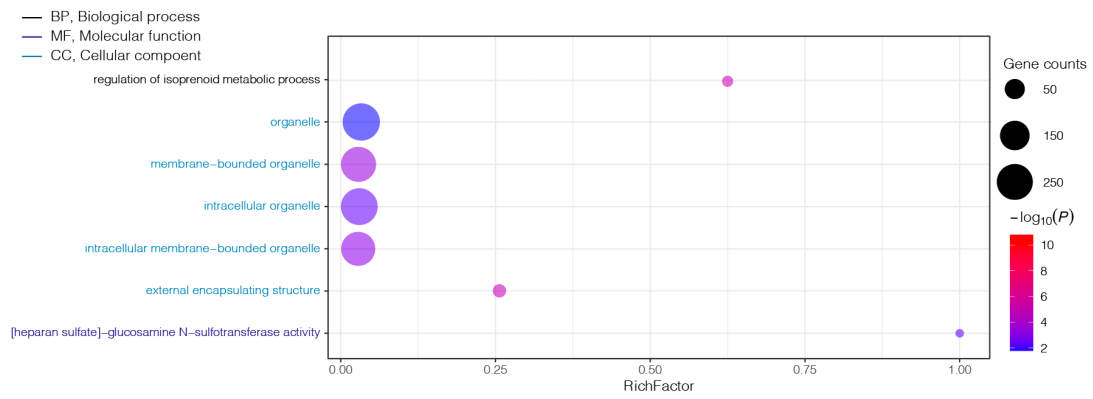


Fig S11. Gene ontology (GO) Enrichment analysis of *Arion vulgaris* species specific genes. Only significantly enriched terms with corrected  $P < 0.05$  were indicated. The color and size of each point represented the  $-\log_{10}$  (FDR) values and Gene counts. A higher  $-\log_{10}$  (FDR) value and enrichment score indicated a greater degree of enrichment. Rich factor refers to the ratio of the number of *A. vulgaris* specific genes in the pathway and the number of all annotated genes in the pathway.



Fig S12. Gene ontology (GO) Enrichment analysis of *Arion vulgaris* expansion genes. Only significantly enriched terms with corrected  $P < 0.05$  were indicated. The color and size of each point represented the  $-\log_{10}$  (HDR) values and Gene counts. A higher  $-\log_{10}$  (HDR) value and enrichment score indicated a greater degree of enrichment. Rich factor refers to the ratio of the number of *A. vulgaris* expansion genes in the pathway and the number of all annotated genes in the pathway.

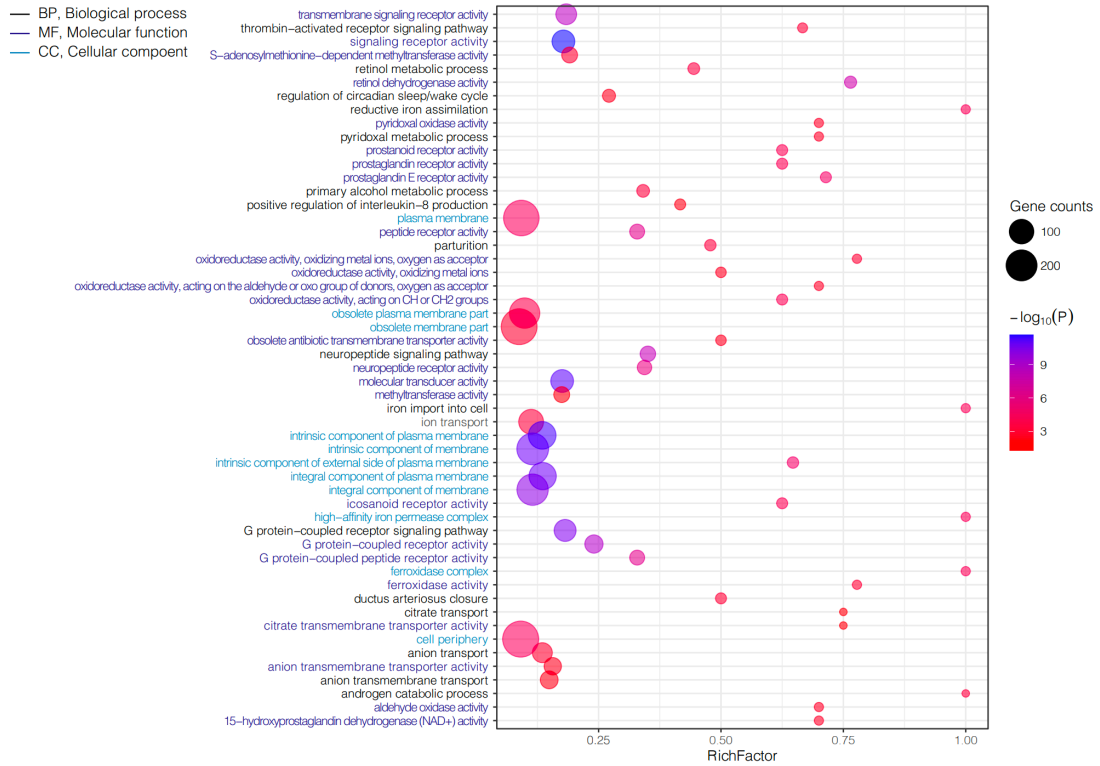


Fig S13. Gene ontology (GO) Enrichment analysis of *Arion vulgaris* contraction genes. Only significantly enriched terms with corrected  $P < 0.05$  were indicated. The color and size of each point represented the  $-\log_{10}$  (FDR) values and Gene counts. A higher  $-\log_{10}$  (FDR) value and enrichment score indicated a greater degree of enrichment. Rich factor refers to the ratio of the number of *A. vulgaris* contraction genes in the pathway and the number of all annotated genes in the pathway.



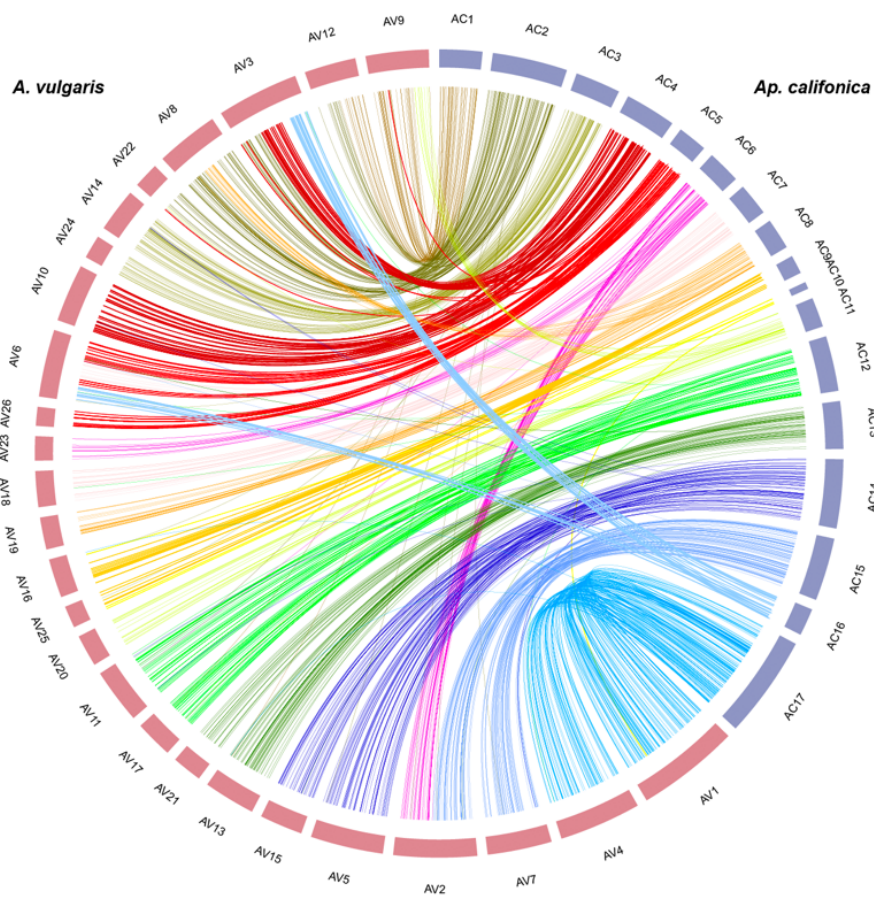


Fig S15. A one to two corresponding relationship in the comparison of *Arion vulgaris* and *Aplysia californica* chromosomes.



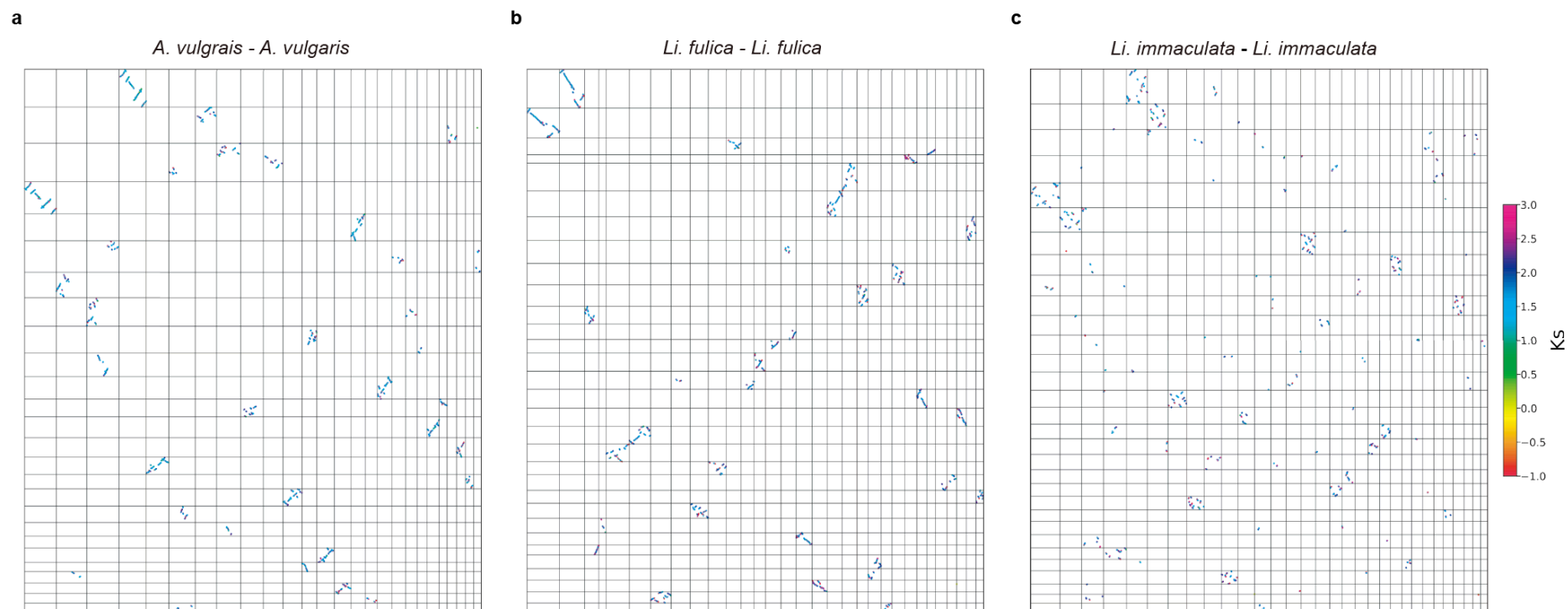


Fig S16. *Ks* dot plot in a) *Arion vulgrais* genome, b) *Lissachatina (Achatina) fulica* genome, and c) *Li. immaculata* genome. The corresponding median *Ks* value is shown for each syntenic block. Each block from horizontally left to right, and vertically from top to bottom in turn represents a) *A. vulgrais* chromosomes from 1-26, b) *Li. fulica* chromosomes from 1-31, and c) *Li. immaculata* chromosomes from 1-31.

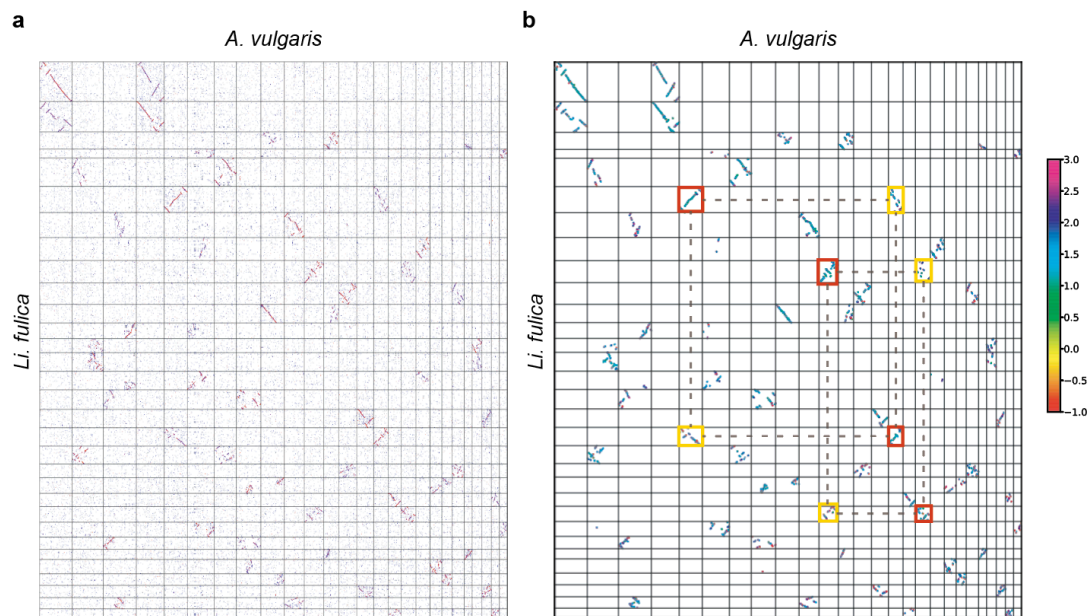


Fig S17. (a) Homologous gene dot plot in comparison of *Arion vulgaris* and *Lissachatina (Achatina) fulica*. Horizontally, from left to right, blocks, in turn represent the chromosomes 1-31 of *Li. fulica*. Vertically, from top to bottom, blocks, in turn represent chromosomes 1-26 of *A. vulgaris*. The color of red, blue, and gray dots represents the BLAST hit scores from high to low. (b)  $K_s$  dot plot of syntenic blocks in comparison of *A. vulgaris* and *Li. fulica*. Horizontally, from left to right, blocks, in turn represent the chromosomes 1-26 of *A. vulgaris*. Vertically, from top to bottom, blocks, in turn represent chromosomes 1-31 of *Li. fulica*. The chromosomes in the red box are more highly conserved than those in the yellow boxes.

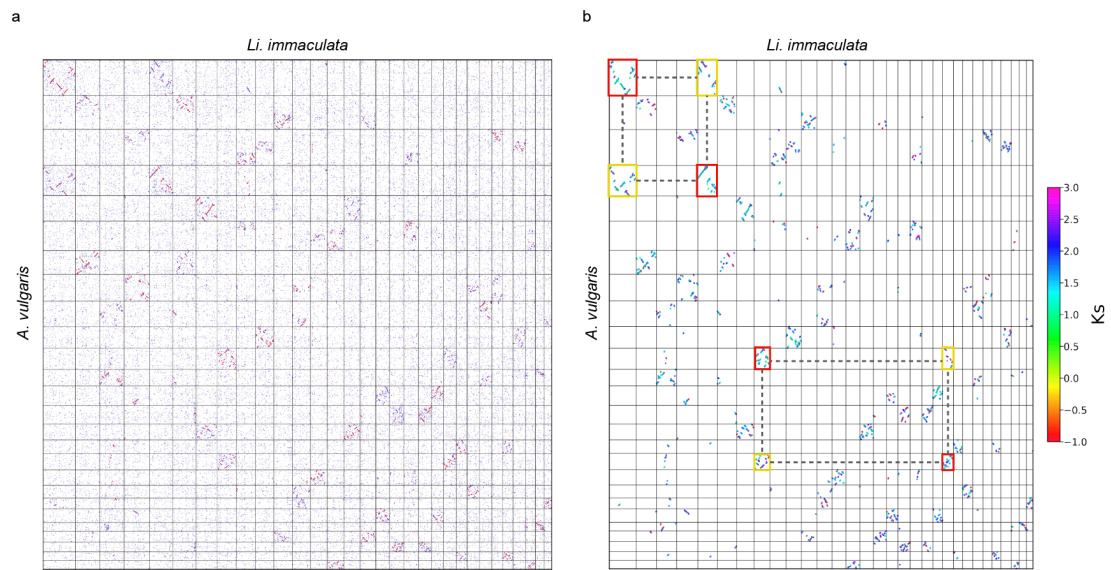


Fig S18. (a) Homologous gene dot plot in comparison of *Arion vulgaris* and *Lissachatina (Achatina) immaculata*. The color of red, blue, and gray dots represents the BLAST hit scores from high to low. (b)  $K_s$  dot plot in comparison of *A. vulgaris* and *Li. immaculata*. Horizontally, from left to right, blocks in turn represent the chromosomes 1-31 of *Li. immaculata*. Vertically, from top to bottom, blocks in turn represent chromosomes 1-26 of *A. vulgaris*. The chromosomes in the red box are more conservative than in the yellow box.

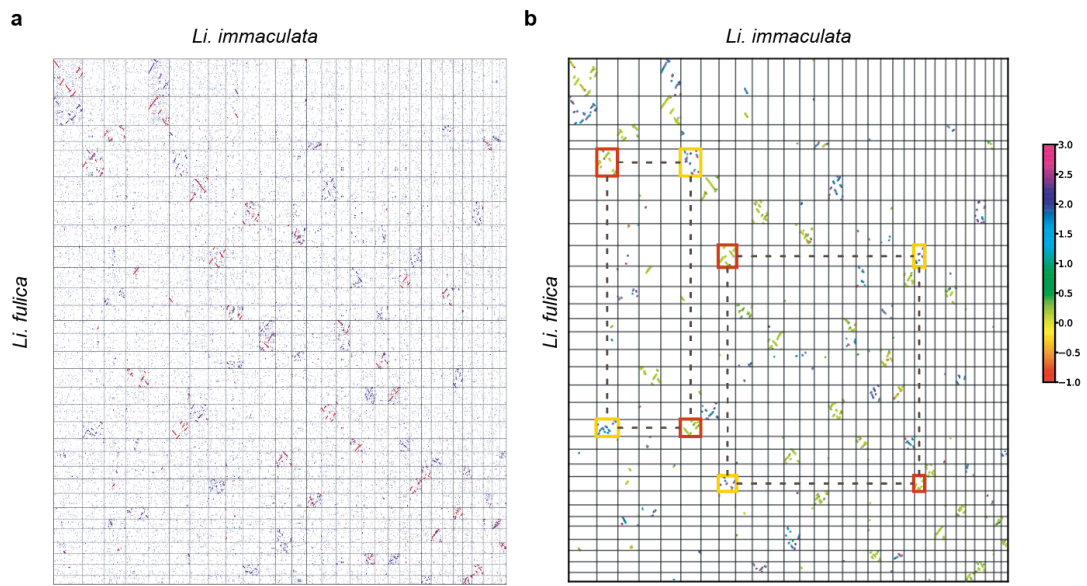


Fig S19. (a) homologous gene dot plot in comparison of *Lissachatina (Achatina) fulica* and *Li. immaculata*. The color of red, blue, and gray dots represents the BLAST hit scores from high to low. (b) *Ks* dot plot of syntenic blocks in comparison of *Li. fulica* and *Li. immaculata*. Horizontally, from left to right, blocks, in turn represent the chromosomes 1-31 of *Li. immaculata*. Vertically, from top to bottom, blocks, in turn represent chromosomes 1-31 of *Li. fulica*. The chromosomes in the red box are more conservative than in the yellow box.

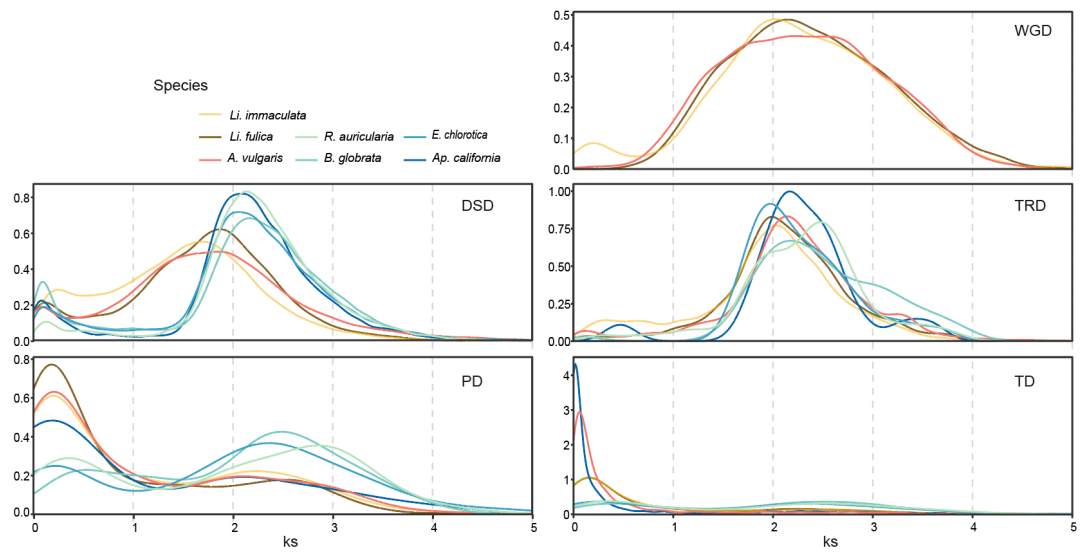


Fig S20.  $K_s$  distributions of gene pairs duplicated by different modes in seven Heterobranchia species. WGD: whole-genome duplication, DSD: dispersed duplication, TRD: transposed duplication, PD: proximal duplication, TD: tandem duplication.

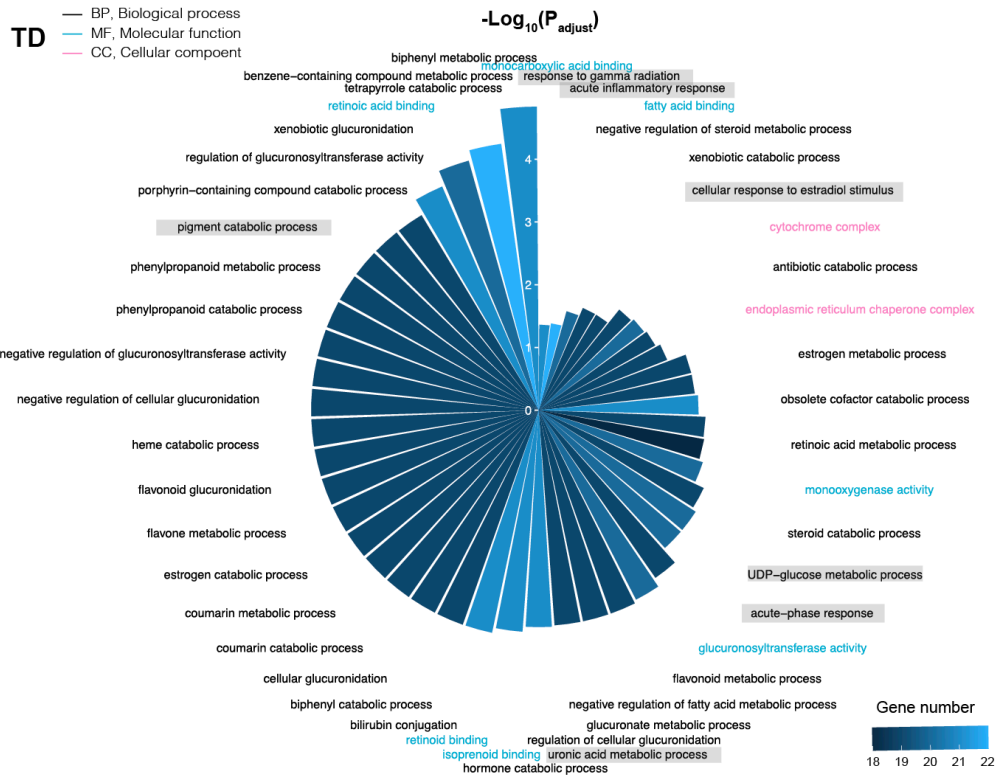


Fig S21. Functional enrichment analysis of genes derived by tandem duplication (TD). The enriched GO terms with corrected P value  $<0.005$  are presented. The color represents the number of genes in a GO term. The radius represents the statistical significance of enriched GO terms. 'P adjust' is the Benjamini–Hochberg false discovery rate (FDR) adjusted P value. A higher  $-\log_{10}$  (FDR) value and enrichment score indicated a greater degree of enrichment. Functions might be related to adaptation are marked with a gray background.



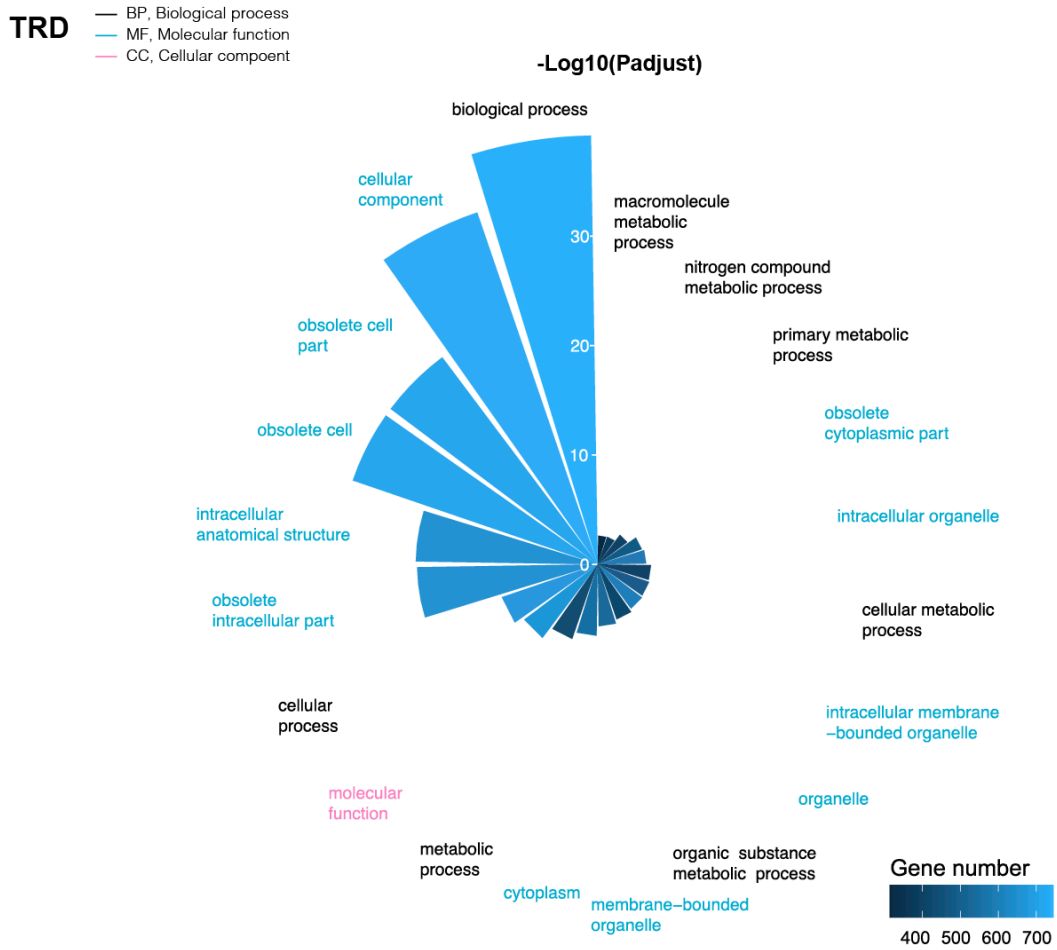


Fig S23. Functional enrichment analysis of genes derived by transposed duplication (TRD). The enriched GO terms with corrected P value <0.005 are presented. The color represents the number of genes in a GO term. The radius represents the statistical significance of enriched GO terms. ‘P adjust’ is the Benjamini–Hochberg false discovery rate (FDR) adjusted P value. A higher  $-\log_{10}$  (FDR) value and enrichment score indicated a greater degree of enrichment. Functions might be related to adaptation are marked with a gray background.



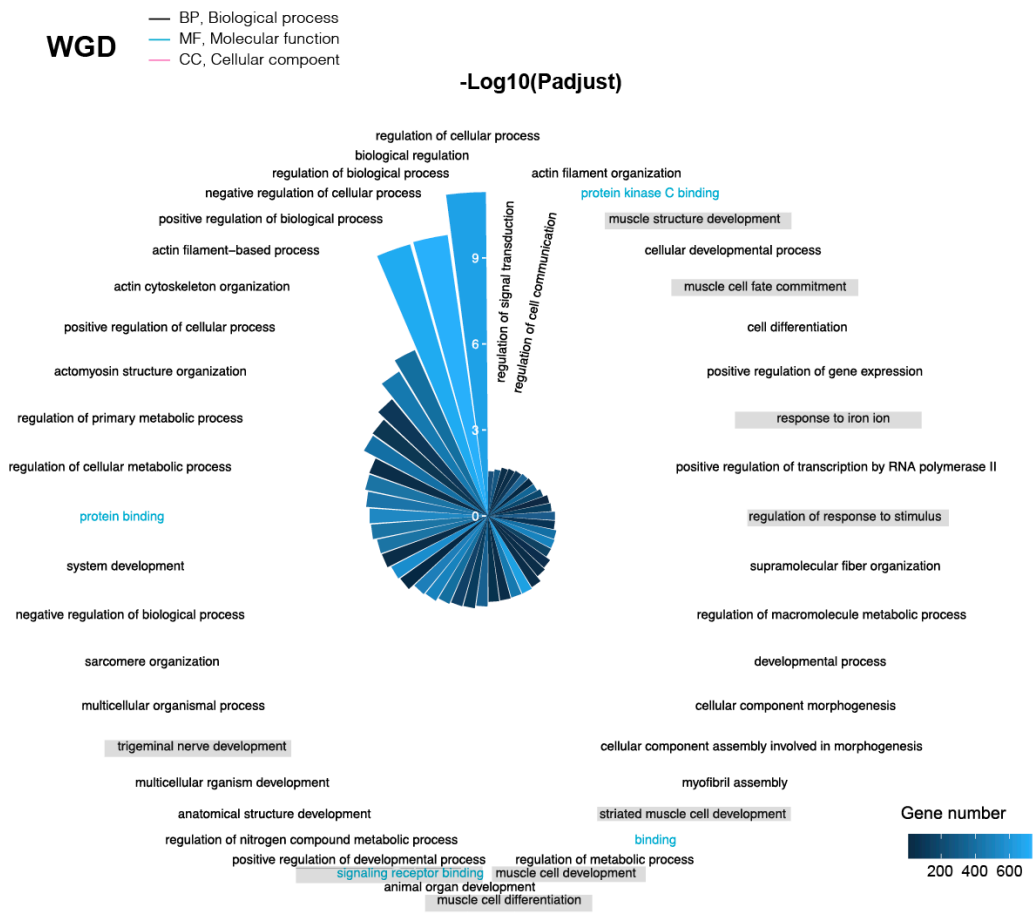


Fig S24. Functional enrichment analysis of genes derived by whole genome duplication (WGD). The enriched GO terms with corrected P value <0.005 are presented. The color represents the number of genes in a GO term. The radius represents the statistical significance of enriched GO terms. ‘P adjust’ is the Benjamini–Hochberg false discovery rate (FDR) adjusted P value. A higher  $-\log_{10}$  (FDR) value and enrichment score indicated a greater degree of enrichment. Functions might be related to adaptation are marked with a gray background.

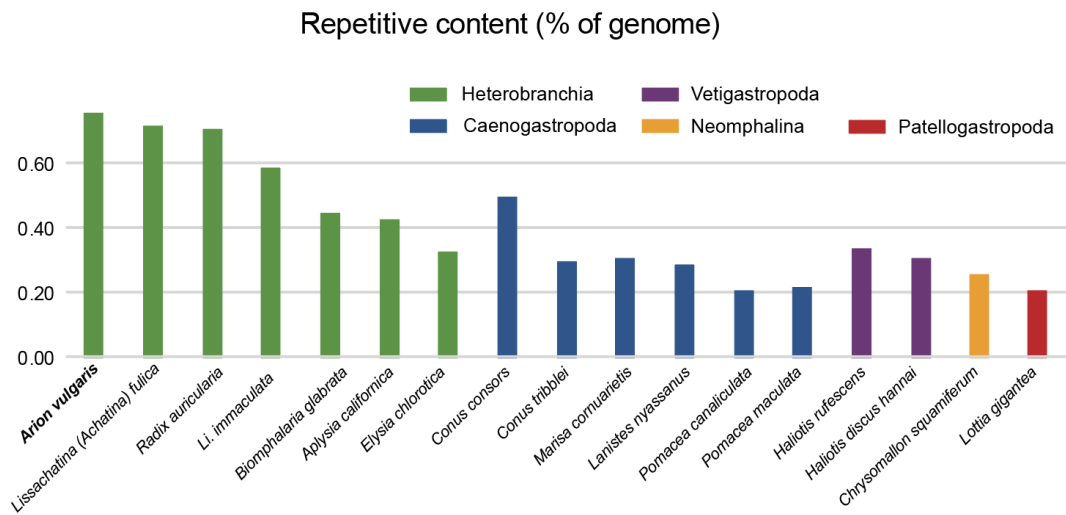


Fig S25. Repetitive content in published gastropod genomes.

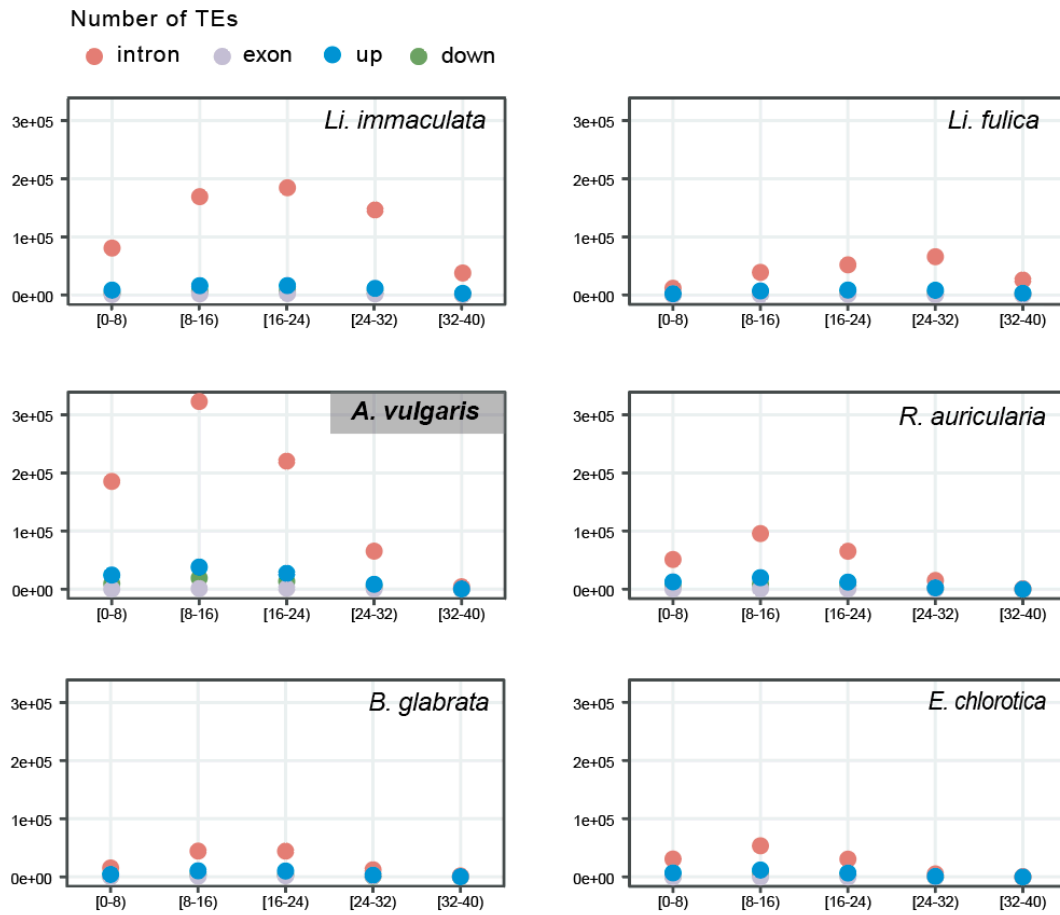


Fig S26. Dot plot shows the number of TEs in different genic regions under different ages. Percentage of divergence from consensus is used as a proxy for age: the older the invasion of the TEs, the more copies will have accumulated mutations. Note, we did not include *Ap. californica* in this analysis, as the gene annotation file and TEs annotation file of *Ap. californica* is inconsistent.

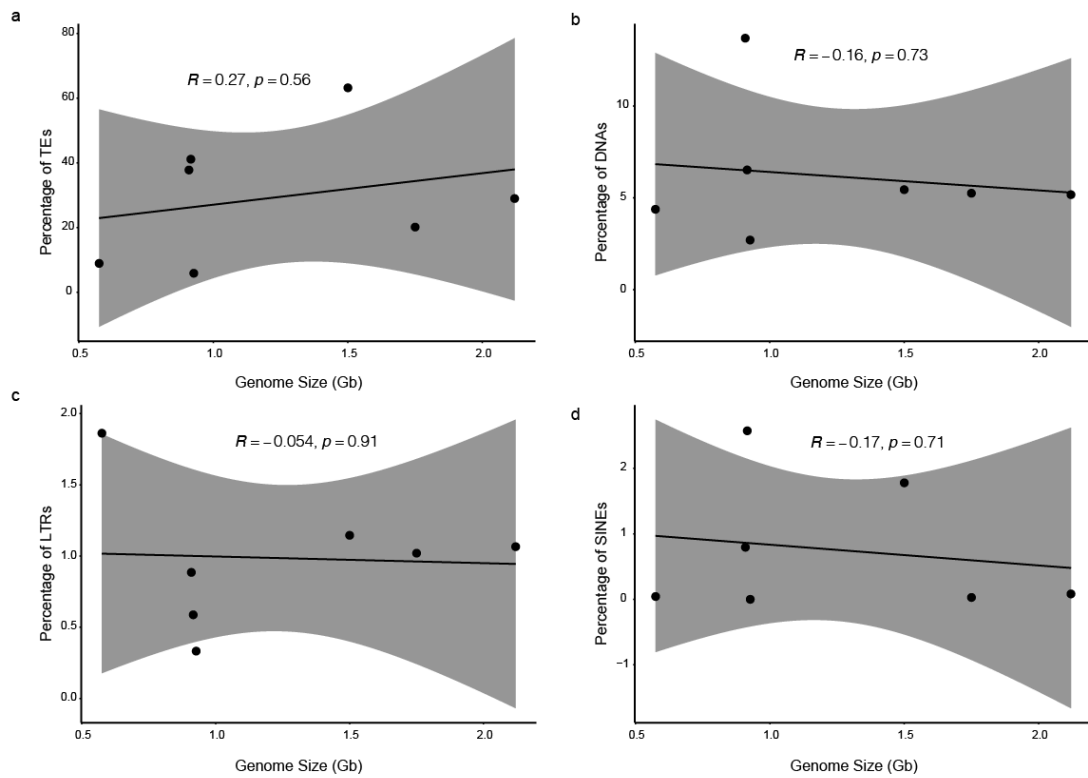


Fig S27. Relationship between total TE coverage (a), DNAs coverage (b), LTRs coverage (c), SINEs coverage (d), and genome size in Heterobranchia species. TE: transposable element; DNAs: DNA transposons; LTR: Long Terminal Repeat; SINEs: Short Interspersed Nuclear Elements.

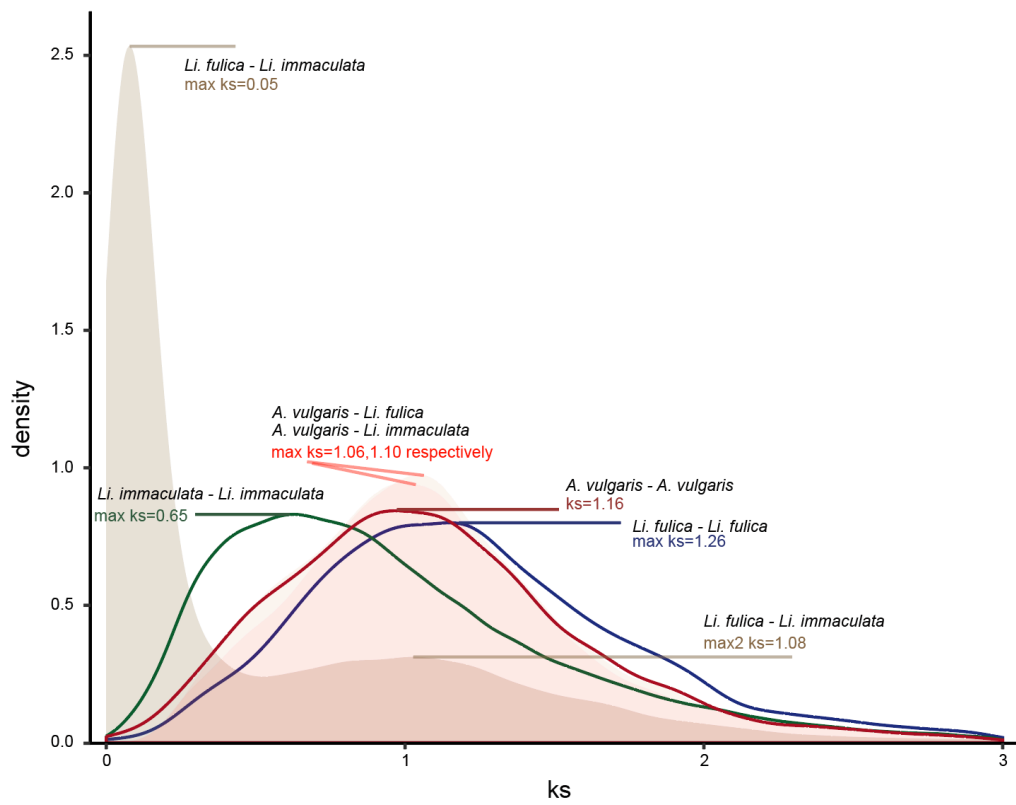


Fig S28. Frequency distributions of synonymous substitutions ( $K_s$ ) for homologous gene pairs identified using MCScanX in comparisons of *Arion vulgaris*, *Lissachatina (Achatina) fulica* and *Li. immaculata*. Shaded areas represent speciation events and solid lines represent whole-genome duplication events.

Table S1. Sequencing data generated for *Arion vulgaris* genome assembly and annotation.

Library type	Platform	Total bases (Gb)	Application
Short reads	HiSeq X Ten	56.83	Genome survey and genomic base correction
Linked reads	HiSeq X Ten	138.21	Genome survey, genomic base correction and genome assembly
Nanopore	Nanopore PromethION	74.94	Genome assembly
HiC	HiSeq X Ten	135.28	Chromosome construction
mRNA-Seq	NovaSeq 6000	6.75	Genome annotation

Table S2. Statistics of assembly at different assemble stages.

	wtdbg2	wtdbg2-ntEdit-Scaff10X	instaGRAAL
Characteristic	Nanopore reads	Nanopore reads, short reads, linked reads	Nanopore reads, short reads, linked reads, HiC reads
Total Contig length (Gb)	1.535 (6,109)	1.543 (6,109)	1.541 (7,076)
Contig N50 (Mb)	4.466 (93)	4.488 (93)	8.603 (49)
Longest Contig (Mb)	23.976	24.090	39.732
Total scaffold length (Gb)	-	1.543 (5,786)	1.543 (6,751)
Scaffold N50(Mb)	-	7.660 (55)	64.342 (10)
Longest Scaffold (Mb)	-	34.170	114.239
N's per 100 kbp (bp)	-	2.09	2.09
Complete BUSCOs	0.84	0.90	0.91
Complete and single-copy BUSCOs	0.82	0.84	0.85
Complete and duplicated BUSCOs	0.02	0.06	0.06
Fragmented BUSCOs (F)	0.03	0.02	0.02
Missing BUSCOs (M)	0.13	0.08	0.07

Note: The number in brackets indicates the number of corresponding scaffolds and contigs.

Table S3: Species used for comparative genomics analysis and data citations in this study.

Clade	Species	GenBank/Database accession
H	<i>Arion vulgaris</i>	GCA_020796225.1
H	<i>Lissachatina (Achatina) immaculata</i> <sup>1</sup>	GCA_009760885.1
H	<i>Li. fulica</i> <sup>2</sup>	<a href="https://gigadb.org/dataset/100647">gigadb.org/dataset/100647</a>
H	<i>Aplysia californica</i> <sup>3</sup>	GCA_000002075.2 (For gene family, phylogeny, repeats analysis) <a href="https://www.dnazoo.org/assemblies/Aplysia_californica">https://www.dnazoo.org/assemblies/Aplysia_californica</a> (For macrosynteny analysis)
H	<i>Biomphalaria glabrata</i> <sup>4</sup>	GCA_000457365.1
H	<i>Elysia chlorotica</i> <sup>5</sup>	GCA_003991915.1
H	<i>Radix auricularia</i> <sup>6</sup>	GCA_002072015.1
C	<i>Pomacea canaliculata</i> <sup>7</sup>	GCA_004794335.1
C	<i>Marisa cornuarietis</i> <sup>7</sup>	GCA_004794655.1
C	<i>Conus consors</i> <sup>8</sup>	GCA_004193615.1
C	<i>Lanistes nyassanus</i> <sup>7</sup>	GCA_004794575.1
V	<i>Haliotis rufescens</i> <sup>9</sup>	GCA_003343065.1
N	<i>Chrysomallon squamiferum</i> <sup>10</sup>	GCA_012295275.1
P	<i>Lottia gigantea</i> <sup>11</sup>	GCA_000327385.1
B	<i>Argopecten purpuratus</i> <sup>12</sup>	<a href="https://gigadb.org/dataset/100419">gigadb.org/dataset/100419</a>
B	<i>Saccostrea glomerata</i> <sup>13</sup>	GCA_003671525.1

Note: Clade H, C, V, N, P represents Heterobranchia, Caenogastropoda, Vetigastropoda, Neomphalina, and Patellogastropoda respectively. B: Bivalvia



Table S4: Evidence weight file used for *Arion vulgaris* gene prediction.

Item	weight
Protein	
<i>Lissachatina (Achatina) fulica</i>	4
<i>Biomphalaria glabrata</i>	2
<i>Elysia chlorotica</i>	2
<i>Haliotis rufescens</i>	2
<i>Pomacea canaliculata</i>	2
<i>Aplysia californica</i>	2
<i>ab initio</i> prediction	1
Transcript	6

Table S5: Statistics of predicted protein-coding genes in the *Arion vulgaris* genome.

Gene set	Number	Average transcript length(bp)	Average CDs length(bp)	Average exon per gene	Average exon length(bp)	Average intron length(bp)
homolog prediction						
<i>Lissachatina (Achatina) fulica</i>	22,718	10,568.96	1,202.44	4.60	261.53	2,603.49
<i>Aplysia californica</i>	12,046	15,833.02	1,412.34	7.39	191.22	2,258.25
<i>Biomphalaria glabrata</i>	16,264	15,628.49	1,325.09	7.06	187.81	2,362.09
<i>Elysia chlorotica</i>	13,069	13,604.99	1,204.99	6.42	187.82	2,289.63
<i>Haliotis rufescens</i>	21,393	5,791.92	863.26	2.82	305.68	2,701.96
<i>Pomacea canaliculata</i>	10,993	21,930.11	1,520.19	9.41	161.60	2,427.71
<i>De novo</i> prediction	44,246	12,668.01	1,146.02	5.08	225.44	2,821.68
transcriptome prediction	161,478	62,36.37	910.84	2.14	424.68	4,652.03
final gene set	32,518	15,429.25	1,291.68	5.70	226.43	3,005.06

Table S6: Functional annotation of the predicted genes models.

	Number	Percent (%)
eggNOG-mapper	16,791	0.51636017
KEGG	9,642	0.296512701
InterPro	26,520	0.815548312
SwissProt	31,728	0.975705763
TrEMBL	31,429	0.966510856
Total	31,763	0.97678209

Table S7: Summary of gene family clustering of 16 mollusk species.

Species	total genes	genes in family	unassigned genes	family number	unique family	genes in unique family	genes per family
<i>Arion vulgaris</i>	32,518	30,636	1,882	13,333	253	881	2.30
<i>Lissachatina (Achatina) fulica</i>	23,726	23,073	653	12,390	79	310	1.86
<i>Li. Immaculata</i>	30,194	28,633	1,561	12,513	124	361	2.29
<i>Biomphalaria glabrata</i>	25,550	24,093	1,457	12,444	366	1,450	1.94
<i>Radix auricularia</i>	17,338	16,237	1,101	10,310	110	367	1.57
<i>Aplysia californica</i>	27,576	26,761	815	12,191	396	1,601	2.20
<i>Elysia chlorotica</i>	24,980	21,936	3,044	12,980	312	1,255	1.69
<i>Pomacea canaliculata</i>	18,273	17,728	545	12,042	47	159	1.47
<i>Marisa cornuarietis</i>	23,827	21,988	1,839	13,181	201	1,005	1.67
<i>Lanistes nyassanus</i>	20,938	19,651	1,287	12,475	138	555	1.58
<i>Conus consors</i>	17,715	17,023	692	7,654	224	611	2.22
<i>Chrysomallon squamiferum</i>	16,917	15,498	1,419	9,904	252	1,047	1.56
<i>Lottia gigantea</i>	23,340	21,940	1,400	12,165	408	2,794	1.80
<i>Haliotis rufescens</i>	48,956	44,060	4,896	14,580	1,622	8,303	3.02
<i>Saccostrea glomerata</i>	26,956	26,327	629	11,516	619	3,375	2.29
<i>Argopecten purpuratus</i>	24,705	22,797	1,908	12,568	501	2,069	1.81

Table S8: GO enrichment analysis of Stylommatophora specific gene families. For each GO subcategory, a  $2 \times 2$  contingency table was constructed by recording the numbers of genes included or not included in a category of ‘genome background’ genes and Stylommatophora-specific genes. Two-tailed Fisher’s exact test was used to calculate statistical significance. BP: biology process; MF: molecular function.

GO	Type	Function	P value	Gene number
GO:0072073	BP	kidney epithelium development	0.001666	17
GO:0001657	BP	ureteric bud development	0.002973	14
GO:0001823	BP	mesonephros development	0.003725	14
GO:0072163	BP	mesonephric epithelium development	0.003725	14
GO:0072164	BP	mesonephric tubule development	0.003725	14
GO:0001656	BP	metanephros development	0.029842	11
GO:2001014	BP	regulation of skeletal muscle cell differentiation	0.000401	8
GO:0009208	BP	pyrimidine ribonucleoside triphosphate metabolic process	0.002059	6
GO:0009147	BP	pyrimidine nucleoside triphosphate metabolic process	0.004520	6
GO:0009193	BP	pyrimidine ribonucleoside diphosphate metabolic process	4.46E-05	5
GO:0046048	BP	UDP metabolic process	4.46E-05	5
GO:0070305	BP	response to cGMP	4.46E-05	5
GO:0071321	BP	cellular response to cGMP	4.46E-05	5
GO:0009138	BP	pyrimidine nucleoside diphosphate metabolic process	0.000202	5
GO:0046036	BP	CTP metabolic process	0.000202	5
GO:0021512	BP	spinal cord anterior/posterior patterning	4.53E-06	4
GO:0045967	BP	negative regulation of growth rate	4.53E-06	4
GO:2000818	BP	negative regulation of myoblast proliferation	6.56E-05	4
GO:0001978	BP	regulation of systemic arterial blood pressure by carotid sinus baroreceptor feedback	0.000559	4
GO:0001982	BP	baroreceptor response to decreased systemic arterial blood pressure	0.000559	4
GO:0003025	BP	regulation of systemic arterial blood pressure by baroreceptor feedback	0.000559	4
GO:0048632	BP	negative regulation of skeletal muscle tissue growth	0.000559	4
GO:0014859	BP	negative regulation of skeletal muscle cell proliferation	0.003230	4
GO:0071444	BP	cellular response to pheromone	0.003230	4
GO:1902723	BP	negative regulation of skeletal muscle satellite cell proliferation	0.003230	4
GO:1902725	BP	negative regulation of satellite cell differentiation	0.003230	4
GO:0019236	BP	response to pheromone	0.048324	4
GO:2001015	BP	negative regulation of skeletal muscle cell differentiation	0.048324	4
GO:0004382	MF	guanosine-diphosphatase activity	1.03E-06	5
GO:0043273	MF	CTPase activity	1.03E-06	5
GO:0045134	MF	uridine-diphosphatase activity	7.82E-06	5

Table S9: GO enrichment analysis of Stylommatophora expansion genes. For each GO subcategory, a  $2 \times 2$  contingency table was constructed by recording the numbers of genes included or not included in a category of 'genome background' genes and Stylommatophora expansion genes. Two-tailed Fisher's exact test was used to calculate statistical significance. BP: biology process; CC: cellular component.

GO	Type	Function	P value	Gene number
GO:0032501	BP	multicellular organismal process	0.007600	832
GO:0050896	BP	response to stimulus	3.06E-05	807
GO:0007154	BP	cell communication	1.91E-08	534
GO:0023052	BP	signaling	1.88E-10	525
GO:0051239	BP	regulation of multicellular organismal process	0.010047	413
GO:0050793	BP	regulation of developmental process	0.004431	386
GO:0032879	BP	regulation of localization	1.15E-06	378
GO:0060429	BP	epithelium development	0.011049	282
GO:0045595	BP	regulation of cell differentiation	0.006087	277
GO:0035295	BP	tube development	0.005169	235
GO:0009791	BP	post-embryonic development	2.19E-09	229
GO:0040012	BP	regulation of locomotion	1.87E-11	219
GO:0042325	BP	regulation of phosphorylation	0.001597	206
GO:0048729	BP	tissue morphogenesis	0.012526	197
GO:0001932	BP	regulation of protein phosphorylation	0.000909	194
GO:0051241	BP	negative regulation of multicellular organismal process	0.021533	190
GO:0007267	BP	cell-cell signaling	6.96E-10	189
GO:0051270	BP	regulation of cellular component movement	5.06E-10	187
GO:0002009	BP	morphogenesis of an epithelium	0.017699	186
GO:0051093	BP	negative regulation of developmental process	0.008294	173
GO:2000145	BP	regulation of cell motility	8.67E-11	172
GO:0030334	BP	regulation of cell migration	5.97E-11	165
GO:0010562	BP	positive regulation of phosphorus metabolic process	0.021489	150
GO:0045937	BP	positive regulation of phosphate metabolic process	0.021489	150
GO:0042327	BP	positive regulation of phosphorylation	0.004501	144
GO:0001934	BP	positive regulation of protein phosphorylation	0.001870	140
GO:0007444	BP	imaginal disc development	0.007684	129
GO:0009314	BP	response to radiation	0.013070	126
GO:0002164	BP	larval development	5.71E-07	125
GO:0002119	BP	nematode larval development	5.18E-10	105
GO:0010212	BP	response to ionizing radiation	1.15E-11	80
GO:0040024	BP	dauer larval development	2.92E-23	75
GO:0010332	BP	response to gamma radiation	2.63E-17	61
GO:0034284	BP	response to monosaccharide	0.023006	46

GO:0009746	BP	response to hexose	0.008245	45
GO:0060259	BP	regulation of feeding behavior	0.004137	39
GO:1903998	BP	regulation of eating behavior	1.29E-05	36
GO:0043051	BP	regulation of pharyngeal pumping	2.01E-07	35
GO:0030952	BP	establishment or maintenance of cytoskeleton polarity	0.000244	30
GO:0007229	BP	integrin-mediated signaling pathway	0.039414	18
GO:0097201	BP	negative regulation of transcription from RNA polymerase II promoter in response to stress	0.037923	10
GO:0097458	CC	obsolete neuron part	1.29E-05	278
GO:0036477	CC	somatodendritic compartment	3.64E-05	186
GO:0044297	CC	cell body	5.61E-05	162
GO:0048471	CC	perinuclear region of cytoplasm	2.10E-07	156
GO:0043025	CC	neuronal cell body	0.000129	150

---

Table S10: GO enrichment analysis of Stylommatophora contraction genes. For each GO subcategory, a  $2 \times 2$  contingency table was constructed by recording the numbers of genes included or not included in a category of 'genome background' genes and Stylommatophora expansion genes. Two-tailed Fisher's exact test was used to calculate statistical significance. BP: biology process; CC: cellular component.

GO	Type	Function	P value	Gene number
GO:1903961	BP	positive regulation of anion transmembrane transport	0.0001013	5
GO:1903959	BP	regulation of anion transmembrane transport	0.0406884	5
GO:2000118	BP	regulation of sodium-dependent phosphate transport	0.0019846	4
GO:0010966	BP	regulation of phosphate transport	0.0110806	4
GO:0030647	BP	aminoglycoside antibiotic metabolic process	0.0124327	3
GO:0042431	BP	indole metabolic process	0.0124327	3
GO:1901128	BP	gentamycin metabolic process	0.0124327	3
GO:1901684	BP	arsenate ion transmembrane transport	0.0124327	3
GO:1903797	BP	positive regulation of inorganic anion transmembrane transport	0.0124327	3
GO:2000187	BP	positive regulation of phosphate transmembrane transport	0.0124327	3
GO:0072733	BP	response to staurosporine	0.0309616	3
GO:0072734	BP	cellular response to staurosporine	0.0309616	3
GO:2000185	BP	regulation of phosphate transmembrane transport	0.0309616	3
GO:0072686	CC	mitotic spindle	0.0025335	8
GO:0031526	CC	brush border membrane	0.0216566	6
GO:0015370	MF	solute:sodium symporter activity	0.0405823	6
GO:0005436	MF	sodium:phosphate symporter activity	0.0011070	4



Table S11: GO enrichment analysis of Stylommatophora positively selected genes. For each GO subcategory, a  $2 \times 2$  contingency table was constructed by recording the numbers of genes included or not included in a category of 'genome background' genes and Stylommatophora positive selected genes. Two-tailed Fisher's exact test was used to calculate statistical significance. BP: biology process; MF: molecular function.

GO	Type	Function	P value	Gene number
GO:0031641	BP	regulation of myelination	0.014147	4
GO:0007130	BP	synaptonemal complex assembly	0.006699	3
GO:0004525	MF	ribonuclease III activity	0.000227	2
GO:0032296	MF	double-stranded RNA-specific ribonuclease activity	0.000227	2
GO:0031643	BP	positive regulation of myelination	3.40E-05	3

Table S12: GO enrichment analysis of *Arion vulgaris* species-specific and unassigned genes. For each GO subcategory, a 2 × 2 contingency table was constructed by recording the numbers of genes included or not included in a category of ‘genome background’ genes and *A. vulgaris* species-specific genes. Two-tailed Fisher’s exact test was used to calculate statistical significance. BP: biology process; CC: cellular component; MF: molecular function.

GO	Type	Function	P value	Gene number
GO:0019747	BP	regulation of isoprenoid metabolic process	1.46E-05	5
GO:0043226	CC	Organelle	0.016755	259
GO:0043229	CC	intracellular organelle	0.004138	250
GO:0043227	CC	membrane-bounded organelle	0.000500	221
GO:0043231	CC	intracellular membrane-bounded organelle	0.001050	205
GO:0030312	CC	external encapsulating structure	1.81E-05	10
GO:0015016	MF	[heparan sulfate]-glucosamine N-sulfotransferase activity	0.005871	3

Fig S13: GO enrichment analysis of *Arion vulgaris* expansion genes. For each GO subcategory, a 2 × 2 contingency table was constructed by recording the numbers of genes included or not included in a category of ‘genome background’ genes and *A. vulgaris* species-specific genes. Two-tailed Fisher’s exact test was used to calculate statistical significance. BP: biology process; CC: cellular component; MF: molecular function.

GO	Type	Function	P value	Gene number
GO:0009607	BP	response to biotic stimulus	0.035756	309
GO:0043207	BP	response to external biotic stimulus	0.00511	308
GO:0051707	BP	response to other organism	0.008506	306
GO:0030029	BP	actin filament-based process	0.034447	251
GO:0002164	BP	larval development	0.000921	243
GO:0030036	BP	actin cytoskeleton organization	0.045313	224
GO:0009617	BP	response to bacterium	0.008316	222
GO:0071396	BP	cellular response to lipid	0.003975	212
GO:0048545	BP	response to steroid hormone	0.009836	175
GO:0009410	BP	response to xenobiotic stimulus	0.030106	162
GO:0002237	BP	response to molecule of bacterial origin	0.010789	133
GO:0031099	BP	Regeneration	8.95E-05	131
GO:0016042	BP	lipid catabolic process	8.40E-05	129
GO:0071383	BP	cellular response to steroid hormone stimulus	5.19E-06	127
GO:0032496	BP	response to lipopolysaccharide	0.012052	126
GO:0061008	BP	hepaticobiliary system development	0.001825	126
GO:0001889	BP	liver development	0.003526	123
GO:0031960	BP	response to corticosteroid	0.000336	123
GO:0010212	BP	response to ionizing radiation	0.011636	120
GO:0051384	BP	response to glucocorticoid	0.000126	118
GO:0042440	BP	pigment metabolic process	2.00E-05	116
GO:0008202	BP	steroid metabolic process	0.035907	111
GO:0071466	BP	cellular response to xenobiotic stimulus	0.009945	101
GO:0097306	BP	cellular response to alcohol	0.031057	99
GO:0010675	BP	regulation of cellular carbohydrate metabolic process	0.002573	91
GO:0040024	BP	dauer larval development	4.26E-09	89
GO:0016999	BP	antibiotic metabolic process	2.91E-05	89
GO:0031100	BP	animal organ regeneration	9.21E-07	85
GO:0033013	BP	tetrapyrrole metabolic process	9.16E-12	84
GO:0006805	BP	xenobiotic metabolic process	3.87E-05	82
GO:0071385	BP	cellular response to glucocorticoid stimulus	5.91E-09	81
GO:0062014	BP	negative regulation of small molecule metabolic process	1.33E-07	81
GO:0071384	BP	cellular response to corticosteroid stimulus	1.33E-07	81
GO:0045833	BP	negative regulation of lipid metabolic process	6.69E-06	80

GO:0006720	BP	isoprenoid metabolic process	4.34E-06	79
GO:0007588	BP	excretion	4.56E-08	78
GO:0010332	BP	response to gamma radiation	2.48E-08	77
GO:0034754	BP	cellular hormone metabolic process	0.001421	77
GO:0006778	BP	porphyrin-containing compound metabolic process	0	76
GO:0019217	BP	regulation of fatty acid metabolic process	0.000113	76
GO:0006721	BP	terpenoid metabolic process	2.62E-08	75
GO:0042168	BP	heme metabolic process	0	74
GO:0001523	BP	retinoid metabolic process	6.92E-10	73
GO:0016101	BP	diterpenoid metabolic process	3.18E-09	73
GO:0002526	BP	acute inflammatory response	1.83E-11	70
GO:0010677	BP	negative regulation of cellular carbohydrate metabolic process	4.12E-11	69
GO:0045912	BP	negative regulation of carbohydrate metabolic process	1.29E-09	69
GO:0009225	BP	nucleotide-sugar metabolic process	1.16E-07	68
GO:0042447	BP	hormone catabolic process	0	66
GO:0051187	BP	obsolete cofactor catabolic process	5.02E-09	64
GO:0006953	BP	acute-phase response	0	63
GO:0071361	BP	cellular response to ethanol	2.40E-07	63
GO:0071392	BP	cellular response to estradiol stimulus	2.33E-08	62
GO:0006706	BP	steroid catabolic process	3.21E-11	61
GO:0045939	BP	negative regulation of steroid metabolic process	3.95E-09	61
GO:0006011	BP	UDP-glucose metabolic process	0	60
GO:0045922	BP	negative regulation of fatty acid metabolic process	0	60
GO:0008210	BP	estrogen metabolic process	0	60
GO:0042537	BP	benzene-containing compound metabolic process	9.16E-12	60
GO:0017001	BP	antibiotic catabolic process	7.33E-11	60
GO:0018879	BP	biphenyl metabolic process	0	59
GO:0009812	BP	flavonoid metabolic process	0	59
GO:0042178	BP	xenobiotic catabolic process	3.18E-09	59
GO:0033015	BP	tetrapyrrole catabolic process	0	58
GO:0006711	BP	estrogen catabolic process	0	57
GO:0006787	BP	porphyrin-containing compound catabolic process	0	57
GO:0006789	BP	bilirubin conjugation	0	57
GO:0009698	BP	phenylpropanoid metabolic process	0	57
GO:0009804	BP	coumarin metabolic process	0	57
GO:0042167	BP	heme catabolic process	0	57
GO:0046149	BP	pigment catabolic process	0	57
GO:0046226	BP	coumarin catabolic process	0	57
GO:0046271	BP	phenylpropanoid catabolic process	0	57
GO:0051552	BP	flavone metabolic process	0	57
GO:0052695	BP	cellular glucuronidation	0	57

GO:0052696	BP	flavonoid glucuronidation	0	57
GO:0052697	BP	xenobiotic glucuronidation	0	57
GO:0070980	BP	biphenyl catabolic process	0	57
GO:1904223	BP	regulation of glucuronosyltransferase activity	0	57
GO:1904224	BP	negative regulation of glucuronosyltransferase activity	0	57
GO:2001030	BP	negative regulation of cellular glucuronidation	0	57
GO:0006063	BP	uronic acid metabolic process	0	57
GO:0019585	BP	glucuronate metabolic process	0	57
GO:2001029	BP	regulation of cellular glucuronidation	0	57
GO:0042573	BP	retinoic acid metabolic process	3.67E-11	57
GO:0034663	CC	endoplasmic reticulum chaperone complex	0	61
GO:0070069	CC	cytochrome complex	2.02E-10	60
GO:0046982	MF	protein heterodimerization activity	0.042895	212
GO:0031406	MF	carboxylic acid binding	0.039347	106
GO:0008194	MF	UDP-glycosyltransferase activity	0.001112	89
GO:0005080	MF	protein kinase C binding	0	85
GO:0005496	MF	steroid binding	7.70E-10	79
GO:0033293	MF	monocarboxylic acid binding	7.33E-10	75
GO:0005504	MF	fatty acid binding	0	66
GO:0001972	MF	retinoic acid binding	0	60
GO:0005501	MF	retinoid binding	9.16E-12	60
GO:0019840	MF	isoprenoid binding	2.29E-11	60
GO:0015020	MF	glucuronosyltransferase activity	5.96E-11	59

---

Table S14: GO enrichment analysis of *Arion vulgaris* contraction genes. For each GO subcategory, a  $2 \times 2$  contingency table was constructed by recording the numbers of genes included or not included in a category of 'genome background' genes and Stylommatophora expansion genes. Two-tailed Fisher's exact test was used to calculate statistical significance. BP: biology process; CC: cellular component.

GO	Type	Function	P value	Gene number
GO:0006811	BP	ion transport	0.002125	105
GO:0007186	BP	G protein-coupled receptor signaling pathway	3.21E-11	75
GO:0006820	BP	anion transport	0.007451	57
GO:0098656	BP	anion transmembrane transport	0.027645	41
GO:0007218	BP	neuropeptide signaling pathway	6.11E-09	27
GO:0042749	BP	regulation of circadian sleep/wake cycle	0.024221	16
GO:0034308	BP	primary alcohol metabolic process	0.001884	15
GO:0042572	BP	retinol metabolic process	0.001194	12
GO:0007567	BP	parturition	0.001688	11
GO:0097070	BP	ductus arteriosus closure	0.003929	10
GO:0032757	BP	positive regulation of interleukin-8 production	0.032495	10
GO:0070493	BP	thrombin-activated receptor signaling pathway	0.003427	8
GO:0033212	BP	iron import into cell	0.000132	7
GO:0033215	BP	reductive iron assimilation	0.000132	7
GO:0042817	BP	pyridoxal metabolic process	0.013186	7
GO:0006710	BP	androgen catabolic process	0.001959	6
GO:0015746	BP	citrate transport	0.048716	6
GO:0071944	CC	cell periphery	0.000104	276
GO:0044425	CC	obsolete membrane part	0.005055	272
GO:0005886	CC	plasma membrane	6.04E-05	266
GO:0031224	CC	intrinsic component of membrane	1.37E-11	201
GO:0016021	CC	integral component of membrane	6.87E-11	194
GO:0044459	CC	obsolete plasma membrane part	0.000562	179
GO:0031226	CC	intrinsic component of plasma membrane	4.58E-12	140
GO:0005887	CC	integral component of plasma membrane	1.37E-11	135
GO:0031233	CC	intrinsic component of external side of plasma membrane	2.26E-05	11
GO:0033573	CC	high-affinity iron permease complex	0.000132	7
GO:1905862	CC	ferroxidase complex	0.000132	7
GO:0038023	MF	signaling receptor activity	2.29E-12	83
GO:0060089	MF	molecular transducer activity	6.87E-12	83
GO:0004888	MF	transmembrane signaling receptor activity	2.29E-09	64
GO:0004930	MF	G protein-coupled receptor activity	1.31E-09	44
GO:0008509	MF	anion transmembrane transporter activity	0.020526	38
GO:0008757	MF	S-adenosylmethionine-dependent methyltransferase	0.006238	29

		activity		
GO:0008168	MF	methyltransferase activity	0.041717	29
GO:0001653	MF	peptide receptor activity	6.24E-07	24
GO:0008528	MF	G protein-coupled peptide receptor activity	6.24E-07	24
GO:0008188	MF	neuropeptide receptor activity	1.55E-06	22
GO:0004745	MF	retinol dehydrogenase activity	2.19E-08	13
GO:0004957	MF	prostaglandin E receptor activity	3.10E-05	10
GO:0004953	MF	icosanoid receptor activity	0.000219	10
GO:0004954	MF	prostanoid receptor activity	0.000219	10
GO:0004955	MF	prostaglandin receptor activity	0.000219	10
GO:0016725	MF	oxidoreductase activity, acting on CH or CH2 groups	0.000219	10
GO:0016722	MF	oxidoreductase activity, oxidizing metal ions	0.016515	9
GO:0042895	MF	obsolete antibiotic transmembrane transporter activity	0.016515	9
GO:0004322	MF	ferroxidase activity	0.004202	7
GO:0016724	MF	oxidoreductase activity, oxidizing metal ions, oxygen as acceptor	0.004202	7
GO:0004031	MF	aldehyde oxidase activity	0.013186	7
GO:0004732	MF	pyridoxal oxidase activity	0.013186	7
GO:0016404	MF	15-hydroxyprostaglandin dehydrogenase (NAD+) activity	0.013186	7
GO:0016623	MF	oxidoreductase activity, acting on the aldehyde or oxo group of donors, oxygen as acceptor	0.013186	7
GO:0015137	MF	citrate transmembrane transporter activity	0.048716	6

---

Fig S15: GO enrichment analysis of *Arion vulgaris* positive selected genes. For each GO subcategory, a  $2 \times 2$  contingency table was constructed by recording the numbers of genes included or not included in a category of 'genome background' genes and *A. vulgaris* species-specific genes. Two-tailed Fisher's exact test was used to calculate statistical significance. BP: biology process; CC: cellular component; MF: molecular function.

GO	Type	Function	P value	Gene number
GO:0051099	BP	positive regulation of binding	0.001346	4
GO:0009794	BP	regulation of mitotic cell cycle, embryonic	0.001047	2
GO:0042509	BP	regulation of tyrosine phosphorylation of STAT protein	0.009004	2
GO:0042531	BP	positive regulation of tyrosine phosphorylation of STAT protein	2.92E-05	2
GO:0051769	BP	regulation of nitric-oxide synthase biosynthetic process	1.01E-06	2
GO:0051770	BP	positive regulation of nitric-oxide synthase biosynthetic process	8.16E-12	2
GO:1900744	BP	regulation of p38MAPK cascade	0.002657	2
GO:0002731	BP	negative regulation of dendritic cell cytokine production	1.54E-10	1
GO:0002825	BP	regulation of T-helper 1 type immune response	0.001017	1
GO:0002826	BP	negative regulation of T-helper 1 type immune response	1.54E-10	1
GO:0006535	BP	cysteine biosynthetic process from serine	1.54E-10	1
GO:0006565	BP	L-serine catabolic process	0.001017	1
GO:0007260	BP	tyrosine phosphorylation of STAT protein	1.54E-10	1
GO:0009092	BP	homoserine metabolic process	0.001017	1
GO:0009093	BP	cysteine catabolic process	0.024452	1
GO:0009403	BP	toxin biosynthetic process	5.28E-06	1
GO:0018343	BP	protein farnesylation	0.024452	1
GO:0019346	BP	transsulfuration	0.001017	1
GO:0019448	BP	L-cysteine catabolic process	0.024452	1
GO:0030223	BP	neutrophil differentiation	0.001017	1
GO:0031959	BP	mineralocorticoid receptor signaling pathway	1.54E-10	1
GO:0032693	BP	negative regulation of interleukin-10 production	0.024452	1
GO:0034050	BP	host programmed cell death induced by symbiont	1.54E-10	1
GO:0034477	BP	U6 snRNA 3'-end processing	1.54E-10	1
GO:0035406	BP	histone-tyrosine phosphorylation	1.54E-10	1
GO:0035409	BP	histone H3-Y41 phosphorylation	1.54E-10	1
GO:0035771	BP	interleukin-4-mediated signaling pathway	1.54E-10	1
GO:0035905	BP	ascending aorta development	0.001017	1
GO:0035910	BP	ascending aorta morphogenesis	0.001017	1
GO:0036015	BP	response to interleukin-3	0.001017	1
GO:0036016	BP	cellular response to interleukin-3	0.001017	1
GO:0038113	BP	interleukin-9-mediated signaling pathway	1.54E-10	1
GO:0038114	BP	interleukin-21-mediated signaling pathway	1.54E-10	1
GO:0038155	BP	interleukin-23-mediated signaling pathway	0.001017	1



GO:0042262	BP	DNA protection	0.024452	1
GO:0042976	BP	activation of Janus kinase activity	5.28E-06	1
GO:0043418	BP	homocysteine catabolic process	5.28E-06	1
GO:0044375	BP	regulation of peroxisome size	5.28E-06	1
GO:0044854	BP	plasma membrane raft assembly	0.024452	1
GO:0044857	BP	plasma membrane raft organization	0.024452	1
GO:0045219	BP	regulation of FasL production	1.54E-10	1
GO:0045221	BP	negative regulation of FasL production	1.54E-10	1
GO:0045625	BP	regulation of T-helper 1 cell differentiation	5.28E-06	1
GO:0045626	BP	negative regulation of T-helper 1 cell differentiation	1.54E-10	1
GO:0051316	BP	attachment of spindle microtubules to kinetochore involved in meiotic chromosome segregation	0.024452	1
GO:0060031	BP	mediolateral intercalation	5.28E-06	1
GO:0060355	BP	positive regulation of cell adhesion molecule production	0.001017	1
GO:0060399	BP	positive regulation of growth hormone receptor signaling pathway	0.001017	1
GO:0070106	BP	interleukin-27-mediated signaling pathway	1.54E-10	1
GO:0070757	BP	interleukin-35-mediated signaling pathway	1.54E-10	1
GO:0070813	BP	hydrogen sulfide metabolic process	0.001017	1
GO:0070814	BP	hydrogen sulfide biosynthetic process	5.28E-06	1
GO:0071104	BP	response to interleukin-9	1.54E-10	1
GO:0071355	BP	cellular response to interleukin-9	1.54E-10	1
GO:0098756	BP	response to interleukin-21	1.54E-10	1
GO:0098757	BP	cellular response to interleukin-21	1.54E-10	1
GO:0106005	BP	RNA 5'-cap (guanine-N7)-methylation	1.54E-10	1
GO:1902724	BP	positive regulation of skeletal muscle satellite cell proliferation	1.54E-10	1
GO:1902728	BP	positive regulation of growth factor dependent skeletal muscle satellite cell proliferation	1.54E-10	1
GO:1905457	BP	negative regulation of lymphoid progenitor cell differentiation	0.001017	1
GO:2000049	BP	positive regulation of cell-cell adhesion mediated by cadherin	0.024452	1
GO:2000668	BP	regulation of dendritic cell apoptotic process	1.54E-10	1
GO:2000670	BP	positive regulation of dendritic cell apoptotic process	1.54E-10	1
GO:2000974	BP	negative regulation of pro-B cell differentiation	0.001017	1
GO:2000978	BP	negative regulation of forebrain neuron differentiation	0.001017	1
GO:0044450	CC	obsolete microtubule organizing center part	0.000180	4
GO:0034451	CC	centriolar satellite	0.000630	2
GO:0005827	CC	polar microtubule	0.024452	1
GO:0005965	CC	protein farnesyltransferase complex	5.28E-06	1
GO:0008275	CC	gamma-tubulin small complex	0.001017	1
GO:0031533	CC	mRNA cap methyltransferase complex	1.54E-10	1
GO:0004527	MF	exonuclease activity	0.000160	3
GO:0004028	MF	3-chloroallyl aldehyde dehydrogenase activity	1.54E-10	1

GO:0004030	MF	aldehyde dehydrogenase [NAD(P)+] activity	0.024452	1
GO:0004122	MF	cystathionine beta-synthase activity	1.54E-10	1
GO:0004124	MF	cysteine synthase activity	1.54E-10	1
GO:0004311	MF	farnesyltranstransferase activity	0.001017	1
GO:0004362	MF	glutathione-disulfide reductase activity	1.54E-10	1
GO:0004482	MF	mRNA (guanine-N7-)-methyltransferase activity	1.54E-10	1
GO:0004660	MF	protein farnesyltransferase activity	5.28E-06	1
GO:0005119	MF	smoothened binding	5.28E-06	1
GO:0005131	MF	growth hormone receptor binding	0.001017	1
GO:0005143	MF	interleukin-12 receptor binding	1.54E-10	1
GO:0015038	MF	glutathione disulfide oxidoreductase activity	1.54E-10	1
GO:0016662	MF	oxidoreductase activity, acting on other nitrogenous compounds as donors, cytochrome as acceptor	5.28E-06	1
GO:0017005	MF	3'-tyrosyl-DNA phosphodiesterase activity	0.001017	1
GO:0031730	MF	CCR5 chemokine receptor binding	5.28E-06	1
GO:0035400	MF	histone tyrosine kinase activity	1.54E-10	1
GO:0035401	MF	histone kinase activity (H3-Y41 specific)	1.54E-10	1
GO:0048020	MF	CCR chemokine receptor binding	0.001017	1
GO:0050421	MF	nitrite reductase (NO-forming) activity	5.28E-06	1
GO:0051428	MF	peptide hormone receptor binding	1.54E-10	1
GO:0070259	MF	tyrosyl-DNA phosphodiesterase activity	0.024452	1
GO:0071568	MF	UFM1 transferase activity	5.28E-06	1
GO:0071820	MF	N-box binding	0.001017	1
GO:0098605	MF	selenocystathionine beta-synthase activity	1.54E-10	1
GO:0098622	MF	selenodiglutathione-disulfide reductase activity	1.54E-10	1
GO:0098809	MF	nitrite reductase activity	5.28E-06	1

---

Table S16. Number of gene pairs derived from different modes of duplication in seven Heterobranchia species. WGD: whole-genome duplication, DSD: dispersed duplication, TRD: transposed duplication, PD: proximal duplication, TD: tandem duplication.

Species	DSD	PD	TD	TRD	WGD
<i>Lissachatina (Achatina) fulica</i>	4,685	1,246	2,660	3,492	5,251
<i>Li. immaculata</i>	10,905	1,056	2,653	3,585	4,021
<i>Arion vulgaris</i>	10,035	1,030	4,432	3,268	4,999
<i>Radix auricularia</i>	6,940	601	1,776	56	0
<i>Biomphalaria glabrata</i>	13,954	663	2,281	128	0
<i>Elysia chlorotica</i>	11,655	573	1,410	78	0
<i>Aplysia californica</i>	6,111	270	7,488	24	0

Table S17. Number of positive selected gene pairs ( $Ka/Ks > 1$ ) derived from different modes in seven Heterobranchia species. The brackets show the percentages of gene pairs with  $Ka/Ks > 1$  in the corresponding duplicated gene pairs. WGD: whole-genome duplication, DSD: dispersed duplication, TRD: transposed duplication, PD: proximal duplication, TD: tandem duplication.

Species	DSD	PD	TD	TRD	WGD
<i>Lissachatina (Achatina) fulica</i>	369 (7.8%)	47 (3.77%)	119 (4.47%)	51 (1.46%)	2 (0.04%)
<i>Li. Immaculata</i>	1,777 (16.3%)	39 (3.69%)	168 (6.33%)	88 (2.45%)	12 (0.30%)
<i>Arion vulgaris</i>	2,041 (20.3%)	47 (4.56%)	648 (14.62%)	46 (1.41%)	4 (0.08%)
<i>Radix auricularia</i>	38 (0.55%)	3 (0.50%)	9 (0.51%)	0	0
<i>Biomphalaria glabrata</i>	391 (2.80%)	2 (0.30%)	19 (0.83%)	0	0
<i>Elysia chlorotica</i>	400 (3.43%)	11 (1.92%)	20 (1.42%)	0	0
<i>Aplysia californica</i>	61 (1.00%)	8 (2.96%)	289 (3.86%)	0	0

Table S18. Identified repeat classes in the *Arion vulgaris* genome. TEs: transposable elements; DNAs: DNA transposons; LTR: Long Terminal Repeat; LINEs: Long Interspersed Nuclear Elements; RC: Rolling Circle; SINEs: Short Interspersed Nuclear Elements;

Repeat type	Repeat size (bp)	% Of genome
TEs	941,228,221	61.08
LINE	561,000,000	36.39
DNA	83,828,750	5.44
SINE	27,404,787	1.78
LTR	17,654,456	1.15
RC	10,002,034	0.65
Satellite	8,990,082	0.58
Low	864,980	0.06
Unknown	274,000,000	17.76
Tandem Repeats	326,134,855	21.14
Total	1,158,337,369	75.09

Table S19. Number of different TEs types in different genic regions in six Heterobranchia species. Down: 1kb down-3'UTR, up: 2kb up-5'UTR. %Total shows the percentage of different type of TEs in all types of TEs in correspond genic region. DNAs: DNA transposons; LTR: Long Terminal Repeat; LINEs: Long Interspersed Nuclear Elements; SINEs: Short Interspersed Nuclear Elements;

	Region	DNA	% Total	LINE	% Total	LTR	% Total	SINE	% Total	Unknown	% Total	Total
<i>Lissachatina (Achatina) fulica</i>	down	4,504	0.20	12,863	0.57	4,231	0.19	9	0.0004	43	0.002	22,615
	exon	1,449	0.19	4,270	0.56	1,572	0.20	3	0.0004	43	0.006	7,686
	intron	119,603	0.19	363,202	0.59	109,786	0.18	236	0.0004	1,442	0.002	618,462
	up	11,519	0.21	30,769	0.56	9,749	0.18	18	0.0003	157	0.003	54,531
<i>Li. immaculata</i>	down	5,076	0.44	3,734	0.32	1,307	0.11	76	0.0066	75	0.006	11,539
	exon	902	0.33	533	0.20	942	0.35	33	0.0121	42	0.015	2,728
	intron	76,437	0.39	80,549	0.41	18,104	0.09	1,151	0.0059	1,210	0.006	194,095
	up	12,532	0.43	9,851	0.34	3,131	0.11	221	0.0076	207	0.007	28,914
<i>Arion vulgaris</i>	down	4,004	0.09	23,634	0.51	1,156	0.02	1,501	0.0322	15,716	0.337	46,632
	exon	1,046	0.27	1,411	0.36	774	0.20	70	0.0180	468	0.120	3,888
	intron	55,905	0.07	421,028	0.53	14,180	0.02	25,272	0.0317	272,362	0.341	797,679
	up	8,980	0.09	51,112	0.52	2,287	0.02	2,839	0.0288	32,124	0.325	98,719
<i>Biomphalaria glabrata</i>	down	2,823	0.13	5,571	0.26	381	0.02	1,857	0.0878	4,455	0.211	21,156
	exon	262	0.12	249	0.11	168	0.08	55	0.0249	120	0.054	2,212
	intron	29,924	0.13	58,573	0.26	3,700	0.02	20,958	0.0920	48,195	0.212	227,804
	up	6,936	0.15	13,318	0.28	870	0.02	3,154	0.0671	10,181	0.217	47,012
<i>Elysia chlorotica</i>	down	4,616	0.45	1,629	0.16	1,525	0.15	45	0.0043	209	0.020	10,365
	exon	1,980	0.33	1,765	0.30	1,113	0.19	5	0.0008	87	0.015	5,919
	intron	51,275	0.43	16,579	0.14	16,836	0.14	659	0.0056	2,896	0.024	118,330

	up	12,641	0.44	4,555	0.16	4,164	0.14	119	0.0041	601	0.021	28,753
	down	3,101	0.36	1,438	0.17	46	0.01	92	0.0106	3,725	0.431	8,645
<i>Radix auricularia</i>	exon	4	0.22	6	0.33	1	0.06	1	0.0556	6	0.333	18
	intron	41,113	0.34	15,372	0.13	1,051	0.01	1,126	0.0094	58,573	0.487	120,394
	up	9,837	0.37	3,762	0.14	171	0.01	206	0.0078	11,845	0.446	26,535

## References

- 1 Liu, C. *et al.* Giant African snail genomes provide insights into molluscan whole-genome duplication and aquatic-terrestrial transition. *Mol. Ecol. Resour.* 21, 478-494 (2020).
- 2 Guo, Y. *et al.* A chromosomal-level genome assembly for the giant African snail *Achatina fulica*. *Gigascience* 8, giz124 (2019).
- 3 Di Palma, F. *et al.* *The Draft Genome of Aplysia californica*, <https://www.ncbi.nlm.nih.gov/nuccore/AASC00000000.3> (2014).
- 4 Adema, C. M. *et al.* Whole genome analysis of a schistosomiasis-transmitting freshwater snail. *Nat. Commun.* 8, 15451 (2017).
- 5 Cai, H. *et al.* A draft genome assembly of the solar-powered sea slug *Elysia chlorotica*. *Sci. Data* 6, 190022 (2019).
- 6 Schell, T. *et al.* An annotated draft genome for *Radix auricularia* (Gastropoda, Mollusca). *Genome Biol. Evol.* 9, 585-592 (2017).
- 7 Sun, J. *et al.* Signatures of divergence, invasiveness, and terrestrialization revealed by four apple snail genomes. *Mol. Biol. Evol.* 36, 1507-1520 (2019).
- 8 Andreson, R. *et al.* Gene content of the fish-hunting cone snail *Conus consors*. Preprint at <https://doi.org/10.1101/590695> (2019).
- 9 Masonbrink, R. E. *et al.* An annotated genome for *Haliotis rufescens* (red abalone) and resequenced green, pink, pinto, black, and white abalone species. *Genome Biol. Evol.* 11, 431-438 (2019).
- 10 Sun, J. *et al.* The Scaly-foot snail genome and implications for the origins of biomineralised armour. *Nat. Commun.* 11, 1657 (2020).
- 11 Simakov, O. *et al.* Insights into bilaterian evolution from three spiralian genomes. *Nature* 493, 526-531 (2013).
- 12 Li, C. *et al.* Draft genome of the Peruvian scallop *Argopecten purpuratus*. *Gigascience* 7, giy031 (2018).
- 13 Powell, D. *et al.* The genome of the oyster *Saccostrea* offers insight into the environmental resilience of bivalves. *DNA Res.* 25, 655-665 (2018).



## 4. Discussion

### 4.1 From *A. vulgaris* to mollusk mitogenomes

The mitogenome of *A. vulgaris* measures 14,547 bp, and the characteristics are mostly consistent with other reported stylommatophoran mitogenomes (**Publication I**, Doğan et al. 2020). However, the phylum Mollusca in particular, is replete with examples of extraordinary variation in mitogenome architecture, molecular functioning and intergenerational transmission (Ghiselli et al. 2021). Mollusk mitogenomes vary widely in size, from approximately 13.6–14.1 kb (Heterobranchia gastropods and scaphopods) to approximately 51.0 kb in length (bivalves, e.g., *Scapharca broughtonii* (Schrenck, 1867)), which is probably a result of expansion of the largest non-coding region (Liu et al. 2013). We detected rearrangement events in the *A. vulgaris* mitogenome: the trnW-trnY and trnE-trnQ-rrnS-trnM-trnL2-ATP8-trnN-ATP6-trnR gene clusters (**Publication I**, Doğan et al. 2020). Indeed, changes in the gene order are most common for tRNAs in mollusks, especially in families like Haliotidae, which exhibit largely conserved synteny of the protein-coding genes but exhibit variable tRNA locations (Xin et al. 2011). Recent evidence has emerged that mitochondrial DNA is more than an evolutionary bystander, and the environment can act as a selective force promoting haplotype variation and potentially altering mitochondrial function and heat production (Ballard & Pichaud 2014; Ruiz-Pesini & Mishmar 2004; Shtolz 2019). Across all mollusks, there has been strong purifying selection in maintaining the minimal set of 37 mitochondrial genes mitogenome (Ghiselli et al. 2021a). However, diversifying selection was observed in our study with different codons in different stylommatophoran taxa (**Publication I**, Doğan et al. 2020). We speculate that this pattern might also be present in other mollusks, and might be related to species' different adaptive potentials.

Due to the abundant number of mitochondria per cell in multicellular eukaryotes (Cole 2016) and relatively short molecular length (11-50 Kb in bilaterians) (Zardoya 2020), mitochondrial DNA is comparatively easy to capture and amplify during sequencing. Thus, the requirement of DNA quality and quantity is less stringent compared to whole-genome sequencing, and large-scale sequencing, e.g., multiplexing of up to 300 samples on the same run is straightforward on the Illumina platform, scalable and portable. Nanopore-based sequencing offers the ability to generate the entire mitochondrial genome in a single contig (Zascavage et al. 2019). Thus, for mollusks, especially the most diverse gastropods, the large-scale sequencing for hundreds of species remains the best option for understanding evolutionary processes and reconstructing their relationships, especially among orders and families. In the near future, the genome of *A. vulgaris*, combined with the upcoming genomes of other stylommatophorans, will help us to resolve the phylogeny of stylommatophoran land snails and slugs, as well as their history of divergence and prosperity on land (Saadi & Wade 2019). However, for more distant groups, long-branch attraction, sequence saturation and substitution rate heterogeneity are well worth caution (Stöger & Schrödl 2013; Varney et al. 2021) and nuclear genomic analyses appear more promising.

## 4.2 From gastropod to molluscan phylogeny

Gastropoda is the most diverse molluscan group with around 100,000 species and their habitats range from the deep sea, to fresh water and land. The relationships among the recognized seven subclasses: Patellogastropoda, Neomphalina, Cocculiniformia, Neritimorpha, Vetigastropoda, Heterobranchia and Caenogastropoda have been extensively discussed for a long time. In recent studies, based on transcriptome data, a division into five subclasses has been suggested: They assigned the deep-sea Neomphalina and Cocculiniformia to Vetigastropoda (*s.l.*) resulting in five subclasses of Gastropoda: Patellogastropoda, Vetigastropoda, Neritimorpha, Caenogastropoda and Heterobranchia (Ponder et al. 2020; Uribe et al. 2022). The key debates are mainly focused on deep relationships, whether Patellogastropoda are sister to all gastropods (Orthogastropoda hypothesis), or Patellogastropoda sister to Vetigastropoda (Psilogastropoda *s.l.* hypothesis) (Cunha & Giribet 2019; Golikov & Starobogatov 1975). Morphological inferences in ambiguously supported Orthogastropoda (Haszprunar 1988; Ponder & Lindberg 1997), while different molecular datasets, e.g., mitogenomic, large transcriptomic, as well as whole-genomic analyses recovered either Orthogastropoda or Psilogastropoda (Chen et al. 2020; Cunha & Giribet 2019; Uribe et al. 2022; Uribe et al. 2019; Zapata et al. 2014). For example, early large-scale transcriptome studies, e.g., Kocot et al. (2011) and Zapata et al. (2014), rejected Orthogastropoda. In contrast, Uribe et al. (2019) recovered Orthogastropoda using newly sequenced patellogastropod mitogenomes. The newly sequenced mitogenomes displayed shorter branches than the one of previously used *Lottia*, and with gene organizations more similar to that of the hypothetical gastropod ancestor. Moreover, Uribe et al. (2022) used transcriptome data from twelve taxa belonging to clades with little or no prior representation in previous studies and again recovered Orthogastropoda. Applying mitogenomes for reconstructing deep phylogeny may have serious complications (e.g., Stöger & Schrödl 2013), but also transcriptomic results depend on and were limited by the quality and quantity of data and analyses.

Phylogenomics, i.e., the use of large arrays of genome sequences to infer phylogenetic relationships, has emerged over the last few years and is increasingly used in molecular studies of taxa relationships (Debray et al. 2019). Single-copy orthologous genes (SOGs) have long been recognized as ideal molecular markers for inferring relationships of previously unresolved lineages (Wu et al. 2006). Genome-wide resources have allowed us to obtain SOGs from gastropods from several major clades and to test their relationship depending on the number of SOGs considered. In our studies, we obtained both of the above mentioned phylogenetic relationships: when using 223 SOGs, we recovered Orthogastropoda, and when the gene number increased to thousands of genes (with less species coverage), the topology changed to Psilogastropoda *s.l.* (**Publication II**, Chen & Schrödl 2022), which is also supported by other whole-genomic studies (Lan et al. 2021; Sun et al. 2020). A recent study by Uribe et al. (2022) explored the impact of missing data on the inferred phylogenetic hypotheses using transcriptome data; removing rapidly evolving sites resulted in Orthogastropoda, while full data recovered Patellogastropoda sister to Vetigastropoda.

It will be crucial for future research to optimize both, the quality and quantity of genomic data.

Furthermore, the currently available genomes/transcriptomes and mitogenomes constitute a far-from-optimal taxon set for resolving deep gastropod phylogeny. The major groups such as Patellogastropoda, Neomphalina and Vetigastropoda have for a long time been represented by single or only few members, and should thus be sampled much more densely, in a best-case scenario including the entire diversity of early branching subclades. As we have shown (**Publication II**, Chen & Schrödl 2022), robust reconstruction of deep gastropod relationships will depend on using sufficient molecular data from a large and balanced taxon set.

On a larger scale, the deep mollusk phylogeny suffers from the same dilemma. The currently available genomes create a non-optimal set of taxa for resolving deeper phylogenetic nodes with frequently studied octopus, bivalves and without e.g., Caudofoveata and Solenogastres for comparison (**Fig 3**). The extreme variability and diversity of mollusks and its more than 500 million years of evolutionary history make mollusk difficult to be represented by a few single groups and species. Thus, a broad and dense sampling and a thoughtful selection of taxa are needed for the future resolving of the biology and evolutionary history of mollusks (Sigwart et al. 2021).

### **4.3 Deep dive into the *Arion vulgaris* genome**

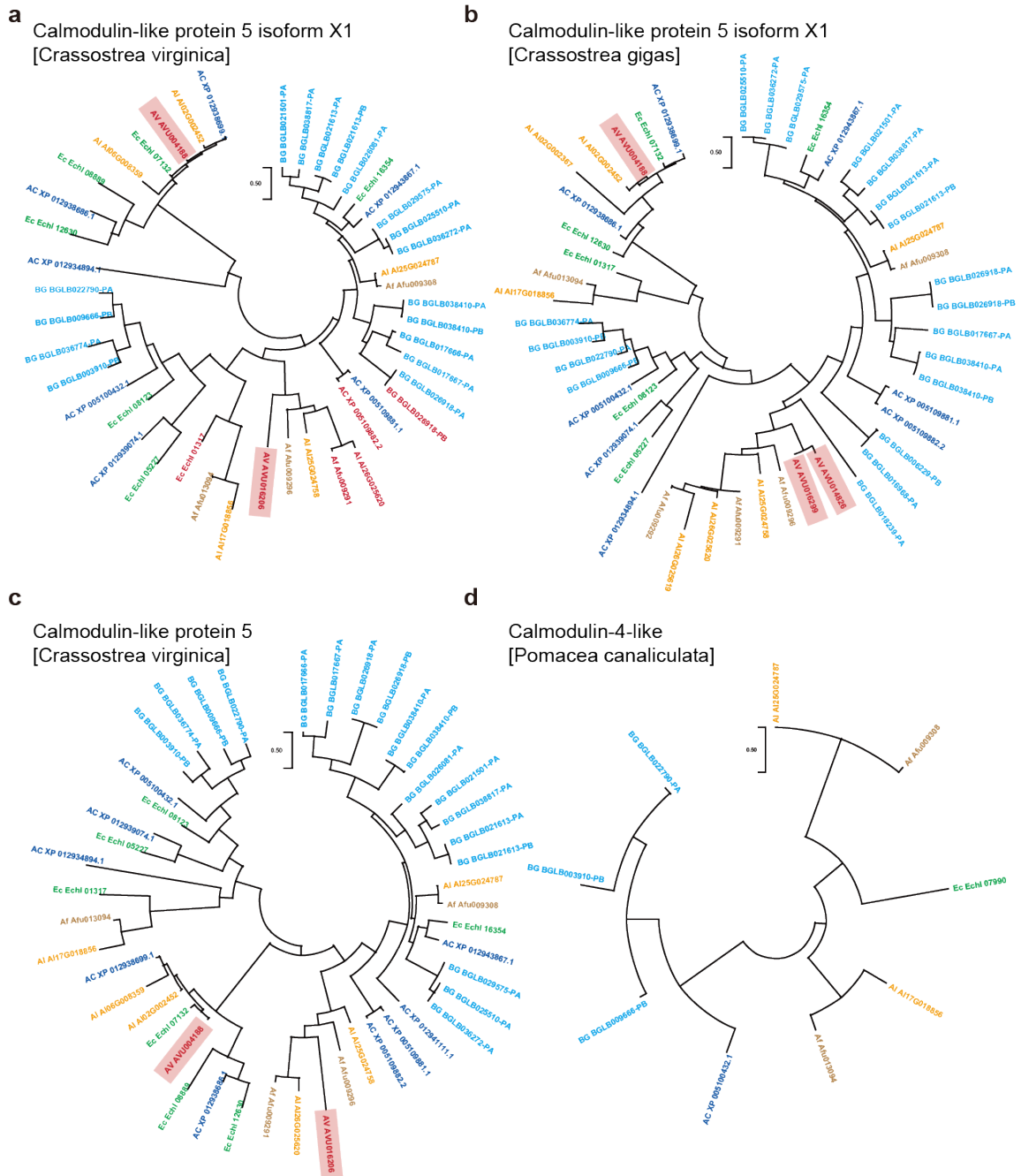
Today, although studies in genome biology tend to be descriptive, sequencing genomes and metagenomes, analyzing epigenomes and transcriptomes allow inference of evolutionary histories, and cataloging potential loci associated with particular functions (McGuire et al. 2020). Genomes provide the most fundamental databases of genetic information that can be directly used for comparison, whether it is between species that are distantly related, different individuals of the same species, or different stages of the same individual. These comparisons offer insights for genetics and evolution, and provide clues for subsequent functional assessment. In this regard, as examples, we explored two gene families of our greater interest (Calmodulin-like protein and Mucin protein), based on the comparative genomics of *A. vulgaris* and other Heterobranchia species.

#### **4.3.1 Calmodulin-like proteins are reduced/lost in *A. vulgaris***

Molluscan shell formation and evolution has been an ongoing concern for malacologists (Addadi et al. 2006; McDougall & Degnan 2018). Recent studies integrated genomics and proteomics into the study of molluscan biomineralization and enabled the identification of genes associated with the shell formation process (Kocot et al. 2016), e.g., an analysis of a deep transcriptome from the mantle tissue of *Patella vulgata* revealed candidate biomineralizing genes (Werner et al. 2013). As member of a basically shelled clade of terrestrial pulmonates, the genome of the slug *A. vulgaris* provides a good material for comparison with snail species to reveal the evolution and

reduction of molluscan shells. In our initial exploration, we downloaded protein sequences associated with mollusk shells from NCBI and searched for homologous genes in six gastropods (GenBank 2021): *A. vulgaris* (no visible shell), *Lissachatina fulica* (Bowdich, 1822) and *L. immaculata* (with coiled external shell), *Biomphalaria glabrata* (Say, 1818) (with coiled external shell), *Elysia chlorotica* Gould, 1870 (with a shell as veliger larva), and *Aplysia californica* Cooper, 1863 (adults have a small, flat, vestigial shell). We found a contraction of Calmodulin-like protein 5 and a complete loss of Calmodulin-like protein 4 in *A. vulgaris*' genome (**Fig 6**). In addition, we found that all these genes were expanded in the fresh water snail *B. glabrata*. Specifically, Calmodulin-like protein 5 isoform X1, identified from *Crassostrea virginica* (Gmelin, 1791), has two copies in *A. vulgaris*, while it has five to six in the land snails *Lissachatina*, 18 copies in *B. glabrata*, seven copies in *E. chlorotica* and eight copies in *A. californica* (**Fig 6a**). For the Calmodulin-like protein 4, it is present in all five species except *A. vulgaris*, with two sea slugs having one copy each, two land snails having two copies each, and *B. glabrata* with three copies (**Fig 6d**). However, further validation is needed as to what the actual effects of the reduction and loss of these genes would be.

In general, shell formation is a complex process controlled by the highly coordinated expression of hundreds of genes, and the regulated secretion of proteins and other macromolecules (Kocot et al. 2016). Molluscan shells have evolved in many different forms and shell loss/ degeneration has occurred independently in several lineages (Clark et al. 2020; Tanner et al. 2017; Zer 2009). Future studies need to combine genomic, transcriptomic, and proteomic data from more representative species sets and may use gene editing technologies e.g., CRISPR/Cas9 (Perry & Henry 2015) to decipher the key genes and pathways of shell formation revealing the secrets of mollusk shells. However, for molluscan functional assessment, additional resources e.g., cell lines are vital for gene manipulation studies and for molecule particle tracking, which are still lacking for mollusks (Clark et al. 2020).

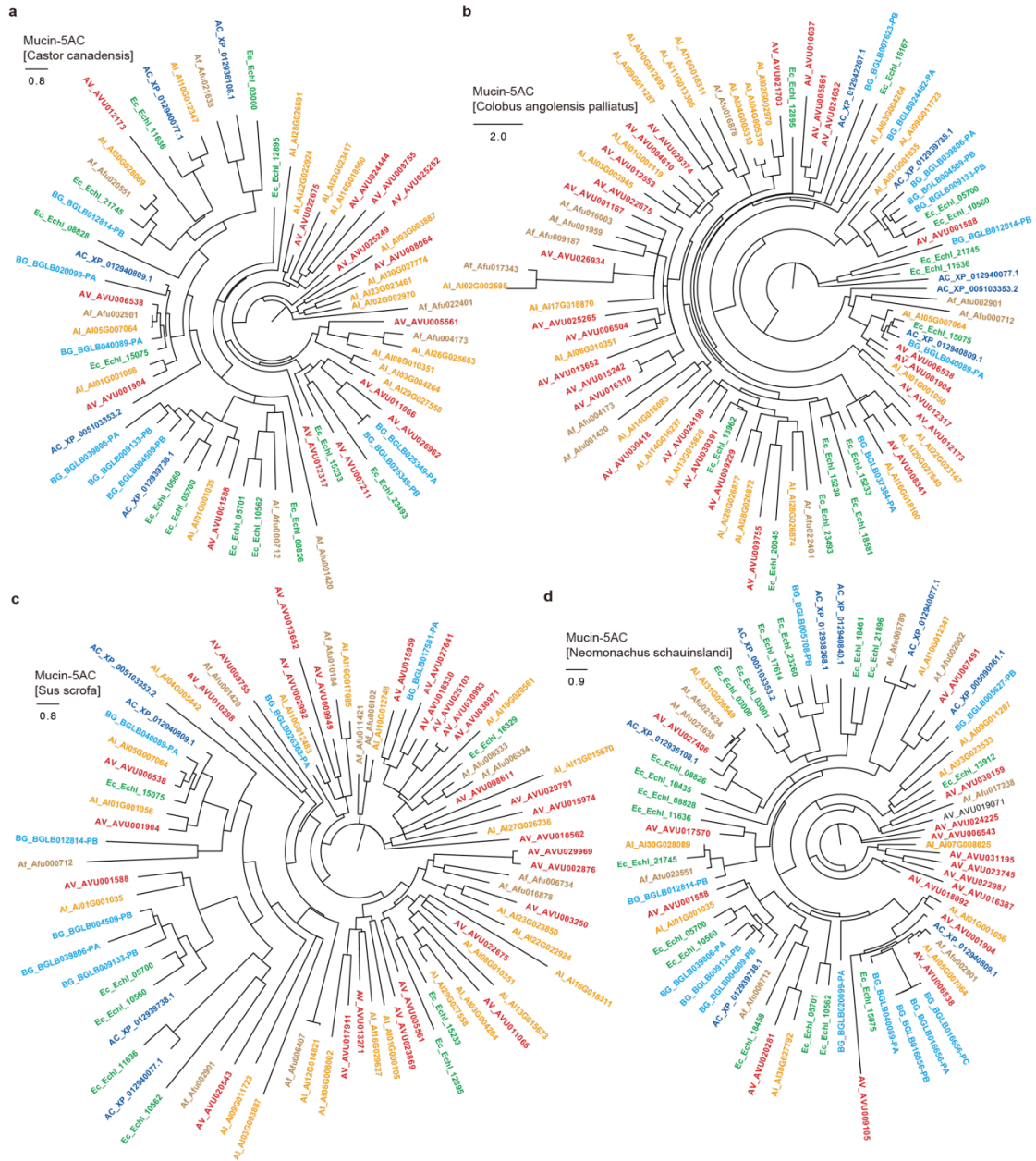


**Fig 6. Calmodulin-like proteins are reduced/lost in *Arion vulgaris* genome.** Different colors represent different species, red: *A. vulgaris*, brown: *Lissachatina fulica*, orange: *L. immaculata*, light blue: *Biomphalaria glabrata*, green: *Elysia chlorotica*, dark blue: *Aplysia californica*.

#### 4.3.2 Mucin related genes are expanded in land snails and slugs

Mucus secretion is probably one of the most important survival and adaptation skills for land snails and slugs. The viscoelastic mucus, with its adhesive and lubricating properties, allows slugs and snails to cling tenaciously to many different surfaces, prevents them from dehydrating and makes them unattractive to potential predators (Hämäläinen & Järvinen 2012). Mucus and its derived components, such as achacin, achatina CRP and mytimacin-AF, show high activity against Gram-positive and Gram-negative bacteria, viruses and yeasts, and prevent gastropods from being infected by microorganisms while also contributing to their innate immunity (Cilia & Fratini 2018; Ito et al. 2011). Recently, snail mucins have also become a lucrative source of innovation with wide ranging applications across chemistry, biology, biotechnology, and biomedicine, e.g., as skin care products, wound healing agents, surgical glues, and to combat gastric ulcers (McDermott et al. 2021).

As with our study of shell-associated genes, we blasted genes of six gastropod species *A. vulgaris*, *L. fulica* and *L. immaculata* (land), *B. glabrata* (fresh water), *E. chlorotica* (intertidal, marine), and *A. californica* (intertidal, marine) against mucin genes from NCBI database (GenBank 2021). In agreement with expectations, we found substantial expansion of the mucin family in land snails and slugs. Interestingly, *A. vulgaris* has even more copies of these genes compared to the two investigated land snails (**Fig 7**). The expansion of mucins thus may be one of the key factors in the water-land transition and territorial adaptation of land snails and slugs. Again, functionality and transcriptomic and proteomic data are needed for further validation. However, as we showed, genomic data can provide us a fundamental resource for identification of possible genes, and to accommodate the molecular hypothesis regarding its biology, physiology, behavior and adaptability.



**Fig 7. Mucin-5AC gene families are expanded in land snails and slugs.** Different colors represent different species, red: *A. vulgaris*, brown: *Lissachatina fulica*, orange: *L. immaculata*, light blue: *Biomphalaria glabrata*, green: *Elysia chlorotica*, dark blue: *Aplysia californica*.

## 5. Conclusion and outlook

During my dissertation, colleagues and I generated the first whole genomic resources for any land slug. Our target species, *Arion vulgaris*, is highly important in terms of economic, ecological, and evolutionary status. By using mitogenomic data, we revealed evolutionary patterns of Stylommatophora land slugs and snails and their phylogenetic relationships. Then, we identified the single-copy orthologous genes from gastropods with whole-genome data available and discussed the influence of using different scales of dataset and representative species on deep gastropod relationships. Furthermore, we explored the molecular evolutionary features, as well as the potential mechanisms of *A. vulgaris* and stylommatophoran land slugs and snails in their aquatic-terrestrial transitions and land adaptations. Overall, our results set the stage for further functional analysis of genomic innovations, and also serve as a foundation for *A. vulgaris* population studies, as well as for comparative genomic studies of mollusks.

As mollusk genomic studies have gradually increased in recent years, we have come to understand the difficulty and complexity of molluscan genomes, which may relate to any potential steps, from difficulties in sampling and sample preservation, DNA extraction, poor sequencing output, and the genome itself being very large and complex with highly repetitive sequences and heterozygous. Nevertheless, mollusk genomic studies have made great advances in uncovering genetic innovations, evolution and adaptations of specific species/groups, as well as in reconstructing complex and controversial genetic relationships in the past few years. Throughout the current molluscan genomic database, available resources remain largely limited, unevenly distributed in quality, and underrepresented in species and most higher taxa. However, this situation should change soon with the ongoing sequencing projects and consortia, e.g., the Earth Biogenome Project (Lewin et al. 2018), which has the ambitious aim to sequence all eukaryotic species in the next decade; the Darwin Tree of Life (Darwin Tree of Life Project 2022), which aims to sequence the genomes of 70,000 species of eukaryotic organisms in Britain and Ireland, the Global Invertebrate Genomics Alliance (Scientists et al. 2014), which expected to significantly boost the genomic resources of non-vertebrates in the near future. Moreover, the genomic database of MolluscDB (Caurcel et al. 2021) also compiles and integrates current molluscan genomic/transcriptomic resources, provides convenient tools for multi-level integrative and comparative genomic analyses.

Future studies, in addition to building a balanced and robust mollusk genomic database, should also focus on mining and re-studying available data. A combination of multi-omics data, functional validation and application is also important for further studies (Klein et al. 2019). I expect that these growing genetic resources, combined with advancing paleontological, systematics, and developmental methods and new discoveries, will help us understand the vivid and fascinating evolutionary history of mollusks and the mysteries of life, genetics, and evolution.



## 6. References

- Addadi L, Joester D, Nudelman F, and Weiner S. 2006. Mollusk shell formation: a source of new concepts for understanding biomineralization processes. *Chemistry* 12:980-987. 10.1002/chem.200500980
- Adema CM. 2021. Sticky problems: extraction of nucleic acids from molluscs. *Philos Trans R Soc Lond B Biol Sci* 376:20200162. 10.1098/rstb.2020.0162
- Aktipis SW, Giribet G, Lindberg DR, and Ponder W. 2008. Gastropoda: an overview and analysis. *Phylogeny and Evolution of the Mollusca* (ed. Ponder W & Lindberg W). University of California Press. 201-237.
- Albertin CB, Medina-Ruiz S, Mitros T, Schmidbaur H, Sanchez G, Wang ZY, Grimwood J, Rosenthal JJC, Ragsdale CW, Simakov O and Rokhsar DS. 2022. Genome and transcriptome mechanisms driving cephalopod evolution. *Nat Commun* 13:2427. 10.1038/s41467-022-29748-w
- Albertin CB, Simakov O, Mitros T, Wang ZY, Pungor JR, Edsinger-Gonzales E, Brenner S, Ragsdale CW, and Rokhsar DS. 2015. The octopus genome and the evolution of cephalopod neural and morphological novelties. *Nature* 524:220-224. 10.1038/nature14668
- Alhakami H, Mirebrahim H, and Lonardi S. 2017. A comparative evaluation of genome assembly reconciliation tools. *Genome Biol* 18:93. 10.1186/s13059-017-1213-3
- Ballard JWO & Pichaud N. 2014. Mitochondrial DNA: More than an evolutionary bystander. *Funct Ecol* 28: 218-231. 10.1111/1365-2435.12177.
- Barghi N, Concepcion GP, Olivera BM, and Lluisma AO. 2016. Structural features of conopeptide genes inferred from partial sequences of the *Conus tribblei* genome. *Mol Genet Genomics* 291:411-422. 10.1007/s00438-015-1119-2
- Barnosky AD, Matzke N, Tomiya S, Wogan GO, Swartz B, Quental TB, Marshall C, McGuire JL, Lindsey EL, Maguire KC, Mersey B, and Ferrer EA. 2011. Has the Earth's sixth mass extinction already arrived? *Nature* 471:51-57. 10.1038/nature09678
- Belcaid M, Casaburi G, McAnulty SJ, Schmidbaur H, Suria AM, Moriano-Gutierrez S, Pankey MS, Oakley TH, Kremer N, Koch EJ, Collins AJ, Nguyen H, Lek S, Goncharenko-Foster I, Minx P, Sodergren E, Weinstock G, Rokhsar DS, McFall-Ngai M, Simakov O, Foster JS, and Nyholm SV. 2019. Symbiotic organs shaped by distinct modes of genome evolution in cephalopods. *Proc Natl Acad Sci U S A* 116:3030-3035. 10.1073/pnas.1817322116
- Botwright NA, Zhao M, Wang T, McWilliam S, Colgrave ML, Hlinka O, Li S, Suwansa-Ard S, Subramanian S, McPherson L, King H, Reverter A, Cook MT, McGrath A, Elliott NG, and Cummins SF. 2019. Greenlip abalone (*Haliotis laevis*) genome and protein analysis provides insights into maturation and spawning. *G3 (Bethesda)* 9:3067-3078. 10.1534/g3.119.400388
- Butler PG, Wanamaker AD, Scourse JD, Richardson CA, and Reynolds DJ. 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. 10.1016/j.palaeo.2012.01.016
- Caurcel C, Laetsch DR, Challis R, Kumar S, Gharbi K, and Blaxter M. 2021. MolluscDB: a genome and transcriptome database for molluscs. *Philos Trans R Soc Lond B Biol Sci* 376:20200157. 10.1098/rstb.2020.0157
- Chakraborty K, and Joy M. 2020. High-value compounds from the molluscs of marine and estuarine ecosystems as prospective functional food ingredients: An overview. *Food Res Int* 137:109637. 10.1016/j.foodres.2020.109637

- Chen Z, Doğan Ö, Guiguelmoni N, Guichard A, and Schrödl M. 2022. Pulmonate slug evolution is reflected in the de novo genome of *Arion vulgaris* Moquin-Tandon, 1855. *Sci Rep* 12:14226. 10.1038/s41598-022-18099-7
- Chen Z, Doğan Ö, Guiguelmoni N, Guichard A, and Schrödl M. 2020. The *de novo* genome of the “Spanish” slug *Arion vulgaris* Moquin-Tandon, 1855 (Gastropoda: Panpulmonata): massive expansion of transposable elements in a major pest species. *bioRxiv*. 10.1101/2020.11.30.403303
- Chen Z, Schrödl M. 2022. How many single-copy orthologous genes from whole genomes reveal deep gastropod relationships? *PeerJ* 10:e13285. 10.7717/peerj.13285
- Cilia G, and Fratini F. 2018. Antimicrobial properties of terrestrial snail and slug mucus. *J Complement Integr Med* 15:20170168. 10.1515/jcim-2017-0168
- Clark MS, Peck LS, Arivalagan J, Backeljau T, Berland S, Cardoso JCR, Caurcel C, Chapelle G, De Noia M, Dupont S, Gharbi K, Hoffman JI, Last KS, Marie A, Melzner F, Michalek K, Morris J, Power DM, Ramesh K, Sanders T, Sillanpaa K, Sleight VA, Stewart-Sinclair PJ, Sundell K, Telesca L, Vendrami DLJ, Ventura A, Wilding TA, Yarra T, and Harper EM. 2020. Deciphering mollusc shell production: the roles of genetic mechanisms through to ecology, aquaculture and biomimetics. *Biol Rev Camb Philos Soc* 95:1812-1837. 10.1111/brv.12640
- Cole LW. 2016. The Evolution of Per-cell Organelle Number. *Front Cell Dev Biol* 4:85. 10.3389/fcell.2016.00085
- Crick RE, Robison RA, and Johnson ME. 2020. Cambrian Period. Available at <https://www.britannica.com/science/Cambrian-Period> (accessed 11th August 2022).
- Cunha TJ, and Giribet G. 2019. A congruent topology for deep gastropod relationships. *Proc Biol Sci* 286:20182776. 10.1098/rspb.2018.2776
- Darwin Tree of Life Project C. 2022. Sequence locally, think globally: The Darwin Tree of Life Project. *Proc Natl Acad Sci U S A* 119: e2115642118. 10.1073/pnas.2115642118
- Davison A, Neiman M. 2021. Mobilizing molluscan models and genomes in biology. *Phil Trans R Soc B*. 376: 20200163. 10.1098/rstb.2020.0163
- Debray K, Marie-Magdelaine J, Ruttink T, Clotault J, Foucher F, and Malecot V. 2019. Identification and assessment of variable single-copy orthologous (SCO) nuclear loci for low-level phylogenomics: a case study in the genus *Rosa* (Rosaceae). *BMC Evol Biol* 19:152. 10.1186/s12862-019-1479-z
- Doğan Ö, Schrödl M, Chen Z. 2020. The complete mitogenome of *Arion vulgaris* Moquin-Tandon, 1855 (Gastropoda: Stylommatophora): mitochondrial genome architecture, evolution and phylogenetic considerations within Stylommatophora. *PeerJ* 8:e8603. 10.7717/peerj.8603.
- 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, and Liu X. 2017. The pearl oyster *Pinctada fucata martensii* genome and multi-omic analyses provide insights into biomineralization. *Gigascience* 6:1-12. 10.1093/gigascience/gix059
- Ebbs ET, Loker ES, and Brant SV. 2018. Phylogeography and genetics of the globally invasive snail *Physa acuta* Draparnaud 1805, and its potential to serve as an intermediate host to larval digenetic trematodes. *BMC Evol Biol* 18:103. 10.1186/s12862-018-1208-z
- Formaggioni A, Plazzi F, and Passamonti M. 2022. Mito-nuclear coevolution and phylogenetic artifacts: the case of bivalve mollusks. *Sci Rep* 12:11040. 10.1038/s41598-022-15076-y

- Fried B, and Huffman JE. 2017. International Encyclopedia of Public Health. *Helminthic Diseases: Foodborne Trematode Infections* (ed. Quah SR). Academic Press. 327-332.
- GenBank. 2021. National Center for Biotechnology Information.
- Ghiselli F, Gomes-dos-Santos A, Adema C, Lopes-Lima M, Sharbrough J, and Boore J. 2021b. Molluscan mitochondrial genomes break the rules. *Phil Trans R Soc B* B376:20200159. 10.1098/rstb.2020.0159
- Giani AM, Gallo GR, Gianfranceschi L, and Formenti G. 2020. Long walk to genomics: History and current approaches to genome sequencing and assembly. *Comput Struct Biotechnol J* 18:9-19. 10.1016/j.csbj.2019.11.002
- Golikov AN and Starobogatov Y. 1975. Systematics of prosobranch gastropods. *Malacologia*. 15:185–232.
- Gomes-dos-Santos A, Lopes-Lima M, Castro LFC, and Froufe E. 2019. Molluscan genomics: the road so far and the way forward. *Hydrobiologia* 847:1705-1726. 10.1007/s10750-019-04111-1
- Guiglielmoni N, Rivera-Vicéns R, Koszul R, and Flot J-F. 2022. A deep dive into genome assemblies of non-vertebrate animals. *PCI* 2:e29. 10.24072/pcjournal.128
- Gundappa MK, Peñalosa C, Regan T, Boutet I, Tanguy A, Houston RD, Bean TP, and Macqueen DJ. 2022. Chromosome level reference genome for European flat oyster (*Ostrea edulis* L.). *Evol Appl* 00: 1-17. 10.1111/eva.13460
- Gupta AK, and Gupta UD. 2020. Animal Biotechnology. *Next generation sequencing and its applications* (ed. Verma AS & Singh A). Academic Press. 395-421.
- Hämäläinen EM, and Järvinen S. 2012. Snails: biology, ecology, and conservation. *Melanoides tuberculata: The history of an invader*. Nova Science Publisher's. 65-85.
- Haszprunar G. 1985. The Heterobranchia – a new concept of the phylogeny of the higher Gastropoda. *J Zool Syst Evol* 23:15-37. 10.1111/j.1439-0469.1985.tb00567.x
- Haszprunar G. 1988. On the origin and evolution of major gastropod groups, with special reference to the streptoneura. *J Molluscan Stud* 54:367-441. 10.1093/mollus/54.4.367
- Haszprunar G. 1992. The first molluscs - small animals. *Bolletino di zoologia* 59:1-16. 10.1080/11250009209386641
- Haszprunar G. 2000. Is the Aplacophora monophyletic? A cladistic analysis. *Am. Malacol. Bull* 15:115-130.
- Haszprunar G, Schander C, and Halanych KM. 2008. Relationships of Higher Molluscan Taxa. *Phylogeny and Evolution of the Mollusca* (ed. Ponder WF & Lindberg DR). University of California Press. 18-32.
- Haszprunar G, and Wanninger A. 2012. Molluscs. *Curr Biol* 22:R510-514. 10.1016/j.cub.2012.05.039
- Herbert RJH, Humphreys J, Davies CJ, Roberts C, Fletcher S, Crowe TP. 2016. Ecological impacts of non-native Pacific oysters (*Crassostrea gigas*) and management measures for protected areas in Europe. *Biodivers Conserv* 25: 2835-2865.
- Hochner B, and Glanzman DL. 2016. Evolution of highly diverse forms of behavior in molluscs. *Curr Biol* 26:R965-R971. 10.1016/j.cub.2016.08.047
- Hotaling S, Kelley JL, and Frandsen PB. 2021. Toward a genome sequence for every animal: Where are we now? *Proc Natl Acad Sci U S A* 118: e2109019118. 10.1073/pnas.2109019118
- Ip JC, Xu T, Sun J, Li R, Chen C, Lan Y, Han Z, Zhang H, Wei J, Wang H, Tao J, Cai Z, Qian PY, and Qiu JW. 2021. Host-endosymbiont genome integration in a deep-sea chemosymbiotic clam. *Mol Biol Evol* 38:502-518. 10.1093/molbev/msaa241

- Ito S, Shimizu M, Nagatsuka M, Kitajima S, Honda M, Tsuchiya T, and Kanzawa N. 2011. High molecular weight lectin isolated from the mucus of the giant African snail *Achatina fulica*. *Biosci Biotechnol Biochem* 75:20-25. 10.1271/bbb.100389
- IUCN. 2021. Red List. Available at <https://www.iucnredlist.org/resources/summary-statistics> (accessed 2022.08.20).
- Jörger KM, Stöger I, Kano Y, Fukuda H, Kneblsberger T, Schrödl M. 2010. On the origin of Acochlidia and other enigmatic euthyneuran gastropods, with implications for the systematics of Heterobranchia. *BMC Evol Biol* 10:323. 10.1186/1471-2148-10-323
- Kappes H, and Haase P. 2011. Slow, but steady: dispersal of freshwater molluscs. *Aquat Sci* 74:1-14. 10.1007/s00027-011-0187-6
- Kern EMA, Kim T, and Park J-K. 2020. The mitochondrial genome in nematode phylogenetics. *Front Ecol Evol* 8:250. 10.3389/fevo.2020.00250
- Klein AH, Ballard KR, Storey KB, Motti CA, Zhao M, and Cummins SF. 2019. Multi-omics investigations within the phylum Mollusca, class Gastropoda: from ecological application to breakthrough phylogenomic studies. *Brief Funct Genomics* 18:377-394. 10.1093/bfpg/elz017
- Kocot KM. 2016. On 20 years of Lophotrochozoa. *Org Divers Evol* 16:329-343. 10.1007/s13127-015-0261-3
- Kocot KM, Aguilera F, McDougall C, Jackson DJ, and Degnan BM. 2016. Sea shell diversity and rapidly evolving secretomes: insights into the evolution of biomineralization. *Front Zool* 13:23. 10.1186/s12983-016-0155-z
- Kocot KM, Cannon JT, Todt C, Citarella MR, Kohn AB, Meyer A, Santos SR, Schander C, Moroz LL, Lieb B & Halanych KM. 2011. Phylogenomics reveals deep molluscan relationships. *Nature* 477: 452-456. 10.1038/nature10382
- Kocot KM, Poustka AJ, Stöger I, Halanych KM, Schrödl M. 2020. New data from Monoplacophora and a carefully-curated dataset resolve molluscan relationships. *Sci Rep* 10: 101. 10.1038/s41598-019-56728-w
- Korshunova T, Martynov A, Bakken T, Evertsen J, Fletcher K, Mudianta IW, Saito H, Lundin K, Michael S, and Picton B. 2017. Polyphyly of the traditional family Flabellinidae affects a major group of Nudibranchia: aeolidacean taxonomic reassessment with descriptions of several new families, genera, and species (Mollusca, Gastropoda). *Zookeys* 717:1-139. 10.3897/zookeys.717.21885
- Lan Y, Sun J, Chen C, Sun Y, Zhou Y, Yang Y, Zhang W, Li R, Zhou K, Wong WC, Kwan YH, Cheng A, Bougouffa S, Van Dover CL, Qiu JW, and Qian PY. 2021. Hologenome analysis reveals dual symbiosis in the deep-sea hydrothermal vent snail *Gigantopelta aegis*. *Nat Commun* 12:1165. 10.1038/s41467-021-21450-7
- Leffler EM, Bullaughey K, Matute DR, Meyer WK, Segurel L, Venkat A, Andolfatto P, and Przeworski M. 2012. Revisiting an old riddle: what determines genetic diversity levels within species? *PLoS Biol* 10:e1001388. 10.1371/journal.pbio.1001388
- Lewin HA, Robinson GE, Kress WJ, Baker WJ, Coddington J, Crandall KA, Durbin R, Edwards SV, Forest F, Gilbert MTP, Goldstein MM, Grigoriev IV, Hackett KJ, Haussler D, Jarvis ED, Johnson WE, Patrinos A, Richards S, Castilla-Rubio JC, van Sluys MA, Soltis PS, Xu X, Yang H, and Zhang G. 2018. Earth BioGenome Project: Sequencing life for the future of life. *Proc Natl Acad Sci U S A* 115:4325-4333. 10.1073/pnas.1720115115
- Li Y, Nong W, Baril T, Yip HY, Swale T, Hayward A, Ferrier DEK, and Hui JHL. 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. 10.1186/s12864-020-07027-6
- Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES, and Dekker J. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326:289-293. 10.1126/science.1181369
- Liu C, Ren Y, Li Z, Hu Q, Yin L, Qiao X, Zhang Y, Xing L, Xi Y, Jiang F, Wang S, Huang C, Liu B, Wang H, Liu H, Wan F, Qian W, and Fan W. 2020. Giant African snail genomes provide insights into molluscan whole-genome duplication and aquatic-terrestrial transition. *Mol Ecol Resour* 21:478-494. 10.1101/2020.02.02.930693
- Liu YG, Kurokawa T, Sekino M, Tanabe T, and 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. 10.1016/j.cbd.2012.12.003
- Maeda T, Takahashi S, Yoshida T, Shimamura S, Takaki Y, Nagai Y, Toyoda A, Suzuki Y, Arimoto A, Ishii H, Satoh N, Nishiyama T, Hasebe M, Maruyama T, Minagawa J, Obokata J, and Shigenobu S. 2021. Chloroplast acquisition without the gene transfer in kleptoplastic sea slugs, *Plakobranthus ocellatus*. *Elife* 10: e60176. 10.7554/eLife.60176
- Masonbrink RE, Purcell CM, Boles SE, Whitehead A, Hyde JR, Seetharam AS, and Severin AJ. 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. 10.1093/gbe/evz006
- Masta SE, Longhorn SJ, and Boore JL. 2009. Arachnid relationships based on mitochondrial genomes: asymmetric nucleotide and amino acid bias affects phylogenetic analyses. *Mol Phylogenet Evol* 50:117-128. 10.1016/j.ympev.2008.10.010
- McDermott M, Cerullo AR, Parziale J, Achrak E, Sultana S, Ferd J, Samad S, Deng W, Braunschweig AB, and Holford M. 2021. Advancing discovery of snail mucins function and application. *Front Bioeng Biotechnol* 9:734023. 10.3389/fbioe.2021.734023
- McDougall C, and Degnan B. 2018. The evolution of mollusc shells. *Wiley Interdiscip Rev Dev Biol* 7:e313. 10.1002/wdev.313
- McGuire AL, Gabriel S, Tishkoff SA, Wonkam A, Chakravarti A, Furlong EEM, Treutlein B, Meissner A, Chang HY, López-Bigas N, Segal E & Kim J. 2020. The road ahead in genetics and genomics. *Nat Rev Genet.* 21:581-596. 10.1038/s41576-020-0272-6
- Morton B. 1983. Evolution and adaptive radiation in the Gastrochaenacea (Bivalvia). *J Molluscan Stud* 49:117-121.
- Mun S, Kim YJ, Markkandan K, Shin W, Oh S, Woo J, Yoo J, An H, and Han K. 2017. The whole-genome and transcriptome of the Manila Clam (*Ruditapes philippinarum*). *Genome Biol Evol* 9:1487-1498. 10.1093/gbe/evx096
- Nong W, Yu Y, Aase-Remedios ME, Xie Y, So WL, Li Y, Wong CF, Baril T, Law STS, Lai SY, Haimovitz J, Swale T, Chen SS, Kai ZP, Sun X, Wu Z, Hayward A, Ferrier DEK, and Hui JHL. 2022. Genome of the ramshorn snail *Biomphalaria straminea*-an obligate intermediate host of schistosomiasis. *Gigascience* 11: giac012. 10.1093/gigascience/giac012
- Orland C, Escalona M, Sahasrabudhe R, Marimuthu MPA, Nguyen O, Beraut E, Marshman B, Moore J, Raimondi P, and Shapiro B. 2022. A draft reference genome assembly of the critically endangered black Abalone, *Haliotis cracherodii*. *J Hered* XX:1-8. 10.1093/jhered/esac024

- Pardos-Blas JR, Irisarri I, Abalde S, Afonso CML, Tenorio MJ, and Zardoya R. 2021. The genome of the venomous snail *Lautoconus ventricosus* sheds light on the origin of conotoxin diversity. *Gigascience* 10: giab037. 10.1093/gigascience/giab037
- Parkhaev PY. 2008. The Early Cambrian Radiation of Mollusca. *Phylogeny and Evolution of the Mollusca* (ed. Winston FP & David RL). University of California Press. 33-69.
- Parkhaev PY. 2017. Origin and the early evolution of the phylum Mollusca. *Paleontol J* 51:663-686. 10.1134/s003103011706003x
- Penalzoza C, Gutierrez AP, Eory L, Wang S, Guo X, Archibald AL, Bean TP, and Houston RD. 2021. A chromosome-level genome assembly for the Pacific oyster *Crassostrea gigas*. *Gigascience* 10: giab020. 10.1093/gigascience/giab020
- Peng C, Huang Y, Bian C, Li J, Liu J, Zhang K, You X, Lin Z, He Y, Chen J, Lv Y, Ruan Z, Zhang X, Yi Y, Li Y, Lin X, Gu R, Xu J, Yang J, Fan C, Yao G, Chen JS, Jiang H, Gao B, and Shi Q. 2021. The first *Conus* genome assembly reveals a primary genetic central dogma of conopeptides in *C. betulinus*. *Cell Discov* 7:11. 10.1038/s41421-021-00244-7
- Pereira RB, Andrade PB, Valentão P. 2016. Chemical diversity and biological properties of secondary metabolites from sea hares of *Aplysia* Genus *Mar Drugs* 14:39. 10.3390/md14020039.
- Perry KJ, and Henry JQ. 2015. CRISPR/Cas9-mediated genome modification in the mollusc, *Crepidula fornicata*. *Genesis* 53:237-244. 10.1002/dvg.22843
- Pierce MP. 2019. Filling in the Gaps: Adopting Ultraconserved Elements Alongside COI to Strengthen Metabarcoding Studies. *Front Ecol Evol* 7:469. 10.3389/fevo.2019.00469
- Ponder WF, Lindberg DR, and Ponder JM. 2020. Gastropoda I: Introduction and the Stem Groups. *Biology and Evolution of the Mollusca (1st ed)*. CRC Press, 76.
- Ruiz-Pesini E, Mishmar D, Brandon M, Procaccio V & Wallace DC. 2004. Effects of purifying and adaptive selection on regional variation in human mtDNA. *Science* 303:223-226. 10.1126/science.1088434.
- Saadi AJ, Wade CM. 2019. Resolving the basal divisions in the stylommatophoran land snails and slugs with special emphasis on the position of the Scolodontidae. *Mol Phylogenet Evol* 139:106529. 10.1016/j.ympev.2019.106529.
- Sanger F, Nicklen S, and Coulson A. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 74:5463-5467. 10.1073/pnas.74.12.5463
- Schatte Olivier A, Jones L, Vay LL, Christie M, Wilson J, and Malham SK. 2018. A global review of the ecosystem services provided by bivalve aquaculture. *Rev Aquac* 12:3-25. 10.1111/raq.12301
- Schrödl M, and Stöger I. 2014. A review on deep molluscan phylogeny: old markers, integrative approaches, persistent problems. *J Nat Hist* 48:2773-2804. 10.1080/00222933.2014.963184
- Schwartz DC, Li X, Hernandez LI, Ramnarain SP, Huff EJ, and Wang Y. 1993. Ordered restriction maps of *Saccharomyces cerevisiae* chromosomes constructed by optical mapping. *Science* 262:110-114. 10.1126/science.8211116
- Scientists GCo, Bracken-Grissom H, Collins AG, Collins T, Crandall K, Distel D, Dunn C, Giribet G, Haddock S, Knowlton N, Martindale M, Medina M, Messing C, O'Brien SJ, Paulay G, Putnam N, Ravasi T, Rouse GW, Ryan JF, Schulze A, Worheide G, Adamska M, Bailly X, Breinholt J, Browne WE, Diaz MC, Evans N, Flot JF, Fogarty N, Johnston M, Kamel B, Kawahara AY, Laberge T, Lavrov D, Michonneau F, Moroz LL, Oakley T, Osborne K, Pomponi SA, Rhodes A, Santos SR, Satoh N, Thacker RW, Van de Peer Y, Woolstra CR, Welch DM, Winston J, and Zhou X. 2014. The Global Invertebrate Genomics Alliance (GIGA): developing community

- resources to study diverse invertebrate genomes. *J Hered* 105:1-18. 10.1093/jhered/est084
- Serb JM, and Eernisse DJ. 2008. Charting evolution's trajectory: Using molluscan eye diversity to understand parallel and convergent evolution. *Evol: Educ Outreach* 1:439-447. 10.1007/s12052-008-0084-1
- Sereika M, Kirkegaard RH, Karst SM, Michaelsen TY, Sorensen EA, Wollenberg RD, and Albertsen M. 2022. Oxford Nanopore R10.4 long-read sequencing enables the generation of near-finished bacterial genomes from pure cultures and metagenomes without short-read or reference polishing. *Nat Methods* 19:823-826. 10.1038/s41592-022-01539-7
- Shendure J, Balasubramanian S, Church GM, Gilbert W, Rogers J, Schloss JA, and Waterston RH. 2017. DNA sequencing at 40: past, present and future. *Nature* 550:345-353. 10.1038/nature24286
- Shtolz N and Mishmar D. 2019. The mitochondrial genome on selective constraints and signatures at the organism, cell, and single mitochondrion levels. *Front Ecol Evol* 7:342. 10.3389/fevo.2019.00342.
- Sigwart JD, and Lindberg DR. 2015. Consensus and confusion in molluscan trees: evaluating morphological and molecular phylogenies. *Syst Biol* 64:384-395. 10.1093/sysbio/syu105
- Sigwart JD, Lindberg DR, Chen C, and Sun J. 2021. Molluscan phylogenomics requires strategically selected genomes. *Phil Trans R Soc B* 376:20200161. 10.1098/rstb.2020.0161
- Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, and Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210-3212. 10.1093/bioinformatics/btv351
- Sin TS. 2003. Damage potential of the golden apple snail *Pomacea canaliculata*(Lamarek) in irrigated rice and its control by cultural approaches. *Int J Pest Manag* 49:49-55. 10.1080/713867835
- Slater BJ, and Bohlin MS. 2022. Animal origins: The record from organic microfossils. *Earth Sci Rev* 232:104107. 10.1016/j.earscirev.2022.104107
- Smith SA, Wilson NG, Goetz FE, Feehery C, Andrade SC, Rouse GW, Giribet G, Dunn CW. 2011. Resolving the evolutionary relationships of molluscs with phylogenomic tools. *Nature* 480:364-367. 10.1038/nature10526
- Stöger I, and Schrödl M. 2013. Mitogenomics does not resolve deep molluscan relationships (yet?). *Mol Phylogenet Evol* 69:376-392. 10.1016/j.ympev.2012.11.017
- Stöger I, Sigwart JD, Kano Y, Knebelberger T, Marshall BA, Schwabe E, and Schrödl M. 2013. The continuing debate on deep molluscan phylogeny: evidence for Serialia (Mollusca, Monoplacophora + Polyplacophora). *Biomed Res Int* 2013:407072. 10.1155/2013/407072
- Stoler N, and Nekrutenko A. 2021. Sequencing error profiles of Illumina sequencing instruments. *NAR Genom Bioinform* 3:lqab019. 10.1093/nargab/lqab019
- Sun J, Chen C, Miyamoto N, Li R, Sigwart JD, Xu T, Sun Y, Wong WC, Ip JCH, Zhang W, Lan Y, Bissessur D, Watsuji TO, Watanabe HK, Takaki Y, Ikeo K, Fujii N, Yoshitake K, Qiu JW, Takai K, and Qian PY. 2020. The Scaly-foot Snail genome and implications for the origins of biomineralised armour. *Nat Commun* 11:1657. 10.1038/s41467-020-15522-3
- Sun J, Mu H, Ip JCH, Li R, Xu T, Accorsi A, Alvarado AS, Ross E, Lan Y, Sun Y, Castro-Vazquez A, Vega IA, Heras H, Ituarte S, Bocxlaer BV, Hayes KA, Cowie RH, Zhao Z, Zhang Y, Qian PY, and Qiu JW. 2019. Signatures of divergence, invasiveness, and terrestrialization revealed by four apple snail genomes. *Mol Biol Evol* 36:1507-1520. 10.1093/molbev/msz084
- Tanner AR, Fuchs D, Winkelmann IE, Gilbert M, Thomas P, Pankey MS, Ribeiro ÂM, Kocot KM, Halanych KM, Oakley TH, da Fonseca RR, Pisani D, and Vinther J. 2017. Molecular clocks

- indicate turnover and diversification of modern coleoid cephalopods during the Mesozoic Marine Revolution. *Proc R Soc B* 284:20162818. 10.1098/rspb.2016.2818
- Thomas-Bulle C, Piednoel M, Donnart T, Filee J, Jollivet D, and Bonnivard E. 2018. Mollusc genomes reveal variability in patterns of LTR-retrotransposons dynamics. *BMC Genomics* 19:821. 10.1186/s12864-018-5200-1
- Uliano-Silva M, Dondero F, Dan Otto T, Costa I, Lima NCB, Americo JA, Mazzoni CJ, Prosdocimi F, and Rebelo MF. 2018. A hybrid-hierarchical genome assembly strategy to sequence the invasive golden mussel, *Limnoperna fortunei*. *Gigascience* 7: gix128. 10.1093/gigascience/gix128
- Uribe JE, González VL, Irisarri I, Kano Y, Herbert DG, Strong EE, and Harasewych MG. 2022. A phylogenomic backbone for gastropod molluscs. *Syst Biol*: syac045. 10.1093/sysbio/syac045
- Uribe JE, Irisarri I, Templado J, and Zardoya R. 2019. New patellogastropod mitogenomes help counteracting long-branch attraction in the deep phylogeny of gastropod mollusks. *Mol Phylogenet Evol* 133:12-23. 10.1016/j.ympev.2018.12.019
- Van Bocxlaer B, Ortiz-Sepulveda CM, Gurdebeke PR, and Vekemans X. 2020. Adaptive divergence in shell morphology in an ongoing gastropod radiation from Lake Malawi. *BMC Evol Biol* 20:5. 10.1186/s12862-019-1570-5
- Varney RM, Brenzinger B, Malaquias MAE, Meyer CP, Schrod M, and Kocot KM. 2021a. Assessment of mitochondrial genomes for heterobranch gastropod phylogenetics. *BMC Ecol Evol* 21:6. 10.1186/s12862-020-01728-y
- Varney RM, Speiser DI, McDougall C, Degnan BM, and Kocot KM. 2021b. The iron-responsive genome of the chiton *Acanthopleura granulata*. *Genome Biol Evol* 13. 10.1093/gbe/evaa263
- Villanueva R, Perricone V, and Fiorito G. 2017. Cephalopods as predators: a short journey among behavioral flexibilities, adaptations, and feeding habits. *Front Physiol* 8:598. 10.3389/fphys.2017.00598
- Vinther J. 2015. The origins of molluscs. *Palaeontology* 58:19-34. 10.1111/pala.12140
- Vinther J, Jell P, Kampouris G, Carney R, Racicot RA, and Briggs DEG. 2012. The origin of multiplacophorans - convergent evolution in Aculiferan molluscs. *Palaeontology* 55:1007-1019. 10.1111/j.1475-4983.2012.01180.x
- Wanninger A, and Wollesen T. 2019. The evolution of molluscs. *Biol Rev Camb Philos Soc*. 10.1111/brv.12439
- Wägele H, Kolb AK, Vonnemann V, and Medina M. 2008. Heterobranchia I: The Opisthobranchia. *Phylogeny and Evolution of the Mollusca* (ed. Winston P). California Scholarship Online: 384-408. 10.1525/california/9780520250925.003.0014
- Wenger AM, Peluso P, Rowell WJ, Chang PC, Hall RJ, Concepcion GT, Ebler J, Fungtammasan A, Kolesnikov A, Olson ND, Topfer A, Alonge M, Mahmoud M, Qian Y, Chin CS, Phillippy AM, Schatz MC, Myers G, DePristo MA, Ruan J, Marschall T, Sedlazeck FJ, Zook JM, Li H, Koren S, Carroll A, Rank DR, and Hunkapiller MW. 2019. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat Biotechnol* 37:1155-1162. 10.1038/s41587-019-0217-9
- Werner GD, Gemmell P, Grosser S, Hamer R, and Shimeld SM. 2013. Analysis of a deep transcriptome from the mantle tissue of *Patella vulgata* Linnaeus (Mollusca: Gastropoda: Patellidae) reveals candidate biomineralising genes. *Mar Biotechnol (NY)* 15:230-243. 10.1007/s10126-012-9481-0
- Wu F, Mueller LA, Crouzillat D, Petiard V, and Tanksley SD. 2006. Combining bioinformatics and



- phylogenetics to identify large sets of single-copy orthologous genes (COSII) for comparative, evolutionary and systematic studies: a test case in the euasterid plant clade. *Genetics* 174:1407-1420. 10.1534/genetics.106.062455
- Wu R, and Kaiser A. 1968. Structure and base sequence in the cohesive ends of bacteriophage lambda DNA. *J Mol Biol* 35:523-537. 10.1016/s0022-2836(68)80012-9
- Xin Y, Ren J, and Liu X. 2011. Mitogenome of the small abalone *Haliotis diversicolor* Reeve and phylogenetic analysis within Gastropoda. *Mar Genomics* 4:253-262. 10.1016/j.margen.2011.06.005
- Yang JL, Feng DD, Liu J, Xu JK, Chen K, Li YF, Zhu YT, Liang X, and Lu Y. 2021. Chromosome-level genome assembly of the hard-shelled mussel *Mytilus coruscus*, a widely distributed species from the temperate areas of East Asia. *Gigascience* 10. 10.1093/gigascience/giab024
- Yoshida MA, Imoto J, Kawai Y, Funahashi S, Minei R, Akizuki Y, Ogura A, Nakabayashi K, Yura K, and Ikeo K. 2020. Genomic and transcriptomic analyses of bioluminescence genes in the enope squid *Watasenia scintillans*. *Mar Biotechnol (NY)* 22:760-771. 10.1007/s10126-020-10001-8
- Young ND, Stroehlein AJ, Wang T, Korhonen PK, Mentink-Kane M, Stothard JR, Rollinson D, and Gasser RB. 2022. Nuclear genome of *Bulinus truncatus*, an intermediate host of the carcinogenic human blood fluke *Schistosoma haematobium*. *Nat Commun* 13:977. 10.1038/s41467-022-28634-9
- Zapata F, Wilson NG, Howison M, Andrade SC, Jörger KM, Schrödl M, Goetz FE, Giribet G, and Dunn CW. 2014. Phylogenomic analyses of deep gastropod relationships reject Orthogastropoda. *Proc Biol Sci* 281:20141739. 10.1098/rspb.2014.1739
- Zardoya R. 2020. Recent advances in understanding mitochondrial genome diversity. *F1000Res* 9: F1000 Faculty Rev-270. 10.12688/f1000research.21490.1
- Zascavage RR, Hall CL, Thorson K, Mahmoud M, Sedlazeck FJ, and Planz JV. 2019. Approaches to whole mitochondrial genome sequencing on the Oxford Nanopore MinION. *Curr Protoc Hum Genet* 104:e94. 10.1002/cphg.94
- Zemanova MA, Knop E, and Heckel G. 2017. Introgressive replacement of natives by invading *Arion* pest slugs. *Sci Rep* 7:14908. 10.1038/s41598-017-14619-y
- Zer V. 2009. Evolution of shell loss in Opisthobranch gastropods: sea hares (Opisthobranchia, Anaspidia) as a model system. *UC Merced*: ProQuest ID: 2009\_zvue.
- Zhang Y, Mao F, Mu H, Huang M, Bao Y, Wang L, Wong NK, Xiao S, Dai H, Xiang Z, Ma M, Xiong Y, Zhang Z, Zhang L, Song X, Wang F, Mu X, Li J, Ma H, Zhang Y, Zheng H, Simakov O, and Yu Z. 2021. The genome of *Nautilus pompilius* illuminates eye evolution and biomineralization. *Nat Ecol Evol* 5:927-938. 10.1038/s41559-021-01448-6

## 7. Additional file: Statistics for Mollusca genomes.

B: Bivalvia, C: Cephalopoda. G: Gastropoda. P: Polyplacophora. Chromosome-level assemblies are shown behind the column of Class, with the number of assembled chromosomes shown in parentheses. BUSCOs: Bench marking Universal Single-Copy Orthologs. The column of BUSCOs shows the percentage of: C: Complete BUSCOs; S: Complete Single-Copy BUSCOs, D: Complete Duplicated BUSCOs; F: Fragmented BUSCOs; M: Missing BUSCOs; n: Total BUSCO groups searched. Citations for unpublished work are filled in by the genome submitter and the year in NCBI. Statistics date until August 2022.

Class	Species	Assembly size (Mb)	Heterozygosity	Repeat content	Contig N50	Scaffold N50	BUSCOs	Reference
B (10)	<i>Crassostrea ariakensis</i>	662.90	1.30%	54.24%	5,932,265	66,336,742	C:97.1%[S:96.2%;D:0.9%], F:0.6%,M:2.3%,n:5295	(Wu et al., 2021)
B (10)	<i>Crassostrea gigas</i>	647.89	3.20%	43%	1,813,842	58,462,999	C:95.6%[S:94%;D:1.6%], n:978	(Penaloza et al., 2021)
B (10)	<i>Crassostrea hongkongensis</i>	608.62	1.08%	45%	2,576,225	55,855,599	C:95.8%, F:0.8%,M:3.4%,n:978	(Peng et al., 2020)
B (10)	<i>Ostrea edulis</i>	1,035.80	1.02%	49.8%	95,771,753	95,771,753	C:91.8%[S:90.5%;D:1.3%] F:1.6%,M:6.6%, n:954	(Gundappa et al., 2022)
B (14)	<i>Mytilus coruscus</i>	1,903.83	1.39%	47.40%	817,337	898,347	C:89.4% [S:88.1%;D:1.3%], F:1.9%,M:8.7%,n:978	(Yang et al., 2021)
B (14)	<i>Mytilus edulis</i>	1,903.83	1.64%	36.35%	664,188	898,347	C:91.1%[S:77.5%;D:13.6%], F:5.35%, M:3.56%,n:954	(Li et al., 2020)
B (14)	<i>Pinctada fucata martensii</i>	990.98	1.30%	48.50%	21,518	59,032,463	C:81.7%[S:72.5%;D:9.2%], F:7.8%,M:10.5%,n:978	(Du, Fan, et al., 2017)
B (15)	<i>Limnoperna fortunei</i>	1,335.58	-	-	1,498,882	97,051,362	-	Wellcome Sanger Institute, 2022

B (16)	<i>Dreissena polymorpha</i>	1,798.01	2.13%	47.43%	1,111,654	117,515,028	C:92.3%[S:88.5%,D:3.8%], n:978	(McCartney et al., 2021)
B (18)	<i>Mercenaria mercenaria</i>	1,777.63	1.34%	49.11%	1,779,571	91,379,220	C:90.5% [S:88.4%,D:2.1%], F:0.9%, M:8.6%, n:954	(Song et al., 2021)
B (18)	<i>Tridacna crocea</i>	1,048.99	-	-	32,149,973	60,802,674	-	Wellcome Sanger Institute, 2022
B (19)	<i>Anadara kagoshimensis</i>	1,115.24	-	-	1,935,399	60,635,260	-	Liaoning Ocean and Fisheries Science Research Institute, 2021
B (19)	<i>Archivesica marissinica</i>	1,545.00	0.41%	42%	79,144	74,312,544	C:91.8%[S:90.1%,D:1.7%], F:1.7%,M:6.5%,n:978	(Ip et al., 2021)
B (19)	<i>Crassostrea virginica</i>	684.74	-	-	1,971,208	75,944,018	C:94.5%, F:1.2%,M:4.3%,n:978	McDonnell Genome Institute - Washington University School of Medicine, 2017
B (19)	<i>Gari tellinella</i>	1,597.63	-	-	19,181,371	85,279,272	-	Wellcome Sanger Institute, 2021
B (19)	<i>Mercenaria mercenaria</i>	1,858.20	-	45%	82,432,825	82,914,371	C:91.8%[S:86.2%,D:5.6%], F:4.0%,M:4.2%,n:954	(Farhat et al.,2022)
B (19)	<i>Pecten maximus</i>	918.31	1.71%	26.95%	1,258,799	44,824,366	C:94.5%[S:91.2%;D:3.3%], F:1.0%,M:4.5%,n:978	(Kenny et al., 2020)
B (19)	<i>Ruditapes philippinarum</i>	1,123.16	1.04%	38.30%	29,238	345,005	C:92.2%[S:90.3%;D:1.9%], F:1.6%,M:6.2%,n:978	(Yan et al., 2019)
B (19)	<i>Sinonovacula constricta</i>	1,220.85	1.53%	53.12%	976,936	65,929,677	C:88.8%[S:85.1%;D:3.7%], F:4.0%,M:7.2%,n:978	(Dong et al., 2020)
B (19)	<i>Solen grandis</i>	1,324.49	-	-	50,000	67,678,117	-	Quanzhou Normal University, 2021
B	<i>Anadara broughtonii</i>	884.57	-	46.10%	1,797,717	44,995,656	C:91.7%[S:87.2,D:4.5], F:1.1%,M:7.2%,n:978	(Bai et al., 2019)

B	<i>Argopecten irradians concentricus</i>	874.78	-	47.19%	63,725	1,246,717	C: 91% [S:87.1;D:3.9%], F:5.5%, M:3.5%, n:843	(Liu et al., 2020)
B	<i>Argopecten irradians irradians</i>	835.60	0.90%	46.63%	78,654	1,533,165	C:91% [S:86.7;D:4.3%], F:6.6%, M:2.4%, n:843	(Du, Song, et al., 2017)
B	<i>Bathymodiolus platifrons</i>	1,658.19	1.24%	47.90%	12,602	343,341	C:84.67%[S:83%,D:1.7%], F:10.4%,M:4.9%,n:978	(Sun et al., 2017)
B	<i>Corbicula fluminea</i>	1,520	2.41%	69.66%	521,060	70,620,000	C:86.7%[S:73%,D:13.6%], F:1.49%,M:11.86%,n:5295	(Zhang et al., 2021)
B	<i>Crassostrea gigas</i>	558.60	0.73%	36.10%	19,387	401,319	C:84.6%[S:81.1%;D:3.5%], F:5.7%,M:9.7%,n:978	(Zhang et al., 2012)
B	<i>Cyclina sinensis</i>	903.12	1.53%	43.14%	2,587,078	46,470,132	C:92.7%[S:91.6%,D:1.1%], F:1.3%,M:6.0%,n:978	(Wei et al., 2020)
B	<i>Dreissena rostriformis</i>	1,241.70	2.40%	31.88%	45,905	131,390	C:83.23%[S:80.2%,D:3.1%], F:11.66%,M:5.11%,n:978	(Calcino et al., 2019)
B	<i>Limnoperna fortunei</i>	1,673.22	2.30%	33.40%	32,203	309,123	C:81.9%[S:78.6%,D:3.3%], F:7.4%,M:10.7%,n:978	(Uliano-Silva et al., 2018)
B	<i>Lutraria rhynchaena</i>	543.90	0.9%-1.6%	29.40%	2,143,760	-	C:95.8%[S:94.3%,D:1.5%] F:0.8%,M:3.4%,n:978	(Thai et al., 2019)
B	<i>Magallana hongkongensis</i>	757.93	-	41.12%	49,472	72,332,161	C:94.6%,n:978	(Li et al., 2020)
B	<i>Margaritifera margaritifera</i>	2,472.07	0.127– 0.105%	59.07%	16,891	288,726	C:84.9%[S:83.8%,D:1.1%], F: 4.9%, M:10.2%,n:954	(Gomes-Dos-Santos et al., 2021)
B	<i>Megalonaias nervosa</i>	2,365.22	0.78%	25.00%	50,186	50,649	C:83%[S:80.9%,D:2.1%], F:9%, M:8%, n:978	(Rogers et al., 2021)
B	<i>Mizuhopecten yessoensis</i>	987.69	1.04%	38.87%	6,859	803,631	C:88.9%[S:88.2%;D:0.7%], F:1.5%,M:9.6%,n:5295	(Wang et al., 2017)

B	<i>Modiolus philippinarum</i>	2,629.65	2.02%	62.00%	19,700	100,161	C:70.3%[S:67.0%;D:3.3%], F:17.3%,M:12.4%,n:978	(Sun et al., 2019) (Calcino et al., 2019)
B	<i>Mya arenaria</i>	1,324.96	4.6%-6.6%	35.00%	10,552	14,639	C:76%[S:63%;D:13%], F:7%,M:17%,n:978	(Plachetzki, Pankey, MacManes, Lesser, & Walker, 2020)
B	<i>Mytilus californianus</i>	1,651.97	-	-	16,323,199	117,871,512	-	UCLA, 2022
B	<i>Mytilus galloprovincialis</i>	1,282.21	1.73 ± 0.24%	43.00%	77,157	207,642	C:69%[S:56%;D:13%], F:8.5%, M:22.5%,n:843	(Gerdol et al., 2020)
B	<i>Ostrea lurida</i>	1,140.79	-	-	7,815	12,947	-	University of Washington School of Aquatic and Fishery Sciences, 2020
B	<i>Panopea generosa</i>	942.35	-	-	14,495	57,743,597	C:66.70%	University of Washington School of Aquatic and Fishery Sciences, 2020
B	<i>Perna viridis</i>	731.87	0.63%	19.57%	3,015	4,106,954	C:99.4%[S:98.3%;D:1.1%], F:0.4%,M:0.2%,n:978	(Inoue et al., 2021)
B	<i>Pinctada fucata</i>	815.30	-	43.44%	1629	167,048	C:91.10%	(Takeuchi et al., 2016)
B	<i>Pinna nobilis</i>	586.48	1.00%	36.20%	5,822	7,576	C:27%, F:13%, M: 60%,n:978	(Bunet et al., 2021)
B	<i>Potamilus streckersoni</i>	1,776.76	0.57%	51.03%	2,032,685	2,051,244	C: 94.6%[S:93.7%;D:0.9%], F:1.2%,M:4.2%,n:978	(Smith, 2021)
B	<i>Saccostrea glomerata</i>	788.12	0.51%	45.03%	39,800	804,232	C:79.7%, F:13.5%,M:6.6%,n:843	(Calcino et al., 2019)
B	<i>Saxidomus purpurata</i>	1,161.00			511,514	52,225,674		Genome Research Foundation,2022
B	<i>Tegillarca granosa</i>	797.65	1.17%	53.75%	605,873	42,616,908	C:93.3% [S:88.7%;D:4.6%], F:1.3%,M:5.4%,n:978	(Bao et al., 2021)

B	<i>Venustaconcha ellipsiformis</i>	1,590.01	0.60%	37.81%	2813	6657	C:67.9%[S:66.7%;D:1.2%], F:21.2%,M:10.9%,n:978	(Renaut et al., 2018)
C (30)	<i>Octopus sinensis</i>	2,719.15	0.34%	42.26%	490,217	105,892,736	C:80.2%[S:72.3%;D:7.9%] F:9.9%,M:9.9%,n:303	(Li et al., 2020)
C (46)	<i>Doryteuthis (Loligo) pealeii</i>	4,598.00	1.2%	62.00%	230,938	107,400,528	-	(Albertin et al, 2022) <sup>3</sup>
C (48)	<i>Euprymna scolopes</i>	5,116.32	-	-	3,700,000	120,300,000	-	(Albertin et al, 2022)
C	<i>Architeuthis dux</i>	3,155.39	-	49.10%	9000	5,478,336	C:88.5%[S:87.6%;D:0.9%], F:3.6%,M:7.9%,n:978	(da Fonseca et al., 2020)
C	<i>Euprymna scolopes</i>	5,280.01	-	50.00%	3558	3,549,550	C:72.94%,n:978	(Belcaid et al., 2019)
C	<i>Hapalochlaena maculosa</i>	4,009.60	-	-	1,130	931,835	-	(Greve et al., 2017)
C	<i>Nautilus pompilius</i>	729.02	-	25.26%	1,094,646	-	C:91.3%[S:89.3%;D:2.0%], F:1.02%,M:7.68%,n:978	(Zhang et al., 2021)
C	<i>Octopus bimaculoides</i>	2,338.19	0.08%	45.00%	5532	475,182	C:86.50%	(Albertin et al., 2015)
C	<i>Octopus minor</i>	5,090.35	-	44.43%	41,584	196,941	C:76.2%[S:64.2%;D:12%], F:8.4%,M:15.4%,n:978	(Kim et al., 2018)
C	<i>Octopus vulgaris</i>	1,772.96	1.10%	50.00%	3,040	265,914	C:51.6%, F:9%, M:39.4%,n:954	(Zarrella et al., 2019)
C	<i>Sepia pharaonis</i>	4,785.53	0.35%	77.30%	1,926,397	-	C:79.5%[S:73.6%;D:5.9%], F:10.2%,M:10.3%,n:954	(Song et al., 2021)
C	<i>Watasenia scintillans</i>	649.18	4.9–5.9%	19.20%	1,283	-	C:58.9%	(Yoshida et al., 2020)
G (9)	<i>Patella pellucida</i>	750.48	2.14%	49.65%	9,229,741	-	C:96.4%[S:93.3%;D:3.1%], F:1.1%,M:2.5%,n:978	Wellcome Sanger Institute, 2021
G (9)	<i>Patella vulgata</i>	694.48	-	-	18,376,887	87,764,759	-	Wellcome snager Institute, 2022

G (14)	<i>Pomacea canaliculata</i>	447.67	1.41%	20.53%	81,153	32,644,854	C:96.4%[S:95.1%,D:1.3%], F:0.7%,M:2.9%,n:978	(Sun et al., 2019)
G (15)	<i>Chrysomallon squamiferum</i>	404.61	1.38%	25.00%	1,883,489	30,197,426	C:96.6%,n:978	(Sun et al., 2020)
G (15)	<i>Gigantopelta aegis</i>	1,149.61	0.50%	51.00%	461,769	81,591,406	C:94.0%,n:954	(Lan et al., 2021)
G (17)	<i>Aplysia californica</i>	927.64	-	-	9,329	52,647,889	-	(Di Palma et al., 2020)
G (18)	<i>Gibbula magus</i>	1,470.40	-	-	3,416,053	80,454,948	-	Wellcome sanger institute, 2022
G (18)	<i>Phorcus lineatus</i>	996.33	1.06%	29.38%	4,947,473	-	C:85.4%[S:84.6%,D:0.8%], F:4.5%,M:10.1%,n:5295	Wellcome Sanger Institute, 2021
G (18)	<i>Steromphala cineraria</i>	1,462.73	3.55%	55.73%	4,992,915	-	C:85.5%[S:78.3%,D:7.2%], F:5.1%,M:9.4%,n:5295	Wellcome Sanger Institute, 2021
G (26)	<i>Arion vulgaris</i>	1,541.03	1.55%	75.09%	8,603,329	64,342,731	C:90.6%[S:85.0%,D:5.6%], F:1.9%,M:7.5%,n:978	(Chen et al, 2022)
G (31)	<i>Lissachatina (Achatina) fulica</i>	1,855.89	0.47%	71.00%	721,038	59,589,303	C:91.5%[S:84.6%,D:6.9%], F:2.5%,M:6.0%,n:978	(Guo et al., 2019)
G (31)	<i>Lissachatina (Achatina) immaculata</i>	1,653.15	0.24%	57.70%	3,802,429	56,367,627	C:92.0%[S:85.5%,D:6.5%], F:1.2%,M:6.8%,n:978	(Liu et al., 2020)
G (35)	<i>Conus ventricosus</i>	3,591.51	1.05–1.08%	53.36%	114,325	93,519,712	C:84.9%[S:82%,D:2.9%], F:4.3%,M:10.8%,n:978	(Pardos-Blas et al., 2021)
G	<i>Alviniconcha marisindica</i>	829.61	-	-	727,552	-	-	HKUST, 2021
G	<i>Ampullaceana balthica</i>	1,105.89	-	-	435,188	816,280	-	Senckenber Biodiversity and Climate Research Centre Frankfurt 2022
G	<i>Anentome helena</i>	1,720.19	-	-	56,088	2,075,175	-	Iruuion Genomes, 2020

G	<i>Aplysia californica</i>	927.31	0.33%	30.00%	9,586	917,541	C:92.4%[S:91.9%,D:0.5%], F:1.9%,M:5.7%,n:978	Broad Institute, 2013
G	<i>Babylonia areolata</i>	1,108.40	-	-	5,933	6,321	-	Fisheries and Technical, Economic College, 2020
G	<i>Batillaria attramentaria</i>	717.57	-	-	1,290,776	-	-	Ewha Womans University, 2021
G	<i>Biomphalaria glabrata</i>	916.38	0.22-0.69%	44.80%	7,298	48,059	C:88.4%[S:86.5%,D:1.9%], F:4.9%,M:6.7%,n:978	(Adema et al., 2017) (Wethington et al., 2007)
G	<i>Biomphalaria glabrata</i>	852.54	-	-	7,313,524	15,088,501	-	Oregon State University, 2022
G	<i>Biomphalaria straminea</i>	1,004.75	-	40.68%	9,530	25,272,813	C:87.0%	(Nong et al., 2022)
G	<i>Bulinus truncatus</i>	1,221.78	1.36%	51.03%	234,305	4,956,851	C:95.8%[S:86.2%,D:9.6%], F:0.8%,M:3.4%,n:954	(Young et al., 2022)
G	<i>Candidula unifasciata</i>	1,286.46	1.09%	61.10%	246,413	246,413	C:92.4%[S:85.3%,D:7.1%], F:1.6%,M:6.0%,n:978	(Chueca, Schell, & Pfenninger, 2021)
G	<i>Cepaea nemoralis</i>	3,490.92	1.43%	76.40%	330,079	333,110	C:87.2%[S:74.3%,D:12.9%], F:3.8%,M:9.0%,n:954	(Saenko, Groenenberg, Davison, & Schilthuizen, 2021)
G	<i>Colubraria reticulata</i>	67.10	-	-	890	-	-	University of Konstanz, 2016
G	<i>Conus betulinus</i>	3,430.83	-	38.56%	171,480	232,607	C:89.8%[S:78%;D:11.8%], F:3.2%,M:7.0%,n:978	(Peng et al., 2021)
G	<i>Conus consors</i>	2,049.32	-	-	749	1,128	-	(Andreson et al., 2019)
G	<i>Conus tribblei</i>	2,160.49	-	-	854	2,681	-	(Barghi, Concepcion, Olivera, & Lluisma, 2016)
G	<i>Dracogyra subfuscus</i>	1,160.00	0.50%	50.00%	4,083	5,907	C:97.1%[S:96.2%,D:0.9%], F:0.6%,M:2.3%,n:5295	(Lan et al., 2021)
G	<i>Elysia chlorotica</i>	557.48	3.66%	32.60%	30,474	441,954	C:94.7%[S:93.3,D:1.4], F:2.5%,M:2.8%,n:978	(Cai et al., 2019)



G	<i>Elysia marginata</i>	790.32	-	29.00%	6,205	225,654	C:91.1%[S:89.6%,D:1.5%], F:6.7%,M:2.2%,n:978	(Maeda et al., 2021)
G	<i>Haliotis cracherodii</i>	1,182.25	1.56%	-	17,462,865	60,096,789	C:97.4 % [ S:97.2%,D:0.2%]	(Orland et al., 2022)
G	<i>Haliotis discus</i>	1,865.48	-	30.76%	41,000	200,099	F:1.9%,M:0.7%, n:954 C:73.8%[S:68.4%,D:5.4], F:14.6%,M:11.6%,n:891	(Nam et al., 2017)
G	<i>Haliotis laevigata</i>	1,762.66	0.68%	-	3,353	81,233	C:86.8%[S:84.6%,D:2.2%], F:8.7%,M:4.5%,n:978	(Botwright et al., 2019)
G	<i>Haliotis rubra</i>	1,378.27	1.27-1.44%	-	1,177,711	1,227,833	C:94.6%[S:91.6%,D:3%], F:1.6%,M:3.8%,n:978	(Gan et al., 2019)
G	<i>Haliotis rufescens</i>	1,779.96	0.95%	33.06%	283,651	1,895,871	C:95.1%[S:88.7%,D:6.4%], F:1%,M:3.9%,n:978	(Masonbrink et al., 2019)
G	<i>Haliotis rufescens</i>	1,334.45	-	-	8,868,657	45,695,856	-	UCLA,2022
G	<i>Lanistes nyassanus</i>	509.78	0.60%	28.87%	25,785	317,839	C:95.0%[S:93.5%,D:1.5%], F:1.2%,M:3.8%,n:978	(Sun et al., 2019)
G	<i>Limacina bulimoides</i>	2,901.93	-	-	893	-	C:30.3%[S:26.8%,D:3.5%], F:29.9%,M:39.8%,n:978	(Choo et al., 2020)
G	<i>Lottia gigantea</i>	359.51	-	22.25%	96,027	1,870,055	C:96.0%[S:94.9%,D:1.1%], F:1.63%,M:2.35%,n:978	(Simakov et al., 2013) (Nam et al., 2017)
G	<i>Lymnaea stagnalis</i>	833.23	-	-	5,751	-	-	BANG, 2016
G	<i>Marisa cornuarietis</i>	535.45	0.08%	30.82%	4,359,112	-	C:96.4%[S:94.8%,D:1.6%], F:0.6%,M:3.0%,n:978	(Sun et al., 2019)
G	<i>Oreohelix idahoensis</i>	5,404.39	-	-	394,446	404,192	-	University of Idaho, 2022
G	<i>Phymorhynchus buccinoides</i>	2,114.59	-	-	336,037	-	-	BGI, 2021
G	<i>Physella acuta</i>	764.48	-	-	1,333	1,358	-	(Ebbs, Loker, & Brant, 2018)

G	<i>Plakobranchus ocellatus</i>	927.89	-	-	5,479	1,453,842	C:95.0%[S:93.1%;D:1.9%], F:3.6%,M:1.4%,n:978	(Maeda et al., 2021)
G	<i>Pomacea maculata</i>	432.29	1.22%	21.25%	75,997	375,864	C:96.2%[S:95.0%;D:1.2%], F:0.6%,M:3.2%,n:978	(Sun et al., 2019)
G	<i>Radix auricularia</i>	909.76	0.263%- 0.939%	70.00%	24,354	578,730	C:93.4%, F:1.2%,M:5.4%,n:843	(Schell et al., 2017) (Al-Waaly et al., 2018) et al., 2018)
P	<i>Acanthopleura granulata</i>	606.54	0.65%	23.56%	1,098,986	23,921,462	C:97.4%[S:96.9%;D:0.5%], n:978	(Varney et al., 2021)

## References of additional file

- Adema, C. M. *et al.* Whole genome analysis of a schistosomiasis-transmitting freshwater snail. *Nat. Commun.* 8, 15451 (2017).
- Albertin, C. B. *et al.* Genome and transcriptome mechanisms driving cephalopod evolution. *Nat. Commun.* 13, 2427 (2022).
- Albertin, C. B. *et al.* The octopus genome and the evolution of cephalopod neural and morphological novelties. *Nature* 524, 220-224 (2015).
- Al-Waaly, A.B. *et al.* Genetic diversity in *Radix* species from the middle and south of Iraq based on simple sequence repeats. *Afr. j. biotechnol.* 17, 1119-1128 (2018).
- Andreson, R. *et al.* Gene content of the fish-hunting cone snail *Conus consors*. Preprint at <https://doi.org/10.1101/590695> (2019).
- Bai, C. *et al.* Chromosomal-level assembly of the blood clam, *Scapharca (Anadara) broughtonii*, using long sequence reads and Hi-C. *Gigascience* 8, giz067 (2019)
- Bao, Y. *et al.* Genomic insights into the origin and evolution of molluscan red-bloodedness in the blood clam *Tegillarca granosa*. *Mol. Biol. Evol.* 38, 2351-2365 (2021).
- Barghi, N., Concepcion, G. P., Olivera, B. M., & Lluisma, A. O. Structural features of conopeptide genes inferred from partial sequences of the *Conus tribblei* genome. *Mol. Genet. Genom.* 291, 411-422 (2016).
- Belcaid, M. *et al.* Symbiotic organs shaped by distinct modes of genome evolution in cephalopods. *Proc. Natl. Acad. Sci. U.S.A.* 116, 3030-3035 (2019).
- Botwright, N. A. *et al.* Greenlip abalone (*Haliotis laevigata*) genome and protein analysis provides insights into maturation and spawning. *G3-GENES GENOM GENET* 9, 3067-3078 (2019).
- Bunet, R. *et al.* First insight into the whole genome shotgun sequence of the endangered noble pen shell *Pinna nobilis*: a giant bivalve undergoing a mass mortality event. *J. Molluscan Stud.* 87, eyaa041 (2021).
- Cai, H. *et al.* A draft genome assembly of the solar-powered sea slug *Elysia chlorotica*. *Sci. Data* 6, 190022 (2019).
- Calcino, A. D. *et al.* The quagga mussel genome and the evolution of freshwater tolerance. *DNA Res.* 26, 411-422 (2019).
- Choo, L. Q. *et al.* Novel genomic resources for shelled pteropods: a draft genome and target capture probes for *Limacina bulimoides*, tested for cross-species relevance. *BMC Genomics*, 21, 11 (2020).
- Chueca, L. J., Schell, T., & Pfenninger, M. De novo genome assembly of the land snail *Candidula unifasciata* (Mollusca: Gastropoda). *G3-GENES GENOM GENET* 11 (2021).
- da Fonseca, R. R. *et al.* A draft genome sequence of the elusive giant squid, *Architeuthis dux*. *Gigascience* 9, giz152 (2020).
- Di Palma, F., Alföldi, J., Johnson, J., Berlin, A., Gnerre, S., Jaffe, D., MacCallum, I., Young, S., Walker, B.J., Lindblad-Toh, K. "The Draft Genome of *Aplysia californica*." Unpublished. <https://www.ncbi.nlm.nih.gov/nuccore/AASC00000000.3>.
- Dong, Y. *et al.* The chromosome-level genome assembly and comprehensive transcriptomes of the razor clam (*Sinonovacula constricta*). *Front. Genet.* 11, 664 (2020).
- Du, X. *et al.* The pearl oyster *Pinctada fucata martensii* genome and multi-omic analyses provide insights into biomineralization. *Gigascience* 6, 1-12 (2017).
- Du, X., Song, K., Wang, J., Cong, R., Li, L., & Zhang, G. Draft genome and SNPs associated with carotenoid accumulation in adductor muscles of bay scallop (*Argopecten irradians*). *J. Genom.* 5,

83-90 (2017).

Ebbs, E. T., Loker, E. S., & Brant, S. V. Phylogeography and genetics of the globally invasive snail *Physa acuta* Draparnaud 1805, and its potential to serve as an intermediate host to larval digenetic trematodes. *Evol. Biol.* 18, 103 (2018).

Farhat, S. *et al.* Comparative analysis of the *Mercenaria mercenaria* genome provides insights into the diversity of transposable elements and immune molecules in bivalve mollusks. *BMC Genomics* 23, 192 (2022).

Gan, H. M. *et al.* Best foot forward: nanopore long reads, hybrid meta-assembly, and haplotig purging optimizes the first genome assembly for the southern hemisphere blacklip abalone (*Haliotis rubra*). *Front. Genet.* 10, 889 (2019).

Gerdol, M. *et al.* Massive gene presence-absence variation shapes an open pan-genome in the Mediterranean mussel. *Genome Biol.* 21, 275 (2020).

Gomes-Dos-Santos, A. *et al.* The crown pearl: a draft genome assembly of the European freshwater pearl mussel *Margaritifera margaritifera* (Linnaeus, 1758). *DNA Res.* 28 (2021).

Greve, C. *et al.* Snails in the desert: species diversification of *Theba* (Gastropoda: Helicidae) along the Atlantic coast of NW Africa. *Ecol. Evol.* 7, 5524-5538 (2017).

Gundappa, M. K. *et al.* Chromosome level reference genome for European flat oyster (*Ostrea edulis* L.). *Evol. Appl.* 00, 1-17 (2022).

Guo, Y. *et al.* A chromosomal-level genome assembly for the giant African snail *Achatina fulica*. *Gigascience* 8, giz124 (2019).

Inoue, K. *et al.* Genomics and transcriptomics of the green mussel explain the durability of its byssus. *Sci. Rep.* 11, 5992 (2021).

Ip, J. C. *et al.* Host-endosymbiont genome integration in a deep-sea chemosymbiotic clam. *Mol. Biol. Evol.* 38, 502-518 (2021).

Kenny, N. J. *et al.* The gene-rich genome of the scallop *Pecten maximus*. *Gigascience* 9, giaa037 (2020).

Kim, B. M. *et al.* The genome of common long-arm octopus *Octopus minor*. *Gigascience* 7, giy119 (2018).

Lan, Y. *et al.* Hologenome analysis reveals dual symbiosis in the deep-sea hydrothermal vent snail *Gigantopelta aegis*. *Nat. Commun.* 12, 1165 (2021).

Li, F. *et al.* Chromosome-level genome assembly of the East Asian common octopus (*Octopus sinensis*) using PacBio sequencing and Hi-C technology. *Mol. Ecol. Res.* 20, 1572-1582 (2020).

Li, R. *et al.* The whole-genome sequencing and hybrid assembly of *Mytilus coruscus*. *Front. Genet.* 11, 440 (2020).

Li, Y. *et al.* 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 (2020).

Liu, C. *et al.* Giant African snail genomes provide insights into molluscan whole-genome duplication and aquatic-terrestrial transition. *Mol. Ecol. Res.* 21, 478-494 (2020).

Liu, X. *et al.* Draft genomes of two Atlantic bay scallop subspecies *Argopecten irradians irradians* and *A. i. concentricus*. *Sci. Data* 7, 99 (2020).

Maeda, T. *et al.* Chloroplast acquisition without the gene transfer in kleptoplastic sea slugs, *Plakobranthus ocellatus*. *Elife* 10, e60176 (2021).

Masonbrink, R. E. *et al.* An annotated genome for *Haliotis rufescens* (red abalone) and resequenced

green, pink, pinto, black, and white abalone species. *Genome Biol. Evol.* 11, 431-438 (2019).

McCartney, M. A. *et al.* The Genome of the zebra mussel, *Dreissena polymorpha*: a resource for comparative genomics, invasion genetics, and biocontrol. *G3-GENES GENOM GENET* 12, jkab423 (2021).

Nam, B. H. *et al.* Genome sequence of pacific abalone (*Haliotis discus hannai*): the first draft genome in family Haliotidae. *Gigascience* 6, 1-8 (2017).

Nong, W. *et al.* Genome of the ramshorn snail *Biomphalaria straminea*-an obligate intermediate host of schistosomiasis. *Gigascience* 11, giac012 (2022).

Pardos-Blas, J. R. *et al.* The genome of the venomous snail *Lautoconus ventricosus* sheds light on the origin of conotoxin diversity. *Gigascience* 10, giab037 (2021).

Penaloza, C. *et al.* A chromosome-level genome assembly for the Pacific oyster *Crassostrea gigas*. *Gigascience* 10, giab020 (2021).

Peng, C. *et al.* The first *Conus* genome assembly reveals a primary genetic central dogma of conopeptides in *C. betulinus*. *Cell Discov.* 7, 11 (2021).

Peng, J. *et al.* Chromosome-level analysis of the *Crassostrea hongkongensis* genome reveals extensive duplication of immune-related genes in bivalves. *Mol. Ecol. Res.* 20, 980-994 (2020).

Plachetzki, D. C. *et al.* The genome of the softshell clam *Mya arenaria* and the evolution of apoptosis. *Genome Biol. Evol.* 12, 1681-1693 (2020).

Renaut, S. *et al.* Genome survey of the freshwater mussel *Venustaconcha ellipsiformis* (Bivalvia: Unionida) using a hybrid de novo assembly approach. *Genome Biol. Evol.* 10, 1637-1646 (2018).

Rogers, R. L. *et al.* Gene family amplification facilitates adaptation in freshwater unionid bivalve *Megaloniais nervosa*. *Mol. Ecol.* 30, 1155-1173 (2021).

Saenko, S. V., Groenenberg, D. S. J., Davison, A., & Schilthuisen, M. The draft genome sequence of the grove snail *Cepaea nemoralis*. *G3-GENES GENOM GENET* 11, jkaa071 (2021).

Schell, T. *et al.* An annotated draft genome for *Radix auricularia* (Gastropoda, Mollusca). *Genome Biol. Evol.* 9, 585-592 (2017).

Simakov, O. *et al.* Insights into bilaterian evolution from three spiralian genomes. *Nature* 493, 526-531 (2013).

Smith, C. H. A high-quality reference genome for a parasitic bivalve with doubly uniparental inheritance (Bivalvia: Unionida). *Genome Biol. Evol.* 13, evab029 (2021).

Song, H. *et al.* The hard clam genome reveals massive expansion and diversification of inhibitors of apoptosis in Bivalvia. *BMC Biology* 19, 15 (2021).

Song, W. *et al.* Pharaoh cuttlefish, *Sepia pharaonis*, genome reveals unique reflectin camouflage gene set. *Front. Mar. Sci.* 8 (2021).

Sun, J. *et al.* The Scaly-foot Snail genome and implications for the origins of biomineralised armour. *Nat. Commun.* 11, 1657 (2020).

Sun, J. *et al.* Signatures of divergence, invasiveness, and terrestrialization revealed by four apple snail genomes. *Mol. Biol. Evol.* 36, 1507-1520 (2019).

Sun, J. *et al.* Adaptation to deep-sea chemosynthetic environments as revealed by mussel genomes. *Nat. Ecol. Evol.* 1, 121 (2017).

Takeuchi, T. *et al.* Bivalve-specific gene expansion in the pearl oyster genome: implications of adaptation to a sessile lifestyle. *Zoological Lett.* 2, 3 (2016).

Thai, B. T. *et al.* Whole genome assembly of the snout otter clam, *Lutraria rhynchaena*, using nanopore and illumina data, benchmarked against bivalve genome assemblies. *Front. Genet.* 10,

1158 (2019).

Uliano-Silva, M. *et al.* A hybrid-hierarchical genome assembly strategy to sequence the invasive golden mussel, *Linnoperna fortunei*. *Gigascience* 7, gix128 (2018).

Varney, R. M., Speiser, D. I., McDougall, C., Degnan, B. M. & Kocot, K. M. The Iron-Responsive Genome of the Chiton *Acanthopleura granulata*. *Genome. Biol. Evol.* 13, evaa263 (2021).

Wang, S. *et al.* Scallop genome provides insights into evolution of bilaterian karyotype and development. *Nat. Ecol. Evol.* 1, 120 (2017).

Wei, M. *et al.* Chromosome-level clam genome helps elucidate the molecular basis of adaptation to a buried lifestyle. *iScience* 23, 101148 (2020).

Wethington, A. R. *et al.* Population genetic structure of *Biomphalaria glabrata* in a schistosomiasis-endemic region in Brazil. *J. Molluscan. Stud.* 73, 45-52 (2007).

Wu, B. *et al.* Chromosome-level genome and population genomic analysis provide insights into the evolution and environmental adaptation of Jinjiang oyster *Crassostrea ariakensis*. *Mol. Ecol. Res.* 00, 1-16 (2021).

Yan, X. *et al.* Clam genome sequence clarifies the molecular basis of its benthic adaptation and extraordinary shell color diversity. *iScience* 19, 1225-1237 (2019).

Yang, J. L. *et al.* Chromosome-level genome assembly of the hard-shelled mussel *Mytilus coruscus*, a widely distributed species from the temperate areas of East Asia. *Gigascience* 10, giab024 (2021).

Young, N. D. *et al.* Nuclear genome of *Bulinus truncatus*, an intermediate host of the carcinogenic human blood fluke *Schistosoma haematobium*. *Nat. Commun.* 13, 977 (2022).

Yoshida, M. A. *et al.* Genomic and transcriptomic analyses of bioluminescence genes in the enope squid *Watasenia scintillans*. *Mar. Biotechnol.* 22, 760-771 (2020).

Zarrella, I. *et al.* The survey and reference assisted assembly of the *Octopus vulgaris* genome. *Sci. Data* 6, 13 (2019).

Zhang, G. *et al.* The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* 490, 49-54 (2012).

Zhang, T. *et al.* Dissecting the chromosome-level genome of the Asian Clam (*Corbicula fluminea*). *Sci. Rep.* 11, 15021 (2021).

Zhang, Y. *et al.* The genome of *Nautilus pompilius* illuminates eye evolution and biomineralization. *Nat. Ecol. Evol.* 5, 927-938 (2021).

## 8. Acknowledgements

I would like to express my most sincere gratitude to my supervisor Prof. Dr. Michael Schrödl. I still remember that four years ago, after several emails and the interview about this program, Michi's friendliness, patience, and concern for people made me feel deeply grateful for the experience of getting to know him, even if I had not been selected. Throughout these four years of work, Michi has given me the utmost tolerance, unconditional trust, support, and always the fastest and timely feedback, which I truly appreciated. He has also had a profound impact on my scientific career. I have been inspired by his attitude towards knowledge, guiding students, writing and publishing, and academic communication, which will also be the most valuable experience in my life

I would like to thank Prof. Dr. Gerhard Haszprunar for providing me with work facilities at ZSM, giving me valuable advice and suggestions for my study, as well as sharing his extensive knowledge in malacology.

Thanks Prof. Dr. Gert Wörheide and Dr. Michael Eitel for organizing the IGNITE Comparative Genomics of Non-Model Invertebrates program. Thanks supervisors, Prof. Dr. Andreas Hejnol, Prof. Dr. Denis Tagu, Prof. Dr. Grace McCormack, Prof. Dr. Jean-François Flot, Prof. Agostinho Antunes, ESR peers Ksenia, Ramon, Nadège, Benjamin, Kenneth, Anne, Yihe, Fabian, Ferenc, Mattia, Paschalis, Antonio, Mariya and other IGNITE numbers for their help, support and valuable communications.

I want to thank the warm ZSM colleagues. Thanks, Franzi, for the meticulous checking and revision of my thesis, valuable comments on the content and structure, as well as on the language of the thesis. Thanks, Özgül, for the unforgettable time we were working together, and always fast reply for solving the server problems and giving me feedback. Thanks Isi, Eva, Tina, Nadine (always timely help for my work and life), Basti, Juan, Timea, Kathi, Pete (for help during work and showing me the amazing mollusk world), Alex, Fredi, Uli for nice lunchbreaks and time. I would like to thank to other colleagues from ZSM, thanks for all the kindness, warm smiles, greetings and help.

I would like to express my heartfelt thanks to Jimmy who has always been by my side, sharing my every improvement, small successes, happiness, and giving me the greatest comfort and support when I was not well. And thanks to my parents, my sister, who are far away in China but have always been with me, caring about me, and proud of me.

## 9. Appendix

### 9.1 Declaration of contributions to each publication

Diese Dissertation wurde angefertigt unter der Leitung von Prof. Dr. Michael Schrödl an der Zoologischen Staatssammlung München und der Fakultät für Biologie der Ludwig-Maximilians-Universität München.

#### Contribution to paper I:

Özgül Doğan, Michael Schrödl, Zeyuan Chen (2019): The complete mitogenome of *Arion vulgaris* Moquin-Tandon, 1855 (Gastropoda: Stylommatophora): mitochondrial genome architecture, evolution and phylogenetic considerations within Stylommatophora. *PeerJ* 8:e8603

I conceived and designed this study together with Özgül Doğan and Michael Schrödl. I performed the mitogenome assembly, annotation, dating of stylommatophoran phylogenetic tree, and the analysis of selective pressure on stylommatophorans, as well as the corresponding writings, tables and figures of these parts.

#### Contribution to paper II:

Zeyuan Chen, Michael Schrödl (2022): How many single-copy orthologous genes from whole genomes reveal deep gastropod relationships? *PeerJ* 10:e13285

I conceived this study together with Michael Schrödl. I performed all the analysis, designed and created figures, tables, drafted and wrote the manuscript.

#### Contribution to paper III:

Zeyuan Chen, Ozgul Dogan, Nadège Guiguelmoni, Anne Guichard & Michael Schrödl (2022): Pulmonate slug evolution is reflected in the *de novo* genome of *Arion vulgaris* Moquin-Tandon, 1855. *Scientific Reports* 12, 14226

I conceived this study together with Michael Schrödl. I performed the genome assembly, genome annotation and all of the evolutionary analysis. I designed all the figures and tables, and drafted the manuscript.

I hereby confirm the above statement.

---

(Zeyuan Chen)

---

(Prof. Dr. Michael Schrödl)

---

(Ort, Datum)

---

(Ort, Datum)



## 9.2 Curriculum Vitae

### Personal data

---

Name	Chen Zeyuan
Date of birth	1996.06.22
Place of birth	China, Shanxi province
E-mail	arielzeyuan@gmail.com

### Education/work experience

---

#### Ph.D.

10.2018-present	Ludwig-Maximilians-Universität München Zoologische Staatssammlung München <b>Mollusca genomics, Bioinformatics</b>
-----------------	--

#### Master

09.2015-07.2018	Sichuan University <b>Genomics, Ecology</b>
-----------------	--

#### Bachelor

09.2011-07.2015	Quanzhou Normal University <b>Biotechnology</b>
-----------------	--

### Conference & workshop

---

08.2022, Munich Germany	World Congress of Malacology Munich (Talk+Poster)
06.2022, Split Croatia	IGNITE Final Meeting (Talk)
11.2021, Online	GIGA IV Conference (Talk)
02.2021, Online	IGNITE's 4th Network Wide Training
02.2020, Rennes France	Program DSB 2020
09.2019, Zagreb Croatia	Computation approaches to comparative genomics II (Talk+Poster)
04.2019, Banyuls France	Computation approaches to comparative genomics I (Talk+Poster)
12.2018, Split Croatia	Programming and bioinformatic skills
10.2015, Chengdu China	Species diversity - theory and practice

### Secondment

---

01.2020, Rennes France	Institute for Genetics Environment and Plant Protection <b>Genome assembly</b>
09.2019, Galway Ireland	Molecular Evolution and Systematics Laboratory, National University of Ireland <b>DNA extraction, Flow cytometry</b>

### Grants received

---

03.2022	The Malacological Society of London Grant
10.2018-12.2020	Marie Skłodowska-Curie Grant
09.2015-07.2018	Class II Scholarship of National Master's Degree

## Reviews for journals

---

05.2022

Frontiers in Marine Science

## Scientific publications

---

**Chen, Z.**, Doğan, Ö., Guiglielmoni, N., Guichard, A., Schrödl, M. (2022). Pulmonate slug evolution is reflected in the *de novo* genome of *Arion vulgaris* Moquin-Tandon, 1855. *Sci Rep* 12:14226

**Chen, Z.**, Schrödl, M. (2022). How many single-copy orthologous genes from whole genomes reveal deep gastropod relationships? *PeerJ* 10:e13285

Doğan, Ö., Schrödl, M. & **Chen, Z.** (2020). The complete mitogenome of *Arion vulgaris* Moquin-Tandon, 1855 (Gastropoda: Stylommatophora): Mitogenome architecture, evolution and phylogenetic considerations within Stylommatophora. *PeerJ* 8:e8603

**Chen, Z.**, Ai, F., Zhang, J., Ma, X. ...& Ma, T. (2020). Survival despite isolation, inbreeding and asexual reproduction in the Tropics: insights from the genome of the world's southernmost poplar (*Populus ilicifolia*). *Plant J* 103:430-442

Su, Y., Bai, X., Yang, W., Wang, W., **Chen, Z.**, Ma, J., Ma, T. (2018). Single-base-resolution methylomes of *Populus euphratica* reveal the association between DNA methylation and salt stress. *Tree Genet Genomes* 14:86

Wang, W., Liu, H., He, Q., Yang, W., **Chen, Z.**, Wang, M., Su, Y., Ma, T. (2017). Characterization of the complete chloroplast genome of *Camptotheca acuminata*. *Conservation Genet Resour* 9:241-243

**Chen, Z.**, Wang, W., Yang, W., Ma, T. (2016). Characterization of the complete chloroplast genome of *Populus ilicifolia*. *Conservation Genet Resour* 8:391-393

Yang, W., Wang, W., Zhang, L., **Chen, Z.**, Guo, X., Ma, T. (2016). Characterization of the complete chloroplast genome of *Idesia polycarpa*. *Conservation Genet Resour* 8:271-273

Munich, October 2022

Zeyuan Chen

### **9.3 Eidesstattliche Versicherung und Erklärung**

#### Eidesstattliche Erklärung

Ich versichere hiermit an Eides statt, dass die vorgelegte Dissertation von mir selbständig und ohne unerlaubte Hilfe angefertigt ist.

München, den 01.10.2022

(Zeyuan Chen)

#### Erklärung

Hiermit erkläre ich, dass die Dissertation nicht ganz oder in wesentlichen Teilen einer anderen Prüfungskommission vorgelegt worden ist und dass ich mich nicht anderweitig einer Doktorprüfung ohne Erfolg unterzogen habe.

München, den 01.10.2022

(Zeyuan Chen)

文件名: **Chen\_Zeyuan.docx**  
目录: **/Users/zeyuan/Library/Containers/com.microsoft.Word/Data/Documents**  
模板: **/Users/zeyuan/Library/Group Containers/UBF8T346G9.Office/User Content.localized/Templates.localized/Normal.dotm**  
标题: **A glimpse into mollusk genomics from the de novo genome of Arion vulgaris Moquin-Tandon, 1855 (Gastropoda, Stylommatophora)**  
主题: **A glimpse into mollusk genomics from the de novo genome of Arion vulgaris Moquin-Tandon, 1855 (Gastropoda, Stylommatophora)**  
作者: **Zeyuan Chen**  
关键词:  
备注:  
创建日期: **2023/3/2 18:14:00**  
修订号: **2**  
上次保存日期: **2023/3/2 18:14:00**  
上次保存者: **Zeyuan Chen**  
编辑时间总计: **1 分钟**  
上次打印时间: **2023/3/2 18:14:00**  
打印最终结果  
页数: **203**  
字数: **30,388 (约)**  
字符数: **168,961 (约)**