Received: 26 November 2022 | Revised: 8 March 2023 | Accepted: 27 March 2023

DOI: 10.1111/1755-0998.13793

RESOURCE ARTICLE

Target-capture probes for phylogenomics of the Caenogastropoda

Tricia C. Goulding | **Ellen E. Strong** | **Andrea M. Quattrini**

Smithsonian Institution, National Museum of Natural History, Washington, District of Columbia, USA

Correspondence
Tricia C. Goulding, National Museum of Natural History, PO Box 37012, MRC 163, Washington, DC 20013, USA.
Email: tc.goulding@gmail.com

Funding information
Smithsonian Institution, National Museum of Natural History

Handling Editor: Angus Davison

Abstract

Target-capture approaches have facilitated a rapid growth in the field of phylogenomics but few probe sets exist for molluscs, an exceptionally rich phylum with unparalleled ecological and morphological diversity. We designed and tested the first universal probe set using Phyluce to capture ultraconserved elements (UCEs) and exon loci from the Subclass Caenogastropoda - one of six major lineages of gastropods. The probe set consists of 29,441 probes designed to target 1142 UCE loci and 1933 exon loci (3075 total). In silico analyses of our probe set yielded an average of 2110 loci from genomes and 1389 loci from transcriptomes of diverse caenogastropods, from which an average of 1669 and 849 loci were retained respectively after screening to remove those that matched multiple contigs. Phylogenetic analyses of the loci extracted from transcriptomes produced well-supported trees very similar to those published based on transcriptomic analyses. Phylogenetic relationships estimated from loci extracted from genomes recover similar phylogenetic relationships, and indicate that the loci targeted with this probe set are informative for resolving deep phylogenetic relationships. An in vitro analysis of the probe set with the Epitoniidae, a diverse caenogastropod family of uncertain affinity and with poorly resolved evolutionary relationships, recovered a total of 2850 loci. Although preliminary, the analysis of loci captured by our probe set for a small number of epitoniid taxa produced a well-resolved tree indicating that this probe set is also able to resolve relationships at shallower hierarchical scales. Together, the in silico and in vitro analyses indicate that target-capture enrichment with this probe set is a useful tool for reconstructing phylogenetic relationships across taxonomic levels and evolutionary time scales.

KEYWORDS
exons, molluscs, target capture, ultraconserved elements

1 | INTRODUCTION

Genomic data are rapidly improving our understanding of deep evolutionary relationships among metazoans and have the potential to help address many questions about the diversification of invertebrates. Mollusks are the second most diverse phylum of animals, but this diversity is vastly under-represented among available genomics resources, with approximately 33 genomes published (Schell et al., 2017; Yang et al., 2020) for the estimated 85,000 recognized extant species (MolluscaBase). Genomic data

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.
© 2023 The Authors. *Molecular Ecology Resources* published by John Wiley & Sons Ltd.

from transcriptomes have greatly illuminated ancient relationships among mollusks, including relationships between classes (Kocot et al., 2011; Smith et al., 2011) and other major lineages (Cunha & Giribet, 2019; Kocot et al., 2013, 2019; Lemer et al., 2019; Lindgren & Anderson, 2018; Uribe et al., 2022; Zapata et al., 2014). However, obtaining transcriptome data for systematic studies is difficult as it requires intensive sampling effort to collect and preserve specimens for RNA extraction. In contrast, hundreds to thousands of genomic loci can be readily obtained from specimens in natural history collections using sequence capture methods and high-throughput sequencing (Jones & Good, 2016; Sproul & Maddison, 2017).

The long-accepted division of Gastropoda into Prosobranchia, Opisthobranchia and Pulmonata is obsolete, as is the division of Prosobranchia into Archaeo-, Meso- and Neogastropoda. The current phylogenetic classification of the Gastropoda (Bouchet et al., 2017) recognizes six clades at the rank of subclass, namely Patellogastropoda, Vetigastropoda, Neomphaliones, Neritimorpha, Caenogastropoda, and Heterobranchia. Of these, the Caenogastropoda includes over 34,000 valid marine, freshwater and terrestrial species, accounting for 47% of all Recent gastropod species (MolluscaBase). With a long evolutionary history extending to the Paleozoic, relationships within the group have remained challenging to resolve with morphology and standard molecular loci owing in part to the antiquity of its origins. It also contains many extinct lineages and many hyperdiverse clades that have radiated quickly, further complicating efforts to resolve their phylogeny (Ponder et al., 2008, 2020). Apart from a well-established, large, mostly marine, crown clade Neogastropoda (15,181 species, or 44% of all caenogastropods), the Caenogastropoda includes several groups of which the monophyly and/or relationships are unstable and debated, and in some cases evidently constitute grades rather than clades: the land and freshwater Architaenioglossa (3923 species); the marine and freshwater Cerithioidea (1546 species); the marine, land and freshwater Littorinimorpha (11,025 species); and the exclusively marine Ptenoglossa (Triphoroidea 1774 species, and Epitonioidea 788 species), Abyssochrysoidea (56 species) and Campaniloidea (16 species). Based on traditional Sanger sequencing, there has been significant progress with regards to the circumscription and evolutionary relationships within several of these groups (e.g., Criscione et al., 2017; Fedosov et al., 2018, 2019; Galindo et al., 2016; Kantor et al., 2017, 2022; Osca et al., 2015; Puillandre et al., 2008; Strong et al., 2011, 2019; Takano et al., 2022; Takano & Kano, 2014; Wang et al., 2021; Wilke et al., 2013). However, a backbone phylogeny is still wanting, especially for what was formerly called the Mesogastropoda, i.e. the non-neogastropod part of the Caenogastropoda, within which the phylogenetic validity of such taxa as Hypsogastropoda, Sorbeoconcha, and Ptenoglossa is not adequately established. Thus far, analyses of exons from transcriptomes have been used to investigate relationships among caenogastropods, but data are available for only a limited subset of taxa (Cunha & Giribet, 2019; Krug et al., 2022; Zapata et al., 2014).

Target-capture enrichment is a cost-effective method to sequence genomic data from diverse taxa for phylogenomic studies (Faircloth et al., 2012), and has been utilized to enrich DNA for high-throughput sequencing from plants, vertebrates, insects, and spiders (Branstetter et al., 2017; Hamilton et al., 2016; Hughes et al., 2021; McLay et al., 2021; Smith et al., 2014). In mollusks, sequence capture approaches are beginning to be used for phylogenomics of genera, families, and superfamilies (Abdelkrim et al., 2018; Layton et al., 2020; Ortiz-Sepulveda et al., 2022; Pfeiffer et al., 2019). However, there have only been a few universal probe sets developed for non-arthropod invertebrates, including ophiuroids (Hugall et al., 2016), anthozoans (Cowman et al., 2020; Erickson et al., 2021; Quattrini et al., 2018), and Heterobranchia, the second-most diverse lineage of gastropods with ~44% of extant species (Moles & Giribet, 2021; MolluscaBase) that, like Caenogastropoda, also has marine, freshwater, and terrestrial members.

Here, we designed a universal probe set to target ultraconserved elements (UCEs) and exons from the Caenogastropoda. We used published genomes and transcriptomes from the Architaenioglossa, Cerithioidea, and diverse representatives of the Hypsogastropoda to identify highly conserved regions and design probes to sequence these and their variable flanking regions for phylogenomic analyses. These probes were then tested in silico using published genomic data, and in vitro with the Epitoniidae (wentletraps). The latter is a diverse family with over 788 species recognized in the recent fauna (MolluscaBase) of mostly benthic species found from intertidal depths to the deep sea, but which also includes the bubble-rafting *Janthina* (Bouchet et al., 2017). Species diversity of epitoniids is relatively well characterized from shallow water habitats, but the higher order classification has been difficult to resolve with morphology due to convergence in shell characters (Bouchet & Warén, 1986; Kilburn, 1985).

2 | MATERIALS AND METHODS

2.1 | UCE identification and probe design

Genome scaffold assemblies for five species were downloaded from the European Nucleotide Archive, ENA (www.ebi.ac.uk, Table 1) and one unpublished genome was made available for our use (Supplementary Material, Whelan et al., 2022). These genomes were used in probe design and in silico testing, representing three lineages of caenogastropods: Architaenioglossa (*Lanistes nyassanus*, *Pomacea canaliculata*, and *P. maculata*), Cerithioidea (*Leptoxis ampla*), and Neogastropoda (*Anentome helena* and *Babylonia areolata*). Additional genomes for six caenogastropods and two outgroup gastropod lineages were downloaded and utilized for in silico tests. BUSCO scores estimating the completeness of each genome were estimated in BUSCO 5.2.2 using the metazoa_odb10 database (Manni et al., 2021; Simão et al., 2015).

The detailed workflow of Faircloth (2017) combined with the program PHYLUCE version 1.6.7 was used to identify conserved regions and design probes to target these regions following the online tutorial (https://phyluce.readthedocs.io/en/v1.6.8/tutorial-four.html).

17550998, 2023, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1755-0998.13793 by IFREMER Centre Bretagne BLP, Wiley Online Library on [10/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 The number of scaffolds and N50 for each genome and transcriptome used in the in silico analyses.

Family	Clade	Species	Accession	Assembly source	Scaffolds	N50	# Loci recovered	
							Recovered	Final
Genomes								
Ampullariidae	Architaenioglossa	*Lanistes nyassanus*	GCA_004794575.1	Sun et al. (2019)	17,149	317,839	2409	1646
Ampullariidae	Architaenioglossa	*Marisa cornuarietis*	GCA_004794655.1	Sun et al. (2019)	659	4,359,112	2384	1614
Ampullariidae	Architaenioglossa	*Pomacea canaliculata*	GCA_004794335.1	Sun et al. (2019)	3129	32,644,854	2378	1605
Ampullariidae	Architaenioglossa	*Pomacea maculata*	GCA_004794325.1	Sun et al. (2019)	3908	375,864	2389	1605
Pleuroceridae	Cerithioidea	*Leptoxis ampla*	unpublished	Supplementary Material, Whelan et al. (2022)	1,077,730	13,448	2371	2184
Provannidae	Hypsogastropoda	*Alviniconcha marisindica*	GCA_018857735.1	Yang et al. (2022)	3926	727,552	2400	2129
Babyloniidae	Neogastropoda	*Babylonia areolata*	GCA_011634625.1	B., Thai, M. H., Tan, P., Lee, L., Croft, & C. M., Austin, unpublished)	340,123	6321	1931	1733
Conidae	Neogastropoda	*Conus betulinus*	GCA_016801955.1	Peng et al. (2021)	41,413	232,489	1578	1393
Conidae	Neogastropoda	*Conus tribblei*	GCA_001262575.1	Barghi et al. (2016)	1,126,156	2681	1899	1693
Conidae	Neogastropoda	*Conus ventricosus*	ASM1839881v1	Pardos-Blas et al. (2021)	19,399	93,519,712	1822	1614
Nassariidae	Neogastropoda	*Anentome helena*	GCA_009936545.1	S., Pirro & D., Wagner, unpublished)	2,637,315	2,075,175	1645	1332
Raphitomidae	Neogastropoda	*Phymorhynchus buccinoides*	GCA_017654935.1	Z., Liu, unpublished)	18,181	336,037	1710	1479
Planorbidae (Outgroup)	Heterobranchia	*Biomphalaria glabrata*	GCA_014524955.1	Adema et al. (2017)	927	2,598,223	1760	394
Peltospiridae (Outgroup)	Neomphalida	*Chrysomallon squamiferum*	GCA_012295275.1	Zeng et al. (2020)	1	48,234,470	316	74
Transcriptomes								
Ampullariidae	Architaenioglossa	*Pomacea diffusa*	SRR1505132	Cunha and Giribet (2019)	243,447	713	1201	709
Viviparidae	Architaenioglossa	*Cipangopaludina catayensis*	GCEL01000000.1	Gerdol et al. (2018)	150,998	1006	2074	715
Viviparidae	Architaenioglossa	*Sinotaia quadrata* (as *Bellamya purificata*)	SRR12456152	This study	423,867	658	2246	909
Batillariidae	Cerithioidea	*Batillaria attramentaria*	SRR6214976	Krug et al. (2022)	169,139	1460	2132	1301
Pachychilidae	Cerithioidea	*Tylomelania sarasinorum*	SRR6214976	Krug et al. (2022)	169,117	1453	2134	1309
Planaxidae	Cerithioidea	*Hinea brasiliana*	SRR1505112	This study	126,005	375	1196	709
Semisulcospiridae	Cerithioidea	*Semisulcospira coreana*	GGNX01000000	Lee et al. (2019)	134,402	1058	2770	1755
Atlantidae	Hypsogastropoda	*Atlanta ariejansseni*	GIOD01000000.1	Wall-Palmer et al. (2021)	97,483	1762	528	283

17550998, 2023, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1755-0998.13793 by IFREMER Centre Bretagne BLP, Wiley Online Library on [10/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 (Continued)

Family	Clade	Species	Accession	Assembly source	Scaffolds	N50	# Loci recovered	
							Recovered	Final
Bithyniidae	Hypsogastropoda	*Bithynia siamensis goniomphalos*	SRR768418	Cunha and Giribet (2019)	133,941	581	1620	803
Calyptraeidae	Hypsogastropoda	*Crepidula navicella*	SRR3168547	Cunha and Giribet (2019)	150,443	858	1839	1063
Charoniidae	Hypsogastropoda	*Charonia tritonis*	SRR5468098	Cunha and Giribet (2019)	168,386	674	1359	808
Epitoniidae	Hypsogastropoda	*Janthina janthina*	SRR1505114	Cunha and Giribet (2019)	38,873	890	1254	766
Littorinidae	Hypsogastropoda	*Echinolittorina malaccana*	SRR1269556	Cunha and Giribet (2019)	154,493	475	1891	1197
Littorinidae	Hypsogastropoda	*Littorina saxatilis*	GHPE01000000	Kaliberdina and Granovich (2003)	32,801	2260	1262	717
Naticidae	Hypsogastropoda	*Euspira heros*	SRR1505131	Cunha and Giribet (2019)	32,775	690	590	360
Naticidae	Hypsogastropoda	*Neverita didyma*	GHHQ01000000	Q., Wang, G., Fu, & H., Guo, unpublished	14,374	2328	1729	850
Pomatiidae	Hypsogastropoda	*Pomatias elegans*	SRR11015453	Krug et al. (2022)	87,292	2166	2529	1227
Pomatiopsidae	Hypsogastropoda	*Oncomelania hupensis*	SRR1284711	Cunha and Giribet (2019)	163,221	1291	2112	1133
Provannidae	Hypsogastropoda	*Alviniconcha strummeri*	GJGI01000000.1	Breusing et al. (2022)	28,848	2082	1966	1140
Rubyspiridae	Hypsogastropoda	*Rubyspira osteovora*	SRR1505136	Cunha and Giribet (2019)	105,445	475	1537	1043
Strombidae	Hypsogastropoda	*Aliger gigas* (as *Lobatus gigas*)	SRR827578	Cunha and Giribet (2019)	70,388	842	478	289
Tateidae	Hypsogastropoda	*Potamopyrgus antipodarum*	GFLZ01000000.1	Bankers et al. (2017)	62,862	1341	1956	1006
Velutinidae	Hypsogastropoda	*Marseniopsis mollis*	SRR3205281-SRR3205287	Krug et al. (2022)	89,072	1008	1890	1065
Colubrariidae	Neogastropoda	*Cumia reticulata*	ERR852086	Cunha and Giribet (2019)	70,506	822	691	391
Conidae	Neogastropoda	*Conus consors*	SRR1958882	Cunha and Giribet (2019)	83,231	633	725	448
Drilliidae	Neogastropoda	*Clavus canalicularis*	SRR11423814	This study	52,418	784	935	488
Melongenidae	Neogastropoda	*Volegalea cochlidium*	SRR1005737	Cunha and Giribet (2019)	89,593	256	169	124
Muricidae	Neogastropoda	*Nucella lapillus*	SRX357400	Krug et al. (2022)	112,163	413	1460	988
Muricidae	Neogastropoda	*Rapana venosa*	GDIA01047641	Song et al. (2019)	246,627	1047	1746	1075
Muricidae	Neogastropoda	*Urosalpinx cinerea*	SRR1505141	Cunha and Giribet (2019)	85,261	753	1309	823
Nassariidae	Neogastropoda	*Anentome helena*	SRR11015439	Krug et al. (2022)	572,055	705	2222	1342
Pseudomelatomidae	Neogastropoda	*Crassispira cerithina*	SRR1574922	Cunha and Giribet (2019)	29,562	486	476	325
Amphibolidae (Outgroup)	Heterobranchia	*Phallomedusa solida*	SRR1505127	Cunha and Giribet (2019)	134,588	991	1516	824
Neritidae (Outgroup)	Neritimorpha	*Nerita melanotragus*	SRR1920139	Cunha and Giribet (2019)	129,362	856	1806	519

Note: The total number of loci recovered per sample is indicated, as well as the final number of nonduplicate loci.

17550998, 2023, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1755-0998.13793 by IFREMER Centre Bretagne BLP, Wiley Online Library on [10/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

All analyses were run on the Smithsonian High Performance Computing Cluster. First, repetitive regions, retroelements, small RNAs and transposons were masked using Repeat Masker with maximum divergence level of a repeat to its consensus sequence set to 50% (Smit et al., 2015). Genomic sequence data were converted into 2bit files using faToTwoBit from the BLAT Suite (Kent, 2002). As part of this workflow, 100-bp paired end reads were simulated from each genome without sequencing error using the program ART_ILLUMINA (Huang et al., 2012). Simulated reads from three exemplar taxa, *B. areolata*, *L. ampla*, and *P. canaliculata*, were subsequently aligned to the base genome *A. helena*, which was selected for its comparatively large scaffold size. Stampy version 1.0.32 was used to map conserved regions of the simulated reads from these taxa to the base genome with less than 5% sequence divergence. The alignment file of mapped reads was converted to a BED format using BEDtools, and then sorted by position and merged into putative conserved regions (Quinlan & Hall, 2010). To remove repetitive regions from the mapped reads, the program PHYLUCE (Faircloth, 2016) was used with the command phyluce_probe_strip_masked_loci_from_set to remove regions where the base genome was less than 80 bp (--min-length 80) and where more than 25% of the base genome was masked by Repeat Masker (--filter-mask 0.25). From this, an SQLite database was created for 205,168 loci across the exemplar genomes. Each genome was queried from the SQLite table to identify loci shared, conserved loci between the base genome and the three exemplar taxa, with a total of 6140 loci identified from all four genomes.

From the 6140 conserved loci, regions up to 160 bp were extracted from the *A. helena* base genome using phyluce_probe_get_genome_sequences_from_bed in Phyluce. This code filters out sequences less than 160 bp, as well as those with 25% or more bases masked by Repeat Masker, and sequences with ambiguous bases. There were 5590 loci retained, which were used to design a temporary probe set. Probes of 120 bp were tiled so that there was 40 bp of overlap between probes. The temporary probe set was screened to remove probes with >25% masked bases or GC content above 70% or below 30%, resulting in 11,008 probes targeting 5523 loci. Potential duplicate probes and paralogs were then removed using lastz. Probes were removed if they were >50% identical over >50% of their length, after which 8427 probes remained. To identify the loci from our base genome in the other three exemplar genomes, the temporary probe set was aligned back to the exemplar genomes with a minimum identity value of 50%. Sequences of 180 bp were extracted from the exemplar genomes to locate the conserved loci targeted with the probe set. An SQLite database was created for the loci targeted with the temporary probe set and queried to identify loci present in all four exemplar genomes, which resulted in a total of 1244 loci.

To allow the probe set to work more consistently across a broad range of caenogastropods, three other exemplar genomes were used to design additional probes targeting the 1244 selected loci. The FASTA files for each locus were used to design probes of 120 bp from each genome, tiling density set to 3x density with probes overlapping in the middle, and duplicates, masked bases, and high or low GC content removed. Duplicate loci were identified by aligning the probes to themselves at >50% identity and > 50% coverage, which filtered out 24 duplicate loci and retained 9627 probes targeting 1220 loci in the UCE probe set.

2.2 | Exon identification and probe design

Transcriptome assemblies for 15 caenogastropod species and two outgroup gastropods were downloaded from the supplementary materials of Cunha and Giribet (2019) and assemblies of six additional transcriptomes analysed in Krug et al. (2022) were obtained from the authors. Transcriptome assemblies were downloaded for eight taxa and unassembled sequence reads were downloaded for another three taxa from the European Nucleotide Archive, www.ebi.ac.uk (Table 1). Unassembled sequence reads were assembled with SPADES version 3.14.1 (Bankevich et al., 2012). BUSCO scores estimating the completeness of each transcriptome were estimated as before. Specimens used in probe design were selected to represent divergent lineages of caenogastropods with nearly complete transcriptomes when possible, although in the case of *Janthina janthina*, a transcriptome with a BUSCO score of 22.8% was used as it was the only representative available for the Epitoniidae. Assembled sequence data were converted into 2bit files using faToTwoBit from BLAT Suite (Kent, 2002). For probe design, 100-bp paired end reads were simulated from each transcriptome using the program ART_ILLUMINA (Huang et al., 2012).

To design probes targeting exon regions, the methods used to identify UCEs were repeated with the available transcriptome data. Four taxa representing diverse caenogastropods were selected to identify loci for probe design: *Crepidula navicella* (Calyptraeidae), *Echinolittorina malaccana* (Littorinidae), *Rapana venosa* (Muricidae), and *Oncomelania hupensis* (Pomatiopsidae). 100-bp reads for each species were simulated and mapped to the exemplar transcriptome of *O. hupensis*. After converting the alignments to BED files, merging overlapping reads, and filtering the data to remove repetitive regions and short sequences, an SQLite table was created based on 96,761 loci.

A total of 13,131 loci shared among *O. hupensis* and the exemplar taxa were extracted from the base genome using Phyluce. After filtering sequences less than 160 bp, with 25% masked bases, and sequences with ambiguous bases, 12,753 loci were retained. A temporary probe set targeting these loci in *O. hupensis* was designed and filtered to remove potentially problematic probes with greater than 25% repeat content and high or low GC content, with a total of 25,411 probes designed targeting 12,725 loci. Potential duplicates were removed with a lastz search of the probes to themselves in Phyluce (>50% identical >50% of their length), after which 15,029 probes remained in the temporary probe set.

The temporary probe set was aligned to the four exemplar transcriptomes plus three additional species, *Janthina janthina* (Epitoniidae), *Semisulcospira coreana* (Semisulcospiridae), and *Rubyspira osteovora* (Rubyspiridae). At this stage, slightly longer

17550998, 2023, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1755-0998.13793 by IFREMER Centre Bretagne BLP, Wiley Online Library on [10/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

sequences of 180 bp were extracted from the seven exemplar genomes to locate the conserved loci targeted with the probe set. An SQLite database was created for 7302 loci targeted with the temporary probe set and queried to identify loci identified in the exemplar taxa. Few loci (85 loci) were present in the transcriptomes of all seven exemplars. The number of loci for subsets of the seven species varied from 613 loci (shared by six taxa) to 6783 loci (shared by only two taxa). We decided to target loci present in five of the seven taxa (2065 loci) for probe design. The FASTA files for each locus were used to design probes of 120 bp from each genome, tiling density set to 3×, with duplicates, masked bases, and high or low GC content removed. Duplicate loci were identified by aligning the probes to themselves at 50% identity and 50% coverage, which filtered out 52 duplicate loci, and retained 21,463 exon probes targeting 2013 loci in the exon probe set.

2.3 | Final probe screening

The exon and UCE probe sets were concatenated in a single set of 31,090 probes targeting 3233 loci. Probes were screened against each other to remove redundant probes (50% identical over 50% of their length) using phyluce_probe_easy_lastz and phyluce_probe_remove_duplicate_hits_from_probes_using_lastz. This probe-set file was sent to Arbor BioSciences, where BLAST analyses were conducted to check for specificity against the genomes of *L. nyassanus* and *B. areolata*, and the transcriptomes for *O. hupensis*, *S. coreana*, and *R. venosa*. Probes that failed relaxed BLAST filtering for multiple hits in either genomes or transcriptomes were removed. Additional BLAST filtering against the mitochondrial genomes of *Thylacodes squamigerus* (NC_014588.1), *Columbella adansoni* (KP716637.2), and *Semisulcospira coreana* (NC_037771.1) was also conducted to ensure no mitochondrial loci were included in the bait set. This stringent filtering pipeline retained 29,441 probes targeting 3075 loci.

2.4 | In silico probe test

An in silico test was performed to check how well the final complete probe set aligned to available genomes and transcriptomes, again following the Phyluce tutorial. First, phyluce_probe_run_multiple_lastzs_sqlite was used to align the final probe set to the six 2bit formatted genomes using an identity value of 50%. For each probe test, the matching FASTA data were sliced out of each genome, plus 200 bp of the 5′ and 3′ flanking regions, using phyluce_probe_slice_sequence_from_genomes. These were then filtered in a final screen of the probe set to remove loci that matched multiple contigs with the minimum match between the probes and contigs set to 67% (Table 1). This process was repeated with the 2bit formatted transcriptomes.

Sequence data were extracted from the genomes and transcriptomes separately using the probe set, exported as FASTA files, and aligned separately with MAFFT (Katoh et al., 2002) in Phyluce. The alignments were trimmed internally using GBlocks (Castresana, 2000) as implemented in Phyluce, and a 50% complete matrix was assembled for each in Phyluce. The resulting matrices were analysed in IQTree version 2.1.2 (Minh et al., 2020) with model selection via ModelFinder (Kalyaanamoorthy et al., 2017) and 1000 ultrafast bootstraps (Hoang et al., 2018).

The consensus tree was visualized in FIGTREE version 1.4.4 and rooted with *Chrysomallon squamiferum* for the tree based on sequences extracted from genomes, and with *Nerita melanotragus* for the tree based on sequences from the transcriptomes. Bayesian analyses were conducted using ExaBayes v1.5.1 with two runs and two coupled chains run in parallel for 1,000,000 generations (Aberer et al., 2014).

2.5 | Target enrichment with probes

Following the in silico tests, the final probe set of 120 bp probes was synthesized by Arbor BioSciences as a custom MyBaits kit. DNA from 21 caenogastropods, including 16 epitoniids, was extracted with either the AutoGen platform (phenol-chloroform) or the EZNA Mollusc DNA extraction kit (Omega Bio-Tek) with overnight tissue digestion at 56°C. All specimens were collected between 2010 and 2021 and were preserved in 95% EtOH (Table 2). The concentration of DNA isolated from each specimen was quantified using a Qubit 2.0 fluorometer. Genomic DNA from 16 epitoniids (50–1200 ng per specimen, mean = 440 ng) was sent to Arbor BioSciences for library preparation, target enrichment, and DNA sequencing. Individual libraries were prepared by Arbor targeting an average insert size of approximately 500 nucleotides with Illumina TruSeq adapters and custom dual indexes. These were pooled in groups of 8 for target enrichment and concentrated to 7 µL by vacuum centrifugation. The myBaits version 5.02 protocol was used by Arbor for target enrichment with an overnight hybridization at 65°C. Post-capture, half of the volume of the capture reactions was amplified for 10 cycles. Products were quantified with a spectrofluorimetric assay and a quantitative PCR assay. For captures that did not generate sufficient DNA for equimolar pooling, the second half of the capture volume was amplified for 14 cycles. Captures were pooled in approximately equilmolar ratios and sequenced on an Illumina NovaSeq 6000 on a partial S4 PE150 lane targeting approximately 1 Gbp per library.

Libraries for five caenogastropods belonging to the families Aclididae, Cymatiidae, Eulimidae, and Vanikoridae were prepared at the Laboratories Analytical Biology (LAB) at the Smithsonian Institution's National Museum of Natural History. Sera-mag SpeedBeads were prepared for library preparation following Faircloth and Glenn (2011). Prior to library preparation, up to 500 ng of DNA (mean 277 ng) was fragmented to 400–800 bp using sonication with the QSonica Inc. Sonicator Q800R. The size of the DNA fragments was checked via gel electrophoresis in a 1.5% agarose gel run at 100 V for 45 min. A DNA cleanup was performed after fragmentation with 3X Sera-mag SpeedBeads. The protocol of Faircloth (2015) was followed for library preparation using the Kapa

17550998, 2023, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1755-0998.13793 by IFREMER Centre Bretagne BLP, Wiley Online Library on [10/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 Collection data for specimens of Epitoniidae sequenced for in vitro analyses.

Species	Voucher	Expedition	Collection year	Accession
Alexania inazawai	MNHN-IM-2013-81265	KOUMAC 2.1	2018	SAMN33415858
Cirsotrema pumiceum	MNHN-IM- 2019-12542	CORSICABENTHOS2	2020	SAMN33415850
Cirsotrema sp. 1	MNHN-IM-2013-44940	MADEEP 2014	2014	SAMN33415852
Cirsotrema sp. 2	MNHN-IM-2013-63248	KANACONO	2016	SAMN33415860
Epidendrium sordidum	MNHN-IM-2013-50954	KAVIENG 2014	2014	SAMN33415851
Epitonium sp. 1	MNHN-IM-2019-3232	KOUMAC 2.3	2019	SAMN33415862
Epitonium sp. 2	MNHN-IM-2019-7772	KOUMAC 2.3	2019	SAMN33415856
Epitonium sp. 3	MNHN-IM-2019-7768	KOUMAC 2.3	2019	SAMN33415853
Epitonium sp. 4	MNHN-IM-2013-61919	ZhongSha 2015	2015	SAMN33415855
Gyroscala sp.	MNHN-IM-2013-72805	MADIBENTHOS	2016	SAMN33415859
Janthina exigua	MNHN-IM-2013-68245	KANACONO	2016	SAMN33415861
Opalia sp. 1	MNHN-IM-2019-7770	KOUMAC 2.3	2019	SAMN33415848
Opalia sp. 2	MNHN-IM-2013-60227	KARUBENTHOS 2	2015	SAMN33415854
Opalia burryi	MNHN-IM-2013-72421	MADIBENTHOS	2016	SAMN33415863
Opaliopsis sp.	MNHN-IM-2013-45613	MADEEP 2014	2014	SAMN33415857
Surrepifungium costulatum	MNHN-IM-2013-53656	KAVIENG 2014	2014	SAMN33415849
Costaclis sp.[a]	MNHN-IM-2019-2138	KANADEEP 2	2019	SAMN33415865
Gyrineum lacunatum[a]	MNHN-IM-2009-19703	ATIMO VATAE	2010	SAMN33415864
Kimberia sp.[a]	MNHN-IM-2019-20019	SPANBIOS	2021	SAMN33415866
Monoplex sp.[a]	MNHN-IM-2009-19701	ATIMO VATAE	2010	SAMN33415867
Vanikoro sp.[a]	MNHN-IM-2009-19643	ATIMO VATAE	2010	SAMN33415868

[a]Caenogastropod outgroups.

HyperPrep Kit (Kapa Biosystems) with half of the reaction volume. This corresponds to a 30 μL reaction for end-repair/A-tailing, with 3.5 μL end repair and A-Tailing buffer, 1.5 μL A-tailing enzyme, and 25 μL of DNA suspended in PCR-grade water. For the ligation of y-yoke adapters, a master mix with universal y-yoke oligonucleotide adapters was made from 15 μL ligation buffer, 5 μL DNA ligase, 2.5 μL PCR-grade water, and 2.5 μL of y-yoke adapter stubs, and added to the 30 μL end-repaired/A-tailed DNA. Following ligation, a master mix of 25 μL of HiFi HotStart polymerase (Kapa Biosystems) and 5 μL of PCR-grade water was mixed for each sample and added to 15 μL of adapter-ligated library and 5 μL of iTru dual-indexed primer mix (iTru5 and iTru7) (5 μM, Glenn et al., 2019). The following thermal protocol was used: 98°C for 45 s followed by 10–13 cycles of 98°C for 15 s, 60°C for 30 s, 72°C for 30 s; and a final extension of 72°C for 1 min. The resulting reactions were purified using 1× Sera-mag SpeedBeads cleanup, and resuspended in 23 μL of 10 mM Tris–HCL. The concentration of the libraries was quantified using a Qubit fluorometer (2 μL, post- PCR amplification). The custom MyBaits 5.02 kit from Arbor Biosciences was used following the manufacturer's standard protocol for target enrichment with a hybridization temperature of 65°C. Five outgroups were pooled with other gastropods in groups of eight for target enrichment. An Agilent TapeStation was used to estimate the size of DNA fragments and the concentration of DNA was measured with the Qubit fluorometer. A final 1× bead cleanup was performed if adapter dimers were present in the pooled libraries. Enriched libraries were combined in equimolar ratios into one pool and sent to Oklahoma Medical Research Facility for sequencing with an Illumina NovaSeq (150 bp paired end reads).

2.6 | Post-sequencing analyses

Demultiplexed Illumina reads were processed using Phyluce version 1.7.1 following the workflow in the online tutorial. The reads were first trimmed to remove adapters and low quality bases using illumiprocessor (Faircloth, 2013) modified on the Smithsonian High Performance Computing Cluster to use TrimGalore (https://github.com/FelixKrueger/TrimGalore) with the standard illumiprocessor options plus the options --tg-length and --tg-quality to access filtering. Reads were then assembled using SPADES version 3.14.1 (Bankevich et al., 2012). Probes were matched to the assemblies of each sample using phyluce_assembly_match_contigs_to_probes to locate loci with the minimum identity and minimum coverage set to 70%. Loci were then extracted using phyluce_assembly_get_match_counts and phyluce_assembly_get_fastas_from_match_counts, exported into FASTA files and aligned with MAFFT. Alignments were trimmed with GBlocks using default parameters and data matrices of locus alignments were created for a 70% complete data matrix in Phyluce. Maximum likelihood and Bayesian analyses were run

17550998, 2023, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1755-0998.13793 by IFREMER Centre Bretagne BLP, Wiley Online Library on [10/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

in IQTREE version 2.1.2 (Minh et al., 2020) and EXABAYES version 1.5.1 (Aberer et al., 2014) with the same settings as in the in silico analyses.

3 | RESULTS

3.1 | Probe set

The final probe set contained 29,624 probes to target 3075 loci: 1220 UCE loci and 2013 exon loci. After BLAST filtering by Arbor BioSciences, 29,441 probes were retained: 8872 UCE probes and 20,569 exon probes (0.62% of the probes were removed). The probe set included roughly equal numbers of probes designed from each of the four genomes used: 2241 probes from *Anentome helena*, 2147 from *Babylonia areolata*, 2212 from *Leptoxis ampla*, and 2272 from *Pomacea canaliculata*. The number of probes targeting exons was slightly more variable, with 3239 probes from *Crepidula navicella*, 2976 from *Echinolittorina malaccana*, 2062 from *Janthina janthina*, 3365 from *Oncomelania hupensis*, 2763 from *Rapana venosa*, 2441 from *Rubyspira osteovora*, and 3723 from *Semisulcospira coreana*.

3.2 | In silico test

An average of 2110 loci were extracted from caenogastropod genomes (316–1760 loci for the outgroup taxa) and an average of 1389 loci from transcriptomes (1413–1646 loci for the outgroup taxa). After screening, an average of 1669 loci ± standard deviation (SD) of 256 were retained from the genomes (min. 1332, max. 2184), compared to 74 to 394 loci for the outgroups. From the transcriptomes, an average of 849 ± 337 SD (min. 124, max 1755) nonduplicate exon loci were retained after screening, compared to 332 and 519 loci for the two outgroups. The final number of loci recovered in the in silico analyses was plotted against the BUSCO score of each transcriptome and genome analysed. The number of loci extracted from transcriptomes was correlated with the BUSCO score (R^2 = .428, Figure 1b) but this correlation was weak for the loci extracted from genomes (R^2 = .061, Figure 1a). There is significant variability in the relationship between loci recovered from transcriptomes and BUSCO scores, but the transcriptomes with the lowest scores, *Volegalea cochlidium*, *Crassispira cerithina*, and *Euspira heros*, were each characterized by fewer than 400 loci recovered, while those with the highest scores, *Semisulcospira coreana*, *Oncomelania hupensis*, and *Rapana venosa*, were among those with the highest number of loci recovered (Figure 1b).

Three alignment matrices were generated, one for the UCE loci and two for the exon loci (exon data sets #1–2). The UCE data set included 12 caenogastropod taxa and two outgroups. In this data set, 2361 loci were each aligned and the mean alignment length was 458 bp (95% CI: 8.0) per locus. The 50% matrix for UCE loci included 1706 alignments with a total length of 672,594 bp, including 290,810 informative sites, and a mean alignment length of 394 (95% CI: 8.1). Exon data set #1 included 15 caenogastropod taxa and two outgroups included in a previous phylogenomic study of gastropod relationships (Cunha & Giribet, 2019), while exon data set #2 included sequence data from 34 taxa. A total of 1666 loci were aligned from exon data set #1 (17 taxa) with a mean alignment length of 443 bp (95% CI: 9.8). The 50% matrix for this data set included 389 loci with a total length of 166,357 bp, 48,716 informative sites, and a mean length of 427 (95% CI: 19.6). For exon data set #2 (34 taxa), a total of 2266 loci were aligned, with a mean length of 486 bp (95% CI: 8.6). The 50% matrix for data set #2 included 565 of these loci with a total length of 308,634 bp, 132,249 informative sites, and a mean length of 546 bp (95% CI: 15.7).

The ExaBayes analysis was stopped after 1,000,000 generations, with the average standard deviation of split frequencies for trees at 0.00% for the UCE data set and exon data set #1, and 1.64% for exon data set #2. The model selected in IQTree for analysis of the UCE data set was a transversion model of nucleotide substitutions with empirical base frequencies and the FreeRate model of heterogeneity (TVM + F + R3). The model selected for exon data set #1 was a transversion model of nucleotide substitutions with empirical base

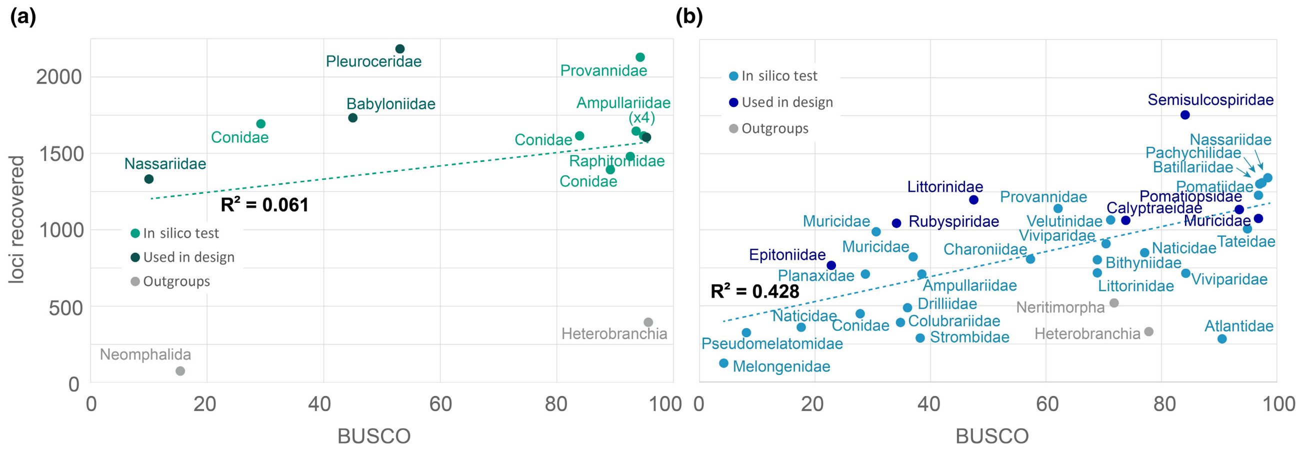


FIGURE 1 Relationship between number of loci recovered from in silico analyses of genomic data and BUSCO score after final screening of (a) genomes and (b) transcriptomes. Data points are labelled by taxonomic group.

17550998, 2023, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1755-0998.13793 by IFREMER Centre Bretagne BLP, Wiley Online Library on [10/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

frequencies and the FreeRate model of heterogeneity (TVM+F+R4) and exon data set #2 was general-time reversible model of nucleotide substitution with empirical base frequencies and the FreeRate model of heterogeneity (GTR+F+R5).

The phylogeny inferred from UCE data set #1 containing 12 caenogastropod taxa was highly supported at all nodes in both maximum likelihood and Bayesian analyses (Figure 2). The phylogeny estimated had posterior probability values (PP) of 1.0 for all nodes, while the maximum likelihood tree had bootstrap support of 95 or higher for all but two nodes (Figure 3a). The phylogeny estimated from data set #2 had PP of 1.0 at all but three nodes, one of which had very low support. In the maximum likelihood analysis, five nodes had bootstrap support less than 95 and all other nodes had full support (Figure 3b). The maximum likelihood analysis of exon data set #2 showed the same phylogenetic relationships as the Bayesian tree except the Conoidea (*C. consors*, *C. cerithina* and *C. canalicularis*) was sister group to the Buccinoidea (*A. helena*, *C. reticulata*, and *V. cochlidium*) instead of the Muricidaeand in the relationships within the Buccinoidea, which were not well resolved.

3.3 | In vitro analysis

The total number of Illumina reads obtained ranged from 1,143,889 to 24,969,972 per sample (mean 8,856,659 ± 6,187,558). Removal of adapters and low-quality reads led to the removal of 2.48% of reads per sample on average, leaving an average of 8,637,359 ± 6,217,399 SD trimmed reads remaining per sample. The trimmed reads were assembled into an average of 562,594 contigs per sample ± 709,566 with a mean length of 250 ± 95 bp.

Across all samples, the data set included 2850 UCE and exon loci. An average of 2221 loci were extracted per sample, of which an average of 1710 were retained per sample after filtering out loci that matched multiple contigs (Table 3). The mean number of informative sites per locus was 68 with a total of 193,083 informative sites in the complete data set. A 70% complete data matrix included 1328 loci. The ExaBayes analysis was stopped after 1,000,000 generations, with the average standard deviation of split frequencies for trees at 0.00%. The model selected in IQTree for analysis of the data set was a Transition model with empirical base frequencies and the FreeRate model of heterogeneity (TIM+F+R4). All nodes were fully supported in both maximum likelihood and Bayesian analyses (Figure 4).

4 | DISCUSSION

4.1 | In silico analysis of caenogastropod relationships

Over the last decade, the number of caenogastropod taxa for which transcriptomes have been published has increased steadily, while the number of genomes sequenced is still very limited. We used these published genomics resources to develop a universal probe set of 29,441 probes targeting both UCE and exon loci from diverse caenogastropods and found that the sequence data extracted with this probe set are informative for phylogenomic analyses of their relationships across multiple scales. Several hundred to more than 2100 loci were extracted from in silico analyses of published genomes and transcriptomes, although few loci were recovered from some transcriptomes with very low BUSCO scores (Figure 1b). A positive correlation was observed between BUSCO score and the number of loci obtained, although this relationship was weak in the genomes (R^2=.06145, Figure 1a) compared to the transcriptomes (R^2=.428, Figure 1b). The strong correlation in the latter is probably due to the inclusion of incomplete transcriptome assemblies, as indicated by low BUSCO scores, which contain few of the highly-conserved coding regions shared between many taxa. In contrast, UCE probes target highly conserved noncoding regions that might generally be present in genomes of diverse taxa, but are not included in the BUSCO calculations because they are noncoding. Thus, BUSCO scores are a useful metric when selecting transcriptomes for designing exon probe sets, but appear to be of limited utility for selecting genomes to design UCE probes.

We first examined transcriptomes analysed in a phylogenomic study (Cunha & Giribet, 2019) to compare how the phylogenetic relationships estimated with our probe set compared to analyses of the larger transcriptome. In that study, 1059 genes were analysed, which were translated from DNA sequences into amino acid sequences. Despite differences in the loci examined and the method of phylogenetic inference, the relationships estimated among caenogastropods with loci extracted with our probe set is nearly identical in topology to their published phylogeny (Figure 3a, Cunha & Giribet, 2019). Both analyses supported the monophyly of the Hypsogastropoda and Neogastropoda (Muricoidea, Conoidea and Buccinoidea) (Colgan et al., 2007; Cunha & Giribet, 2019), with the Tonnoidea (represented by the Charoniidae) as sister to the neogastropods. The only difference between the two topologies is in the relationship of *Janthina janthina* as sister group to a clade including the Littorinidae, Naticidae, and Rubyspiridae (*R. osteovora*, *E. heros*, and *E. malaccana*). This relationship was supported in some of the analyses by Cunha and Giribet (2019), but was not recovered in our analysis (Figure 3a). However, this relationship was recovered with moderate support in our analysis of an expanded data set including the Atlantidae and Pomatiidae (Figure 3b). This suggests that additional taxon sampling is important and necessary to resolve the relationships in this clade.

With additional taxon sampling in the larger data set (Figure 3b) several relationships were obtained that had not been recovered in previous phylogenomic analyses. First, the sister relationship of the Tonnoidea to the Neogastropoda is no longer recovered, instead it is sister to the Velutinoidea, consistent with other recent transcriptomic analyses (Krug et al., 2022). Relationships within the Neogastropoda also differ, with the Conoidea sister group to the Muricidae in this Bayesian analysis (Figure 3b) as well as published maximum likelihood analyses (Krug et al., 2022). Finally, the addition of cerithioid and viviparid taxa also allowed us to investigate the phylogenetic relationships between the Cerithioidea

17550998, 2023, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1755-0998.13793 by IFREMER Centre Bretagne BLP, Wiley Online Library on [10/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

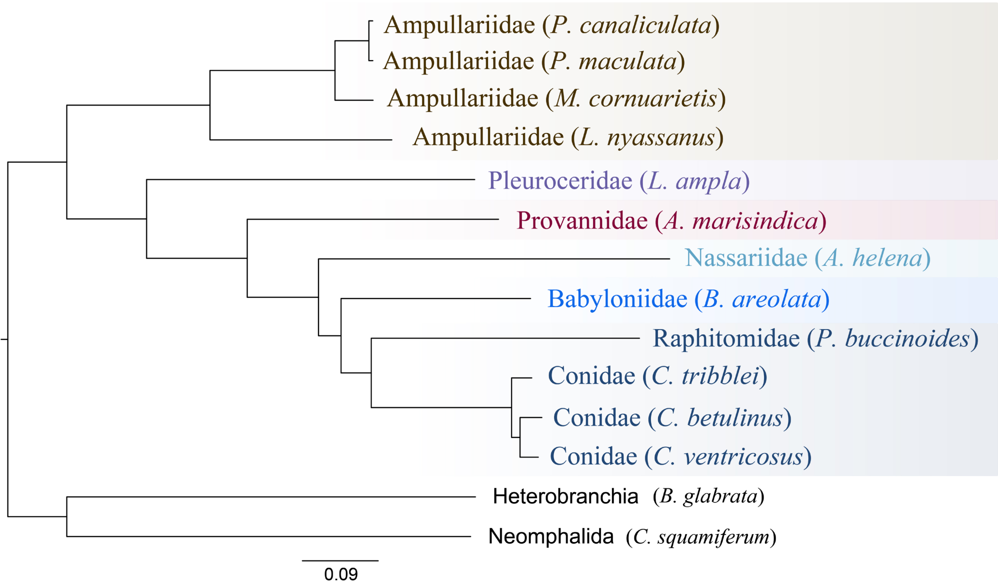


FIGURE 2 Tree from Bayesian analysis estimating phylogenetic relationships among caenogastropods based on 50% complete matrix from genomes. All nodes received full support in both Bayesian and maximum likelihood analyses of the concatenated data set. Each superfamily is highlighted with a different colour.

and the Architaenioglossa for the first time with genomic data. Surprisingly, the Architaenioglossa were found to be sister group to the Cerithioidea, a relationship that has not generally been recovered in morphological analyses, although a sister group relationship between cerithioids and some members of the Architaenioglossa has been recovered in a few analyses of molecular data (Harasewych et al., 1998; Osca et al., 2015; Wang et al., 2021). Our result contradicts the monophyly of the Sorbeoconcha, in which the Cerithioidea is sister to the Hypsogastropoda (Bouchet et al., 2017; Ponder et al., 2020).

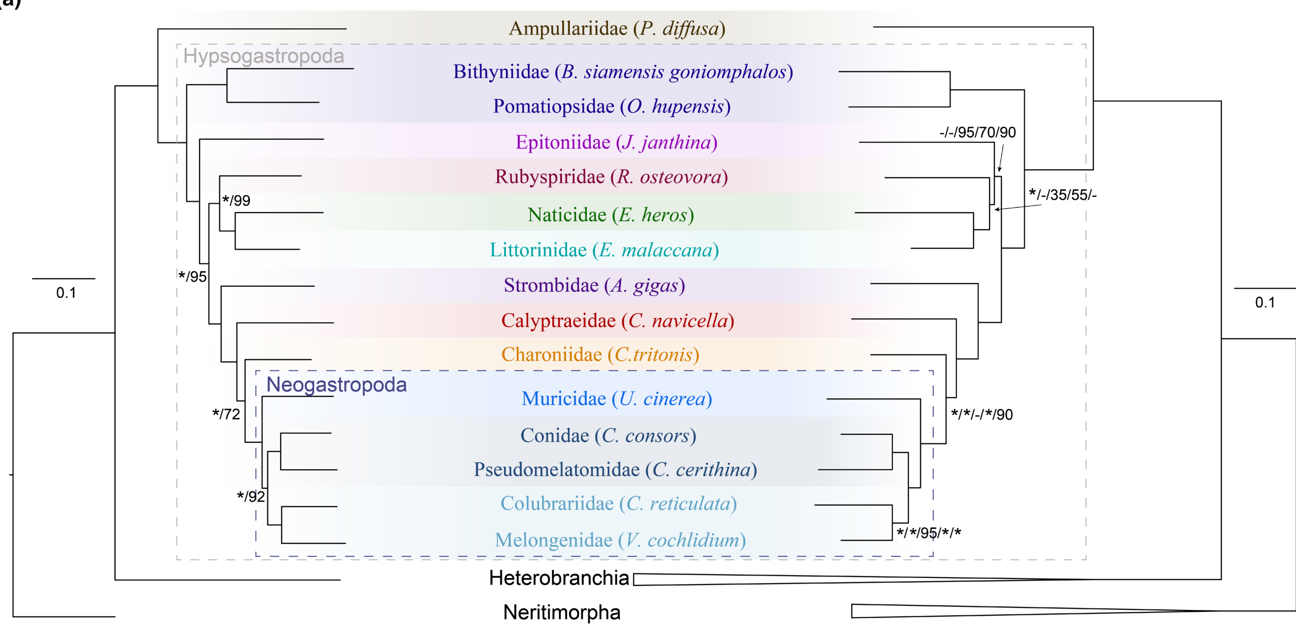
The results of the in silico analyses indicated that the probe set is informative for robustly resolving deep phylogenetic relationships across the Caenogastropoda. There are a few studies of evolutionary relationships of caenogastropods based on transcriptomes (Cunha & Giribet, 2019; Krug et al., 2022; Zapata et al., 2014), but building upon these studies with wider taxon sampling is difficult because the material must be specifically preserved for RNA sequencing. Using target-capture sequencing, we can expand the taxon sampling of these studies using existing museum collections, even using degraded DNA from historical specimens as shown in successful target-capture of fluid-preserved historical invertebrates (50 to 150 years old) such as corals (Untiedt et al., 2021) and spiders (Derkarabetian et al., 2019), or with small-bodied animals with very low input DNA (≤10 ng, Sproul & Maddison, 2017).

4.2 | In vitro analysis of the Epitoniidae

The probe set designed in this study was used to successfully enrich 2850 exon and UCE loci from a caenogastropod family of interest, the Epitoniidae. The deep evolutionary history of the Epitoniidae dating back at least to the early Cretaceous (Durham, 1937; Sohl, 1964) makes it an excellent candidate for phylogenomic study, and the high diversity of extant lineages that have radiated in association with stony corals are also a useful system for exploring ecological speciation in marine invertebrates (Gittenberger et al., 2006; Gittenberger & Gittenberger, 2005). On average, 1710 loci were retained per specimen in the epitoniid data set, many more than the 766 loci recovered from the in silico analysis of the *J. janthina* transcriptome and comparable to the number of loci extracted in the in silico analyses of published genomes (mean 1669 loci). This data set included representatives of 16 epitoniid species in 12 genera and five species from two other superfamilies as outgroups. The relationships among epitoniids and the outgroups are robustly resolved with all nodes receiving strong support in both maximum likelihood and Bayesian analyses (Figure 4).

Previous multilocus phylogenetic studies estimated conflicting relationships between epitoniid genera. A close relationship of *Epidendrium* and *Surrepifungium* to *Epitonium*, first hypothesized based on the mitochondrial phylogeny of Gittenberger et al. (2006) but not recovered in analyses of mitochondrial and nuclear loci (16S, 28S, H3, and H4) by Churchill et al. (2011), is supported here. In Churchill et al. (2011), *Epidendrium* and *Surrepifungium* were recovered as sister to each other, but were estimated to be more closely related to *Cirsotrema* and *Opalia* (0.93/57), epitoniids with much more robust shells, than to *Epitonium*. The relationships estimated here between epitoniid genera are similar to those obtained in the multilocus analyses (COI, 16S, 28S, 18S and H3) of the Hypsogastropoda of Takano and Kano (2014), in that we recover a close relationship between the neustonic *Janthina* and the benthic epitoniid *Alexania inazawai*, and this relationship is fully supported (whereas support for this relationship was moderate, PP 0.95/ bootstrap 72 in analyses of COI and 28S and low in multilocus analyses of Takano & Kano, 2014). *Opalia* and *Opaliopsis* are

17550998, 2023, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1755-0998.13793 by IFREMER Centre Bretagne BLP, Wiley Online Library on [10/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License



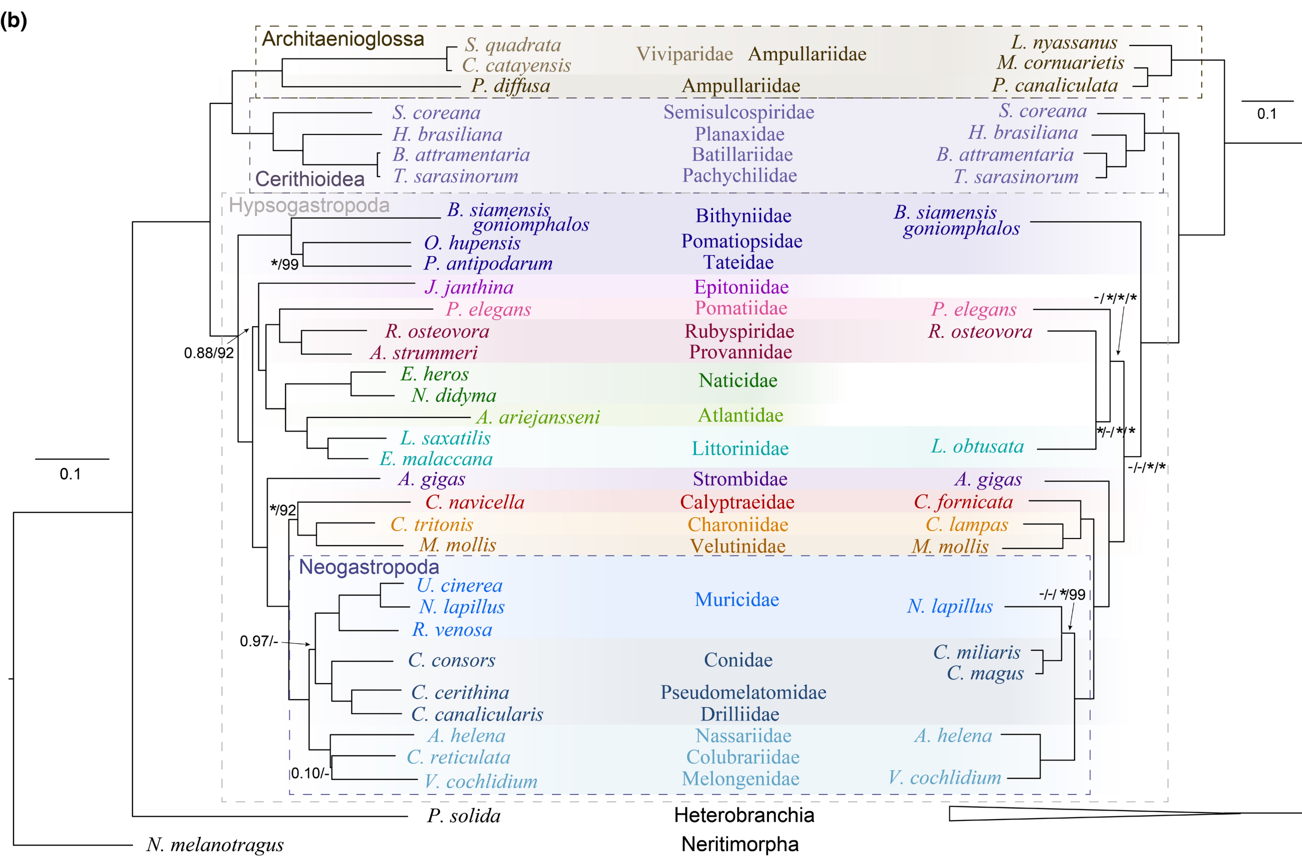


FIGURE 3 Tree from Bayesian analysis estimating phylogenetic relationships among caenogastropods based on 50% complete matrix from transcriptomes. All nodes received full support in all analyses of the concatenated data set unless values are given. "*" indicates PP of 1.0 or bootstrap support of 100 (a) exon data set #1 including 17 taxa (left) with support values from ExaBayes/IQTree compared to the topology from Cunha & Giribet, 2019 (right) with support values from two Bayesian analyses/Astral/two IQTree analyses from their M1 matrix (b) exon data set #2 including 34 taxa (left) with support values from ExaBayes/IQTree compared to caenogastropod outgroups analysed in Krug et al. (2022) (right) with support values for their two Bayesian and two maximum likelihood analyses. Each superfamily is highlighted with a different colour.

17550998, 2023, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1755-0998.13793 by IFREMER Centre Bretagne BLP, Wiley Online Library on [10/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

17550998, 2023, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1755-0998.13793 by IFREMER Centre Bretagne BLP, Wiley Online Library on [10/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 3 The number of contigs and contig length obtained from assembly of sequence reads from in vitro analyses of epitoniids and five caenogastropod outgroups (marked **).

Order: Superfamily	Species	# Contigs	Mean contig len. (bp)	Loci extr.	Mean locus len. (bp)	Loci removed	Loci retained
Unassigned: Epitonioidea	*Alexania inazawai*	273,985	259	2509	1051	723	1786
	Cirsotrema pumiceum	858,784	205	2444	1035	410	2034
	Cirsotrema sp. 1	272,699	189	2246	522	178	2068
	Cirsotrema sp. 2	1,136,584	185	2362	832	208	2154
	Epidendrium sordidum	325,280	222	2500	653	1238	1262
	Epitonium sp. 1	1,824,585	258	2454	1267	331	2123
	Epitonium sp. 2	164,181	512	2400	1385	916	1484
	Epitonium sp. 3	316,959	202	2320	921	425	1895
	Epitonium sp. 4	467,384	213	2514	997	832	1682
	Gyroscala sp.	410,574	342	2154	823	396	1758
	Janthina exigua	260,134	341	2221	1332	514	1707
	Opalia sp. 1	76,905	146	1940	559	148	1792
	Opalia sp. 2	386,792	182	2315	732	440	1875
	Opalia burryi	2,942,441	163	2479	932	451	2028
	Opaliopsis sp.	553,753	152	2442	668	390	2052
	Surrepifungium costulatum	1,161,812	161	2591	890	1228	1363
Littorinimorpha: Tonnoidea	*Gyrineum lacunatum*** (Cymatiidae)	21,867	379	2006	753	557	1449
Littorinimorpha: Tonnoidea	*Monoplex* sp.** (Cymatiidae)	130,787	172	2112	349	680	1432
Littorinimorpha: Vanikoroidea	*Kimberia* sp.** (Aclididae)	178,223	354	2000	1099	361	1639
Littorinimorpha: Vanikoroidea	*Costaclis* sp.** (Eulimidae)	19,027	322	662	376	69	593
Littorinimorpha: Vanikoroidea	*Vanikoro* sp.** (Vanikoridae)	31,714	290	1977	441	243	1734
	Mean	562,594	250	2221	839	511	1710

Note: The number of loci obtained from the Spades assembly is also detailed, with the mean locus length, the number of loci removed for matching multiple contigs, and the number of loci retained for phylogenetic analyses.

supported as sister taxa in our analyses, consistent with Takano and Kano (2014). The relationship of *Gyroscala* as sister group to a clade including *Epitonium*, *Surrepifungium*, *Epidendrium*, *Janthina*, and *Alexania* appears to contrast with a phylogeny based on COI barcode sequences alone (Gittenberger et al., 2006) in which *Gyroscala* sp. clustered with *Cirsotrema* sp., but the tree in that analysis seems to be unrooted, and this relationship might change with the addition of outgroups outside the Epitoniidae. The genera *Epitonium*, *Cirsotrema* and *Opalia* are each monophyletic in this backbone phylogeny (Figure 4), but broader taxonomic sampling is needed to assess the monophyly of each genus, particularly *Epitonium* which was recovered as polyphyletic in Gittenberger et al. (2006). Overall, these results demonstrate the utility of our probe set for resolving the relationships and affinities of both deeply divergent genera and closely related species in a highly diverse clade with higher support than obtained with Sanger-based approaches.

The in vitro analyses of 21 taxa recovered 2850 out of 3075 loci (92.7%) targeted with our probe set. However, only 1328 of these loci are represented in the 70% taxon matrix used for phylogenetic analyses. The low representation of loci in the alignment matrix is related to filtering out a large number of loci as potential duplicates in the Phyluce workflow, which were flagged as matching multiple contigs. Additional sequencing depth could improve the assembly of contigs in the future. With fewer contigs, the number of loci excluded from downstream analyses may be reduced. Modifications of the library preparation methods could also improve the number of loci recovered, such as using focused acoustic shearing to fragment the DNA rather than sonication, as sonication produces DNA fragments of a wide range of sizes, resulting in a large loss of DNA during size selection of the fragments (Bronner & Quail, 2019).

In our initial test of the probe set with five nonepitoniid caenogastropods, we recovered an average of 1369 loci from the Tonnoidea and Vanikoroidea. While this number is lower than the

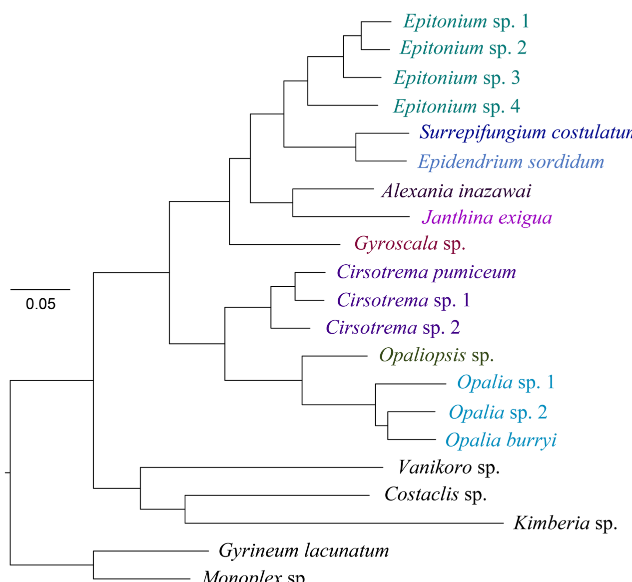


FIGURE 4 Estimated phylogenetic relationships from Bayesian analysis among the Epitoniidae based on a 70% taxon matrix. All nodes received full support in both Bayesian and maximum likelihood analyses of the concatenated data set.

average number recovered from Epitoniidae, the numbers are not directly comparable since methods for the library preparation of the two groups differed. In addition, the sequencing depth differs between these groups, with an average of 8.86 million reads per epitoniid compared to an average of 4.93 million reads from the other caenogastropods. Thus, the difference in the number of loci is likely due to the methods used to generate the data and should not be interpretted as reflecting a disparity in the performance of the probe set.

In the in vitro analyses, 225 loci were not recovered from any taxa. Although an epitoniid was used to design probes to capture exons, the transcriptome of *Janthina* had a low BUSCO score indicating the transcriptome was quite incomplete. Nonetheless, most loci targeted with our probe set were captured in our analyses. As additional taxa are analysed with the probe set, we can determine if these loci remain difficult to capture. If this is the case, additional probes could be designed from another epitoniid or related caenogastropod and added to the probe set.

5 | CONCLUSION

The universal exon and UCE probe set was developed from taxa representing ten families and deeply divergent lineages of the Caenogastropoda, including the Architaenioglossa, the Cerithioidea, and a diversity of Hypsogastropoda. In silico analyses of 30 caenogastropod families indicate that this probe set can be used to extract a high number of UCE and exon loci from deeply divergent caenogastropod lineages informative for phylogenomic analyses at multiple scales. The Epitoniidae is the first group for which the probe set has been used in vitro for library preparation and high throughput sequencing, and resulted in a phylogeny which resolved evolutionary relationships with high support that were inconsistently defined in previous studies. This resource will support phylogenomic studies of a diverse group of gastropods with few genomics tools currently available, and will allow a large diversity of caenogastropod taxa in museum collections amassed over more than a century of collecting to be incorporated into cutting-edge evolutionary and taxonomic studies.

AUTHOR CONTRIBUTIONS

Tricia C. Goulding and ES conceptualized the study, Tricia C. Goulding designed the probe set, analysed sequence data and wrote the manuscript, and Andrea M. Quattrini advised on methodology of probe design, sample preparation and analysis of sequence data. Tricia C. Goulding, Ellen E. Strong, and Andrea M. Quattrini reviewed and revised the manuscript.

ACKNOWLEDGEMENTS

This research was partially funded by the Peter Buck Postdoctoral Fellowship at the Smithsonian Institution. We would like to thank Philippe Bouchet and Nicolas Puillandre (Muséum national d'Histoire naturelle, Paris, France) for providing access to the epitoniids collected during various shore-based expeditions and deep-sea cruises, conducted, respectively, by MNHN and Pro-Natura International (PNI) as part of the Our Planet Reviewed programme, and by MNHN and Institut de Recherche pour le Développement (IRD) as part of the Tropical Deep-Sea Benthos programme (in alphabetic order): ATIMO VATAE (https://doi.org/10.17600/10110040; PI Philippe Bouchet); CORSICABENTHOS; KOUMAC (PI Philippe Bouchet); KANADEEP 2 (https://doi.org/10.17600/18000883; PI Sarah Samadi); KARUBENTHOS 2 (https://doi.org/10.17600/15005400; PIs Philippe Bouchet, Laure Corbari); KAVIENG 2014 (https://doi.org/10.17600/14004400; PI Philippe Bouchet); MADEEP (https://doi.org/10.17600/14004000; PI Sarah Samadi); MADIBENTHOS (PI Philippe Bouchet); and ZhongSha 2015 (PI Wei-Jen Chen). The organizers acknowledge the support of the Flotte Océanographique Française, Province Nord and Gouvernement de la Nouvelle-Calédonie, Collectivité de Corse, Fondation Prince Albert II, Fondation Total, Office Français de la Biodiversité, and Agence Nationale de la Recherche (ANR Taiwan France DeepEvo). All expeditions operated under the regulations then in force in the countries in question and satisfied the conditions set by the Nagoya Protocol for access to genetic resources. We are also grateful to Barbara Buge and Virginie Héros (MNHN) for preparing the loan of the specimens. We also thank Kevin Kocot for providing several transcriptome assemblies. Computations in this study were conducted on the Smithsonian High Performance Cluster (SI/HPC), Smithsonian Institution. https://doi.org/10.25572/SIHPC.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

17550998, 2023, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1755-0998.13793 by IFREMER Centre Bretagne BLP, Wiley Online Library on [10/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

DATA AVAILABILITY STATEMENT

Raw sequence reads have been deposited in the NCBI SRA (BioProject PRJNA937718). Aligned sequence files and the probe set are available through Dryad (https://doi.org/10.5061/dryad.b2rbnzskw).

BENEFIT-SHARING STATEMENT

The research in this publication complies with relevant national laws implementing the Convention on Biological Diversity and Nagoya Protocol.

ORCID

Tricia C. Goulding https://orcid.org/0000-0003-2947-1028
Andrea M. Quattrini https://orcid.org/0000-0002-4247-3055

REFERENCES

Abdelkrim, J., Aznar-Cormano, L., Fedosov, A. E., Kantor, Y. I., Lozouet, P., Phuong, M. A., Zaharias, P., & Puillandre, N. (2018). Exon-capture-based phylogeny and diversification of the venomous gastropods (Neogastropoda, Conoidea). *Molecular Biology and Evolution*, *35*(10), 2355–2374. https://doi.org/10.1093/molbev/msy144

Aberer, A. J., Kobert, K., & Stamatakis, A. (2014). Exabayes: Massively parallel Bayesian tree inference for the whole-genome era. *Molecular Biology and Evolution*, *31*(10), 2553–2556. https://doi.org/10.1093/molbev/msu236

Adema, C. M., Hillier, L. W., Jones, C. S., Loker, E. S., Knight, M., Minx, P., Oliveira, G., Raghavan, N., Shedlock, A., do Amaral, L. R., Arican-Goktas, H. D., Assis, J. G., Baba, E. H., Baron, O. L., Bayne, C. J., Bickham-Wright, U., Biggar, K. K., Blouin, M., Bonning, B. C., … Wilson, R. K. (2017). Whole genome analysis of a schistosomiasis-transmitting freshwater snail. *Nature Communications*, *8*(1), 1–12. https://doi.org/10.1038/ncomms15451

Bankers, L., Fields, P., McElroy, K. E., Boore, J. L., Logsdon, J. M., Jr., & Neiman, M. (2017). Genomic evidence for population-specific responses to co-evolving parasites in a New Zealand freshwater snail. *Molecular Ecology*, *26*(14), 3663–3675. https://doi.org/10.1111/mec.14146

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, *19*(5), 455–477. https://doi.org/10.1089/cmb.2012.0021

Barghi, N., Concepcion, G. P., Olivera, B. M., & Lluisma, A. O. (2016). Structural features of conopeptide genes inferred from partial sequences of the *Conus tribblei* genome. *Molecular Genetics and Genomics*, *291*(1), 411–422. https://doi.org/10.1007/s00438-015-1119-2

Bouchet, P., Rocroi, J. P., Hausdorf, B., Kaim, A., Kano, Y., Nützel, A., Parkhaev, P., Schrödl, M., & Strong, E. E. (2017). Revised classification, nomenclator and typification of gastropod and monoplacophoran families. *Malacologia*, *61*(1–2), 1–526. https://doi.org/10.4002/040.061.0201

Bouchet, P., & Warén, A. (1986). Revision of the northeast Atlantic bathyal and abyssal Aclididae, Eulimidae, Epitoniidae (Mollusca, Gastropoda). *Bollettino Malacologico, Supplemento 2*, 299–576. https://doi.org/10.5962/bhl.title.140732

Branstetter, M. G., Longino, J. T., Ward, P. S., & Faircloth, B. C. (2017). Enriching the ant tree of life: Enhanced UCE bait set for genome-scale phylogenetics of ants and other Hymenoptera. *Methods in Ecology and Evolution*, *8*(6), 768–776. https://doi.org/10.1111/2041-210X.12742

Breusing, C., Genetti, M., Russell, S. L., Corbett-Detig, R. B., & Beinart, R. A. (2022). Horizontal transmission enables flexible associations with locally adapted symbiont strains in deep-sea hydrothermal vent symbioses. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(14), e2115608119. https://doi.org/10.1073/pnas.2115608119

Bronner, I. F., & Quail, M. A. (2019). Best practices for Illumina library preparation. *Current Protocols in Human Genetics*, *102*(1), e86. https://doi.org/10.1002/cphg.86

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, *17*(4), 540–552. https://doi.org/10.1093/oxfordjournals.molbev.a026334

Churchill, C. K. C., Foighil, Ó. D., Strong, E. E., & Gittenberger, A. (2011). Females floated first in bubble-rafting snails. *Current Biology*, *21*(19), 802–803. https://doi.org/10.1016/j.cub.2011.08.011

Colgan, D. J., Ponder, W. F., Beacham, E., & Macaranas, J. (2007). Molecular phylogenetics of Caenogastropoda (Gastropoda: Mollusca). *Molecular Phylogenetics and Evolution*, *42*(3), 717–737. https://doi.org/10.1016/j.ympev.2006.10.009

Cowman, P. F., Quattrini, A. M., Bridge, T. C. L., Watkins-Colwell, G. J., Fadli, N., Grinblat, M., … Baird, A. H. (2020). An enhanced target-enrichment bait set for Hexacorallia provides phylogenomic resolution of the staghorn corals (Acroporidae) and close relatives. *Molecular Phylogenetics and Evolution*, *153*, 106944. https://doi.org/10.1016/j.ympev.2020.106944

Criscione, F., Ponder, W. F., Köhler, F., Takano, T., & Kano, Y. (2017). A molecular phylogeny of Rissoidae (Caenogastropoda: Rissooidea) allows testing the diagnostic utility of morphological traits. *Zoological Journal of the Linnean Society*, *179*(1), 23–40. https://doi.org/10.1111/zoj.12447

Cunha, T. J., & Giribet, G. (2019). A congruent topology for deep gastropod relationships. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1898), 20182776. https://doi.org/10.1098/rspb.2018.2776

Derkarabetian, S., Benavides, L. R., & Giribet, G. (2019). Sequence capture phylogenomics of historical ethanol-preserved museum specimens: Unlocking the rest of the vault. *Molecular Ecology Resources*, *19*(6), 1531–1544. https://doi.org/10.1111/1755-0998.13072

Durham, J. W. (1937). Gastropods of the family Epitoniidae from Mesozoic and Cenozoic rocks of the west coast of North America, including one new species by F. E. Turner and one by R. A. Bramkamp. *Journal of Paleontology*, *11*(6), 479–512.

Erickson, K. L., Pentico, A., Quattrini, A. M., & McFadden, C. S. (2021). New approaches to species delimitation and population structure of anthozoans: Two case studies of octocorals using ultraconserved elements and exons. *Molecular Ecology Resources*, *21*(1), 78–92. https://doi.org/10.1111/1755-0998.13241

Faircloth, B. C. (2013). *Illumiprocessor: A trimmomatic wrapper for parallel adapter and quality trimming*. https://doi.org/10.6079/J9ILL

Faircloth, B. C. (2015). *Illumina TruSeq library prep for target enrichment*.

Faircloth, B. C. (2016). PHYLUCE is a software package for the analysis of conserved genomic loci. *Bioinformatics*, *32*(5), 786–788. https://doi.org/10.1093/bioinformatics/btv646

Faircloth, B. C. (2017). Identifying conserved genomic elements and designing universal bait sets to enrich them. *Methods in Ecology and Evolution*, *8*, 1103–1112. https://doi.org/10.1111/2041-210X.12754

Faircloth, B. C., & Glenn, T. (2011). *Serapure v.2.2. Ecol. and Evol. Biology*. University of California, Los Angeles. https://ethanomics.files.wordpress.com/2012/08/serapure_v2-2.pdf

Faircloth, B. C., McCormack, J. E., Crawford, N. G., Harvey, M. G., Brumfield, R. T., & Glenn, T. C. (2012). Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology*, *61*(5), 717–726. https://doi.org/10.1093/sysbio/sys004

17550998, 2023, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1755-0998.13793 by IFREMER Centre Bretagne BLP, Wiley Online Library on [10/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Fedosov, A., Puillandre, N., Herrmann, M., Kantor, Y., Oliverio, M., Dgebuadze, P., Modica, M. V., & Bouchet, P. (2018). The collapse of *Mitra*: Molecular systematics and morphology of the Mitridae (Gastropoda: Neogastropoda). *Zoological Journal of the Linnean Society*, *183*(2), 253–337. https://doi.org/10.1093/zoolinnean/zlx073

Fedosov, A. E., Caballer Gutierrez, M., Buge, B., Sorokin, P. V., Puillandre, N., & Bouchet, P. (2019). Mapping the missing branch on the neogastropod tree of life: Molecular phylogeny of marginelliform gastropods. *Journal of Molluscan Studies*, *85*(4), 439–451. https://doi.org/10.1093/mollus/eyz028

Galindo, L. A., Puillandre, N., Utge, J., Lozouet, P., & Bouchet, P. (2016). The phylogeny and systematics of the Nassariidae revisited (Gastropoda, Buccinoidea). *Molecular Phylogenetics and Evolution*, *99*, 337–353. https://doi.org/10.1016/j.ympev.2016.03.019

Gerdol, M., Cervelli, M., Oliverio, M., & Modica, M. V. (2018). Piercing fishes: Porin expansion and adaptation to hematophagy in the vampire snail *Cumia reticulata*. *Molecular Biology and Evolution*, *35*(11), 2654–2668. https://doi.org/10.1093/molbev/msy156

Gittenberger, A., & Gittenberger, E. (2005). A hitherto unnoticed adaptive radiation: Epitoniid species (Gastropoda: Epitoniidae) associated with corals (Scleractinia). *Contributions to Zoology*, *74*(1–2), 125–203. https://doi.org/10.1163/18759866-0740102009

Gittenberger, A., Kokshoorn, E., & Gittenberger, E. (2006). *A molecular phylogeny of Epitoniidae (Mollusca: Gastropoda), focusing on the species associated with corals. Gittenberger a: The evolutionary history of parasitic gastropods and their coral hosts in the indo-Pacific*. PhD thesis, Leiden University, 207–213.

Glenn, T. C., Nilsen, R. A., Kieran, T. J., Sanders, J. G., Bayona-Vásquez, N. J., Finger, J. W., Pierson, T. W., Bentley, K. E., Hoffberg, S. L., Louha, S., Garcia-De Leon, F. J., del Rio Portilla, M. A., Reed, K. D., Anderson, J. L., Meece, J. K., Aggrey, S. E., Rekaya, R., Alabady, M., Belanger, M., … Faircloth, B. C. (2019). Adapterama I: Universal stubs and primers for 384 unique dual-indexed or 147,456 combinatorially-indexed Illumina libraries (iTru & iNext). *PeerJ*, *7*, e7755.

Hamilton, C. A., Lemmon, A. R., Lemmon, E. M., & Bond, J. E. (2016). Expanding anchored hybrid enrichment to resolve both deep and shallow relationships within the spider tree of life. *BMC Evolutionary Biology*, *16*(1), 1–20. https://doi.org/10.1186/s12862-016-0769-y

Harasewych, M. G., Adamkewicz, S. L., Plassmeyer, M., & Gillevet, P. M. (1998). Phylogenetic relationships of the lower Caenogastropoda (Mollusca, Gastropoda, Architaenioglossa, Campaniloidea, Cerithioidea) as determined by partial 18S rDNA sequences. *Zoologica Scripta*, *27*(4), 361–372. https://doi.org/10.1111/j.1463-6409.1998.tb00467.x

Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, *35*(2), 518–522.

Huang, W., Li, L., Myers, J. R., & Marth, G. T. (2012). ART: A next-generation sequencing read simulator. *Bioinformatics*, *28*(4), 593–594.

Hugall, A. F., O'Hara, T. D., Hunjan, S., Nilsen, R., & Moussalli, A. (2016). An exon-capture system for the entire class Ophiuroidea. *Molecular Biology and Evolution*, *33*(1), 281–294. https://doi.org/10.1093/molbev/msv216

Hughes, L. C., Ortí, G., Saad, H., Li, C., White, W. T., Baldwin, C. C., Crandall, K. A., Arcila, D., & Betancur-R, R. (2021). Exon probe sets and bioinformatics pipelines for all levels of fish phylogenomics. *Molecular Ecology Resources*, *21*(3), 816–833. https://doi.org/10.1111/1755-0998.13287

Jones, M. R., & Good, J. M. (2016). Targeted capture in evolutionary and ecological genomics. *Molecular Ecology*, *25*, 185–202. https://doi.org/10.1111/mec.13304

Kaliberdina, M. V., & Granovich, A. I. (2003). Infection of the mollusc *Littorina saxatilis* with parthenites of trematodes and their impact on a shell form: Analysis of populations inhabiting the littoral coast of the White Sea. *Parazitologiia*, *37*(1), 69–86.

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., Von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), 587–589.

Kantor, Y., Sirenko, B., Zvonareva, S. S., & Fedosov, A. (2022). Taxonomic status of genera of Buccininae (Neogastropoda, Buccinidae) updated based on molecular data with description of new species and corrections of nomenclature of *Buccinum*. *European Journal of Taxonomy*, *817*, 11–34. https://doi.org/10.5852/ejt.2022.817.1759

Kantor, Y. I., Fedosov, A. E., Puillandre, N., Bonillo, C., & Bouchet, P. (2017). Returning to the roots: Morphology, molecular phylogeny and classification of the Olivoidea (Gastropoda: Neogastropoda). *Zoological Journal of the Linnean Society*, *180*(3), 493–541. https://doi.org/10.1093/zoolinnean/zlw003

Katoh, K., Misawa, K., Kuma, K. I., & Miyata, T. (2002). MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, *30*(14), 3059–3066. https://doi.org/10.1093/nar/gkf436

Kent, W. J. (2002). BLAT—The BLAST-like alignment tool. *Genome Research*, *12*(4), 656–664.

Kilburn, R. N. (1985). The family Epitoniidae (Mollusca: Gastropoda) in southern Africa and Mozambique. *Annals of the Natal Museum*, *27*(1), 239–337.

Kocot, K. M., Cannon, J. T., Todt, C., Citarella, M. R., Kohn, A. B., Meyer, A., Santos, S. R., Schander, C., Moroz, L. L., Lieb, B., & Halanych, K. M. (2011). Phylogenomics reveals deep molluscan relationships. *Nature*, *477*(7365), 452–456. https://doi.org/10.1038/nature10382

Kocot, K. M., Halanych, K. M., & Krug, P. J. (2013). Phylogenomics supports Panpulmonata: Opisthobranch paraphyly and key evolutionary steps in a major radiation of gastropod molluscs. *Molecular Phylogenetics and Evolution*, *69*(3), 764–771. https://doi.org/10.1016/j.ympev.2013.07.001

Kocot, K. M., Todt, C., Mikkelsen, N. T., & Halanych, K. M. (2019). Phylogenomics of Aplacophora (Mollusca, Aculifera) and a solenogaster without a foot. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1902), 20190115. https://doi.org/10.1098/rspb.2019.0115

Krug, P. J., Caplins, S. A., Algoso, K., Thomas, K., Valdés, Á. A., Wade, R., Wong, N. L. W. S., Eernisse, D. J., & Kocot, K. M. (2022). Phylogenomic resolution of the root of Panpulmonata, a hyperdiverse radiation of gastropods: New insight into the evolution of air breathing. *Proceedings of the Royal Society B: Biological Sciences*, *289*(1972), 20211855. https://doi.org/10.1098/rspb.2021.1855

Layton, K. K. S., Carvajal, J. I., & Wilson, N. G. (2020). Mimicry and mitonuclear discordance in nudibranchs: New insights from exon capture phylogenomics. *Ecology and Evolution*, *10*(21), 11966–11982. https://doi.org/10.1002/ece3.6727

Lee, S. Y., Lee, S. M., & Kim, Y. K. (2019). First de-novo transcriptome assembly, annotation and expression profiles of a freshwater snail (*Semisulcospira coreana*) fed with chlorella supplement. *Marine Genomics*, *47*, 100657. https://doi.org/10.1016/j.margen.2019.01.006

Lemer, S., Bieler, R., & Giribet, G. (2019). Resolving the relationships of clams and cockles: Dense transcriptome sampling drastically improves the bivalve tree of life. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1896), 20182684. https://doi.org/10.1098/rspb.2018.2684

Lindgren, A. R., & Anderson, F. E. (2018). Assessing the utility of transcriptome data for inferring phylogenetic relationships among coleoid cephalopods. *Molecular Phylogenetics and Evolution*, *118*, 330–342. https://doi.org/10.1016/j.ympev.2017.10.004

Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., & Zdobnov, E. M. (2021). BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Molecular Biology and*

17550998, 2023, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1755-0998.13793 by IFREMER Centre Bretagne BLP, Wiley Online Library on [10/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Evolution, *38*(10), 4647–4654. https://doi.org/10.1093/molbev/msab199

McLay, T. G. B., Birch, J. L., Gunn, B. F., Ning, W., Tate, J. A., Nauheimer, L., Joyce, E. M., Simpson, L., Schmidt-Lebuhn, A. N., Baker, W. J., Forest, F., & Jackson, C. J. (2021). New targets acquired: Improving locus recovery from the Angiosperms353 probe set. *Applications in Plant Sciences*, *9*(7), 1–9. https://doi.org/10.1002/aps3.11420

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., 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. https://doi.org/10.1093/molbev/msaa015

Moles, J., & Giribet, G. (2021). A polyvalent and universal tool for genomic studies in gastropod molluscs (Heterobranchia). *Molecular Phylogenetics and Evolution*, *155*, 106996. https://doi.org/10.1016/j.ympev.2020.106996

MolluscaBase. https://www.molluscabase.org/

Ortiz-Sepulveda, C. M., Genete, M., Blassiau, C., Godé, C., Albrecht, C., Vekemans, X., & Van Bocxlaer, B. (2022). Target enrichment of long open reading frames and ultraconserved elements to link microevolution and macroevolution in non-model organisms. *Molecular Ecology Resources*, *23*, 659–679. https://doi.org/10.1111/1755-0998.13735

Osca, D., Templado, J., & Zardoya, R. (2015). Caenogastropod mitogenomics. *Molecular Phylogenetics and Evolution*, *93*, 118–128. https://doi.org/10.1016/j.ympev.2015.07.011

Pardos-Blas, J. R., Irisarri, I., Abalde, S., Afonso, C. M., Tenorio, M. J., & Zardoya, R. (2021). The genome of the venomous snail *Lautoconus ventricosus* sheds light on the origin of conotoxin diversity. *Gigascience*, *10*(5), giab037. https://doi.org/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., … Shi, Q. (2021). The first *Conus* genome assembly reveals a primary genetic central dogma of conopeptides in *C. betulinus*. *Cell Discovery*, *7*(1), 11. https://doi.org/10.1038/s41421-021-00244-7

Pfeiffer, J. M., Breinholt, J. W., & Page, L. M. (2019). Unioverse: A phylogenomic resource for reconstructing the evolution of freshwater mussels (Bivalvia, Unionoida). *Molecular Phylogenetics and Evolution*, *137*(February), 114–126. https://doi.org/10.1016/j.ympev.2019.02.016

Ponder, W. F., Colgan, D. J., Healy, J. M., Nützel, A., Simone, L. R. L., & Strong, E. E. (2008). Caenogastropoda. In W. Ponder & D. R. Lindberg (Eds.), *Phylogeny and evolution of the Mollusca* (pp. 331–383).University of California Press.

Ponder, W. F., Lindberg, D. R., & Ponder, J. M. (2020). Gastropoda II – The Caenogastropoda. In *Biology and evolution of the Mollusca*, Vol. 2 (pp. 365–417). CRC Press.

Puillandre, N., Samadi, S., Boisselier, M. C., Sysoev, A. V., Kantor, Y. I., Cruaud, C., Couloux, A., & Bouchet, P. (2008). Starting to unravel the toxoglossan knot: Molecular phylogeny of the "turrids"(Neogastropoda: Conoidea). *Molecular Phylogenetics and Evolution*, *47*(3), 1122–1134. https://doi.org/10.1016/j.ympev.2007.11.007

Quattrini, A. M., Faircloth, B. C., Dueñas, L. F., Bridge, T. C. L., Brugler, M. R., Calixto-Botía, I. F., DeLeo, D. M., Forêt, S., Herrera, S., Lee, S. M. Y., Miller, D. J., Prada, C., Rádis-Baptista, G., Ramírez-Portilla, C., Sánchez, J. A., Rodríguez, E., & McFadden, C. S. (2018). Universal target-enrichment baits for anthozoan (Cnidaria) phylogenomics: New approaches to long-standing problems. *Molecular Ecology Resources*, *18*(2), 281–295. https://doi.org/10.1111/1755-0998.12736

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842.

Schell, T., Feldmeyer, B., Schmidt, H., Greshake, B., Tills, O., Truebano, M., Rundle, S. D., Paule, J., Ebersberger, I., & Pfenninger, M. (2017). An annotated draft genome for *Radix auricularia* (Gastropoda, Mollusca). *Genome Biology and Evolution*, *9*(3), 585–592. https://doi.org/10.1093/gbe/evx032

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Smit, A. F. A., Hubley, R., & Green, P. (2015). *RepeatMasker Open-4.0*.

Smith, B. T., Harvey, M. G., Faircloth, B. C., Glenn, T. C., & Brumfield, R. T. (2014). Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. *Systematic Biology*, *63*(1), 83–95. https://doi.org/10.1093/sysbio/syt061

Smith, S. A., Wilson, N. G., Goetz, F. E., Feehery, C., Andrade, S. C. S., Rouse, G. W., Giribet, G., & Dunn, C. W. (2011). Resolving the evolutionary relationships of molluscs with phylogenomic tools. *Nature*, *480*(7377), 364–367. https://doi.org/10.1038/nature10526

Sohl, N. F. (1964). Gastropods from the coffee sand (Upper Cretaceous) of Mississippi. *U.S. Geological Survey Professional Paper*, *331–C*, 385–394.

Song, H., Yang, M., Yu, Z., & Zhang, T. (2019). Characterization of the whole transcriptome of whelk *Rapana venosa* by single-molecule mRNA sequencing. *Marine Genomics*, *44*, 74–77.

Sproul, J. S., & Maddison, D. R. (2017). Sequencing historical specimens: Successful preparation of small specimens with low amounts of degraded DNA. *Molecular Ecology Resources*, *17*, 1183–1201. https://doi.org/10.1111/1755-0998.12660

Strong, E. E., Colgan, D. J., Healy, J. M., Lydeard, C., Ponder, W. F., & Glaubrecht, M. (2011). Phylogeny of the gastropod superfamily Cerithioidea using morphology and molecules. *Zoological Journal of the Linnean Society*, *162*(1), 43–89. https://doi.org/10.1111/j.1096-3642.2010.00670.x

Strong, E. E., Puillandre, N., Beu, A. G., Castelin, M., & Bouchet, P. (2019). Frogs and tuns and tritons - a molecular phylogeny and revised family classification of the predatory gastropod superfamily Tonnoidea (Caenogastropoda). *Molecular Phylogenetics and Evolution*, *130*, 18–34. https://doi.org/10.1016/j.ympev.2018.09.016

Sun, J., Mu, H., Ip, J. C., Li, R., Xu, T., Accorsi, A., Sánchez Alvarado, A., Ross, E., Lan, Y., Sun, Y., Castro-Vazquez, A., Vega, I. A., Heras, H., Ituarte, S., Van Bocxlaer, B., Hayes, K. A., Cowie, R. H., Zhao, Z., Zhang, Y., … Qiu, J. W. (2019). Signatures of divergence, invasiveness, and terrestrialization revealed by four apple snail genomes. *Molecular Biology and Evolution*, *36*(7), 1507–1520. https://doi.org/10.1093/molbev/msz084

Takano, T., & Kano, Y. (2014). Molecular phylogenetic investigations of the relationships of the echinoderm-parasite family Eulimidae within Hypsogastropoda (Mollusca). *Molecular Phylogenetics and Evolution*, *79*(1), 258–269. https://doi.org/10.1016/j.ympev.2014.06.021

Takano, T., Warén, A., & Kano, Y. (2022). Phylogenetic position of the deep-sea snail family Haloceratidae and new insights into caenogastropod relationships. *Journal of Molluscan Studies*, *88*(2), eyac012. https://doi.org/10.1093/mollus/eyac012

Untiedt, C. B., Quattrini, A. M., McFadden, C. S., Alderslade, P. A., Pante, E., & Burridge, C. P. (2021). Phylogenetic relationships within Chrysogorgia (Alcyonacea: Octocorallia), a morphologically diverse genus of Octocoral, revealed using a target enrichment approach. *Frontiers in Marine Science*, *7*, 1–14. https://doi.org/10.3389/fmars.2020.599984

Uribe, J. E., Gonzalez, V. L., Kano, Y., Herbert, D., Irisarri, I., Strong, E. E., & Harasewych, M. G. (2022). A phylogenomic backbone for gastropod molluscs. *Systematic Biology*, *syac045*, 1–18. https://doi.org/10.1093/sysbio/syac045

Wall-Palmer, D., Mekkes, L., Ramos-Silva, P., Dämmer, L. K., Goetze, E., Bakker, K., Duijm, E., & Peijnenburg, K. T. (2021). The impacts of past, present and future ocean chemistry on predatory

17550998, 2023, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1755-0998.13793 by IFREMER Centre Bretagne BLP, Wiley Online Library on [10/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

planktonic snails. *Royal Society Open Science*, *8*(8), 202265. https://doi.org/10.1098/rsos.202265

Wang, Q., Liu, H., Yue, C., Xie, X., Li, D., Liang, M., & Li, Q. (2021). Characterization of the complete mitochondrial genome of *Ficus variegata* (Littorinimorpha: Ficidae) and molecular phylogeny of Caenogastropoda. *Mitochondrial DNA Part B*, *6*(3), 1126–1128. https://doi.org/10.1080/23802359.2021.1901628

Whelan, N. V., Johnson, P. D., Garner, J. T., Garrison, N. L., & Strong, E. E. (2022). Prodigious polyphyly in Pleuroceridae (Gastropoda: Cerithioidea). *Bulletin of the Society of Systematic Biologists*, *1*(2).

Wilke, T., Haase, M., Hershler, R., Liu, H. P., Misof, B., & Ponder, W. (2013). Pushing short DNA fragments to the limit: Phylogenetic relationships of 'hydrobioid' gastropods (Caenogastropoda: Rissooidea). *Molecular Phylogenetics and Evolution*, *66*(3), 715–736. https://doi.org/10.1016/j.ympev.2012.10.025

Yang, Y., Sun, J., Chen, C., Zhou, Y., Van Dover, C. L., Wang, C., Qiu, J. W., & Qian, P. Y. (2022). Metagenomic and metatranscriptomic analyses reveal minor-yet-crucial roles of gut microbiome in deep-sea hydrothermal vent snail. *Animal Microbiome*, *4*(1), 1–17. https://doi.org/10.1186/s42523-021-00150-z

Yang, Z., Zhang, L., Hu, J., Wang, J., Bao, Z., & Wang, S. (2020). The evo-devo of molluscs: Insights from a genomic perspective. *Evolution and Development*, *22*(6), 409–424. https://doi.org/10.1111/ede.12336

Zapata, F., Wilson, N. G., Howison, M., Andrade, S. C. S., Jörger, K. M., Schrödl, M., Goetz, F. E., Giribet, G., & Dunn, C. W. (2014). Phylogenomic analyses of deep gastropod relationships reject Orthogastropoda. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1794), 20141739. https://doi.org/10.1098/rspb.2014.1739

Zeng, X., Zhang, Y., Meng, L., Fan, G., Bai, J., Chen, J., Song, Y., Seim, I., Wang, C., Shao, Z., Liu, N., Lu, H., Fu, X., Wang, L., Liu, X., Liu, S., & Shao, Z. (2020). Genome sequencing of deep-sea hydrothermal vent snails reveals adaptions to extreme environments. *GigaScience*, *9*(12), giaa139. https://doi.org/10.1093/gigascience/giaa139

How to cite this article: Goulding, T. C., Strong, E. E., & Quattrini, A. M. (2023). Target-capture probes for phylogenomics of the Caenogastropoda. *Molecular Ecology Resources*, *23*, 1372–1388. https://doi.org/10.1111/1755-0998.13793

17550998, 2023, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1755-0998.13793 by IFREMER Centre Bretagne BLP, Wiley Online Library on [10/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License