

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

Zeyuan Chen^{1,2} and Michael Schrödl^{1,2,3}

¹ Mollusca, SNSB-Bavarian State Collection of Zoology, Munich, Bavaria, Germany

² Department Biology II, Ludwig-Maximilians-Universität München, Munich, Bavaria, Germany

³ 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.

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](https://doi.org/10.7717/peerj.13285)

© Copyright

2022 Chen and Schrödl

Distributed under

Creative Commons CC-BY 4.0

OPEN ACCESS

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

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	<i>Biomphalaria glabrata</i>	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	<i>Elysia chlorotica</i>	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>Plakobranchus ocellatus</i>	GCA_019648995.1	Scaffold	–	–	Maeda et al. (2021)
H	<i>Radix Auricularia</i>	GCA_002072015.1	Scaffold	–	–	Schell et al. (2017)
C	<i>Lautoconus ventricosus</i>	GCA_018398815.1	Chromosome	–	–	Pardos-Blas et al. (2021)
C	<i>Pomacea canaliculata</i>	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	<i>Marisa cornuarietis</i>	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	<i>Conus consors</i>	GCA_004193615.1	Scaffold	–	–	Andreson et al. (2019)
C	<i>Conus tribblei</i>	GCA_001262575.1	Scaffold	–	–	Barghi et al. (2016)
C	<i>Lanistes nyassanus</i>	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	<i>Haliotis rufescens</i>	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	<i>Chrysomallon squamiferum</i>	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	<i>Lottia gigantea</i>	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.

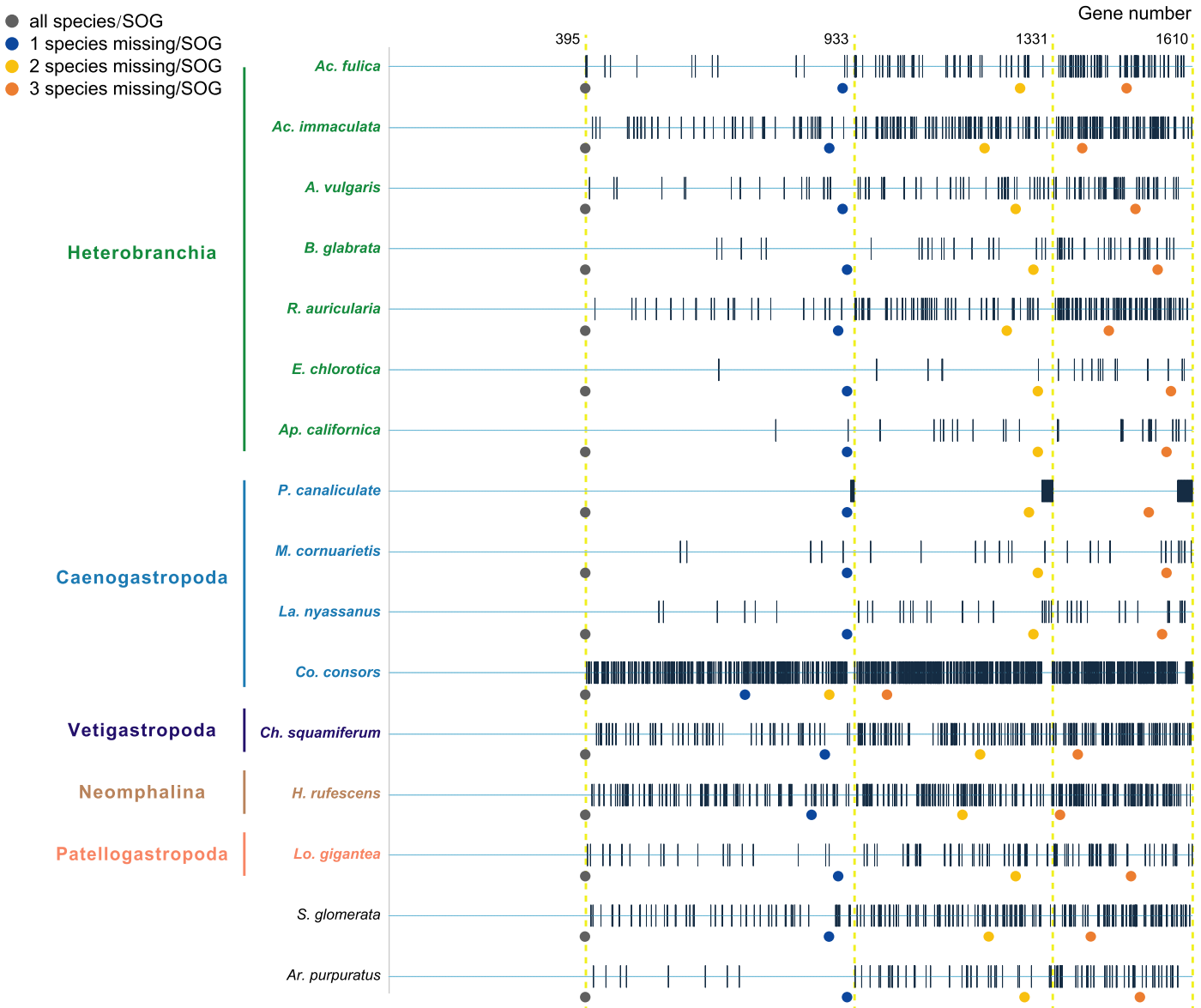


Figure 1 The four data sets used to infer gastropod relationships. The first data set includes single-copy orthologous genes (SOGs) identified in all 16 species per SOG and results in 395 SOGs. 933, 1,331, 1,610 SOGs were identified while allowed 1–3 species missing per SOG respectively. The blue line 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

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).

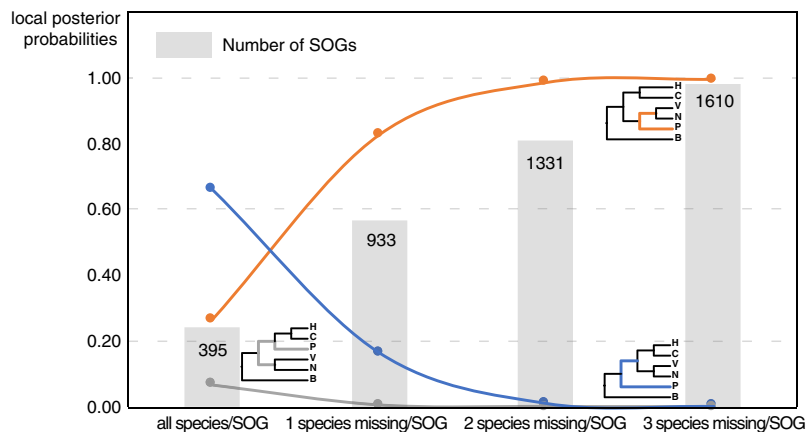


Figure 2 The gastropod relationships constructed with different data sets and the corresponding posterior probabilities of different topology. H: Heterobranchia, C: Caenogastropoda, V: Vetigastropoda, N: Neomphalina, P: Patellogastropoda, B: Bivalve.

Full-size DOI: 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 Patellogastropoda 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):

Patellogastropoda is sister to all other gastropods with a low support (PP = 0.66), and the alternative topologies of Patellogastropoda being sister to Vetigastropoda + Neomphalina (PP = 0.27), Patellogastropoda 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 Patellogastropoda as sister to Vetigastropoda + Neomphalina as dominant (PP = 0.83), the possibility of Patellogastropoda sister to all

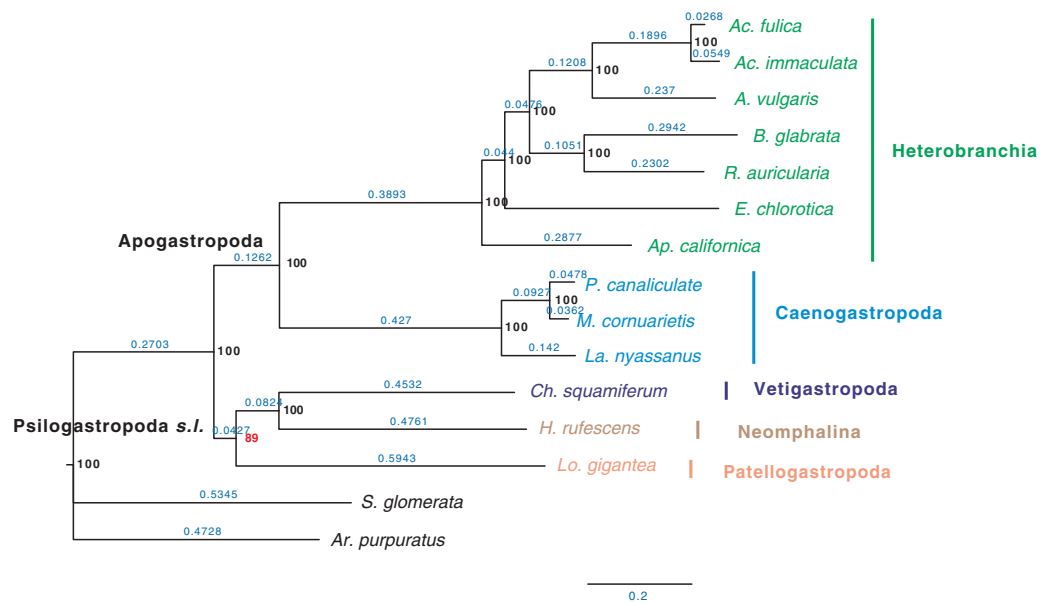


Figure 3 Gastropod phylogeny inferred from 847 single-copy orthologous genes (SOGs) identified from 15 species except *C. consors*. Bootstrap support percentages are indicated for each internal branch. Branch length is marked with blue numbers. Full-size DOI: 10.7717/peerj.13285/fig-3

other gastropods highly decreased (PP = 0.17), and the posterior probabilities of Patellogastropoda sister to Heterobranchia + Caenogastropoda 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 queried 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, Tennessen 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.
- Anderson 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.
- Fryda 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- Kano Y, Brenzinger B, Nutzal 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.
- 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.
- 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.
- 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.
- 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.
- 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.

- 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, *Plakobranthus 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, Bissessur 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, Boclaer 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, Crouzillat 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).