A polyvalent and universal tool for genomic studies in gastropod molluscs (Heterobranchia: Tectipleura)

Juan Moles[1] and Gonzalo Giribet[1]

[1]Harvard University Faculty of Arts and Sciences

April 28, 2020

Abstract

Molluscs are the second most diverse animal phylum and heterobranch gastropods present ˜44,000 species. These comprise fascinating creatures with a huge morphological and ecological disparity. Such great diversity comes with even larger phylogenetic uncertainty and many taxa have been largely neglected in molecular assessments. Genomic tools have provided resolution to deep cladogenic events but generating large numbers of transcriptomes/genomes is expensive and usually requires fresh material. Here we leverage a target enrichment approach to design and synthesize a probe set based on available genomes and transcriptomes across Heterobranchia. Our probe set contains 57,606 70mer baits and targets a total of 2,259 ultra-conserved elements (UCEs). Post-sequencing capture efficiency was tested against 31 marine heterobranchs from major groups, including Acochlidia, Acteonoidea, Aplysiida, Cephalaspidea, Pleurobranchida, Pteropoda, Runcinida, Sacoglossa, and Umbraculida. The combined Trinity and Velvet assemblies recovered up to 2,211 UCEs in Tectipleura and up to 1,978 in Nudipleura, the most distantly related taxon to our core study group. Total alignment length was 525,599 bp and contained 52% informative sites and 21% missing data. Maximum-likelihood and Bayesian inference approaches recovered the monophyly of all orders tested as well as the larger clades Nudipleura, Panpulmonata, and Euopisthobranchia. The successful enrichment of diversely preserved material and DNA concentrations demonstrate the polyvalent nature of UCEs, and the universality of the probe set designed. We believe this probe set will enable multiple, interesting lines of research, that will benefit from an inexpensive and largely informative tool that will, additionally, benefit from the access to museum collections to gather genomic data.

Introduction

Molecular data have played an important role in elucidating molluscan relationships in general and gastropod systematics in particular, to the point the classification of many higher clades is now dominated by molecular-based estimations of phylogeny. Early work on molecular systematics of molluscs started, as in the case of many other animal groups, using a series of markers amplified by PCR, which in another context have been dubbed the "workhorses" of molecular systematics (Sharma & Giribet, 2009) or the "usual suspects" (Dimitrov et al., 2017). From the early days of molecular systematics of molluscs (e.g., Giribet et al., 2006; Passamaneck, Schander, & Halanych, 2004; Winnepenninckx, Backeljau, & De Wachter, 1996), including the early days of gastropod molecular trees (e.g., Colgan, Ponder, Beacham, & Macaranas, 2007, 2003; Colgan, Ponder, & Eggler, 2000; Harasewych et al., 1997; Harasewych, Adamkewicz, Plassmeyer, & Gillevet, 1998; Harasewych & McArthur, 2000; McArthur & Harasewych, 2003), these markers (mainly nuclear ribosomal RNAs, nuclear protein-encoding histone H3, and mitochondrial 16S rRNA and cytochrome *c* oxidase subunit I) have helped to shape the gastropod tree. A first paradigm shift occurred with the generalized used of ESTs (e.g., Dunn et al., 2008) and later, large numbers of transcriptomes soon started accumulating for molluscs (Cunha & Giribet, 2019; González et al., 2015; Kocot et al., 2011; Kocot, Halanych, & Krug, 2013; Kocot, Poustka, Stöger, Halanych, & Schrödl, 2020; Kocot, Todt, Mikkelsen, & Halanych, 2019; Lemer, Bieler, & Giribet, 2019; Lemer, González, Bieler, & Giribet, 2016; Lindgren & Anderson, 2018; Pabst & Kocot, 2018; Smith et al., 2011; Tanner et al., 2017; Zapata et al., 2014). These datasets provided resolution to deep nodes,

previously poorly supported using the standard Sanger markers. But transcriptomes are difficult to obtain for large numbers of taxa, as they require fresh tissue and special preservation to avoid RNA degradation, and are expensive to generate (Zaharias, Pante, Gey, Fedosov, & Puillandre, 2020). A third strategy, able to make use of DNA available from many museum specimens, but avoiding PCR-amplification, are methods based on large numbers of hybridizing probes sequenced with high-throughput techniques, i.e., Illumina sequencing (Crawford et al., 2012; Faircloth et al., 2012; Lemmon, Emme, & Lemmon, 2012; Lemmon & Lemmon, 2012; McCormack et al., 2012). These bait-capture techniques have been recently applied to study gastropod phylogenetics (Abdelkrim et al., 2018; Choo et al., 2020; Zaharias et al., 2020), but available baits have been designed specifically for a genus of Pteropoda and a subset of Neogastropoda. These latter studies were able to include "museum samples" (specimens not collected and preserved for molecular work) (Abdelkrim et al., 2018), and more recently these techniques have been applied to ethanol-preserved specimens older than 100 years (Derkarabetian, Benavides, & Giribet, 2019). Therefore, to capitalize on available museum samples and to be able to sequence thousands of loci across heterobranch gastropods—the most diverse subclass of gastropods, with ca. 44,000 living species (Barker, 2001; Lydeard & Cummings, 2019; WoRMS Editorial Board, 2020)—we have designed a new set of probes for ultra-conserved elements (UCEs) with the aim to apply it to future studies across heterobranch taxa.

Heterobranchs embody a diverse and charismatic group of marine, limnic, and terrestrial snails and slugs with a plethora of ecological and morphological adaptations to all environments, e.g., pelagic (Klussmann-Kolb & Dinapoli, 2006), abyssal (Chaban et al., 2019), meiofaunal (Jörger et al., 2010), parasitic (Dinapoli, Zinssmeister, & Klussmann-Kolb, 2011). They represent the cornerstone of interesting lines of research including chemical ecology and pharmaceutical applications (reviewed in Avila, Núñez-Pons, & Moles, 2018), solar-powered slugs are among the only Metazoa able to incorporate chloroplasts from dietary algae which remain photosynthetically active in their tissues (i.e., kleptoplasty; Wägele et al., 2011), the giant neurones of e.g. *Aplysia* became a key model in neurobiology studies (e.g., Kandel, 1979), some terrestrial snails and slugs are detrimental pests or vectors of snail-borne human parasitic diseases such as angiostrongyliasis, bilharzia or liver rot (Lu et al., 2018), also, many species are indicators of ecosystem wellbeing and climate change (Keul et al., 2017). Still, understanding the evolutionary history of Heterobranchia has been difficult (reviewed in Wägele, Klussmann-Kolb, Verbeek, & Schrödl, 2014), even when the monophyly of the group has been well established.

Although traditionally divided into two large gastropod subclasses, Heterobranchia now includes the polyphyletic Opisthobranchia and the paraphyletic Pulmonata plus some other shelled 'prosobranch' lineages (Schrödl, Jörger, Klussmann-Kolb, & Wilson, 2011). Among the morphological traits that define the taxon, hermaphroditism, a gill of heterogeneous nature, a heterostrophic protoconch, spiral-shaped sperm, and a pallial kidney are shared characteristics (Brenzinger, Haszprunar, & Schrödl, 2013; Haszprunar, 1985; Wägele et al., 2014). Although the monophyly of most major taxa is well supported, some of their interrelationships among and within subgroups remain controversial. For instance, the uncertain systematic placement of some obscure lineages of 'lower heterobranchs', such as Acteonoidea, Rissoelloidea or Rhodopemorpha, that lack an euthyneurous (i.e. detorted) nervous system (Brenzinger et al., 2013; Wägele et al., 2014), awaits resolution. Among Euthyneura, two major clades are accepted: Tectipleura (Panpulmonata + Euopisthobranchia) and Ringipleura (Ringiculoidea + Nudipleura) (Kano, Brenzinger, Nützel, Wilson, & Schrödl, 2016). Panpulmonata includes land snails and slugs (Stylommatophora), many limnic (e.g. Hygrophila), marine intertidal (Siphonarioidea), marine interstitial (Acochlidia), and marine ectoparasitic lineages (Pyramidelloidea), as well as the marine solar-powered slugs (Sacoglossa) (Jörger et al., 2010; Kano et al., 2016; Kocot et al., 2013; Zapata et al., 2014). Euopisthobranchia comprises sea hares (Aplysiida), pelagic sea angels (Pteropoda), bubble snails *sensu lato* (Cephalaspidea), false limpets (Umbraculida), and Runcinida (Jörger et al., 2010; Kano et al., 2016; Zapata et al., 2014). Nudipleura includes the side-gilled slugs (Pleurobranchida) and the colourful sea slugs (Nudibranchia) (Kano et al., 2016; Pabst & Kocot, 2018; Wägele & Willan, 2000; Zapata et al., 2014). The inclusion of novel genomic approaches to better reconstruct the evolutionary history of Heterobranchia from high-ranking to the species level remains crucial (Cunha & Giribet, 2019; Goodheart, Bazinet, Collins, & Cummings, 2015; Kocot et al., 2013; Pabst & Kocot, 2018; Peijnenburg et al., 2019; Za-

pata et al., 2014). Moreover, the possibility to provide molecular evidence from museum-preserved specimens thanks to UCEs will render elusive taxa and/or type material available for study. Hence, a new systematic framework may provide input on the mode and tempo to interesting ecological questions such as the reduction or loss of the shell across Heterobranchia (Medina et al., 2011; Wägele & Klussmann-Kolb, 2005), the acquisition of defensive mechanisms alternative to the shell (Avila et al., 2018; Vonnemann, Schrödl, Klussmann-Kolb, & Wägele, 2005), the adaptation to freshwater and terrestrial habitats (Klussmann-Kolb, Dinapoli, Kuhn, Streit, & Albrecht, 2008; Neusser, Jörger, Lodde-Bensch, Strong, & Schrödl, 2016), morpho-anatomical transitions and adaptations (Brenzinger et al., 2013; Kano et al., 2016) or evolutionary dietary patterns (Goodheart, Bazinet, Valdés, Collins, & Cummings, 2017; Malaquias, Berecibar, & Reid, 2009), among many other hot topics on this hyperdiverse group of molluscs.

Material and Methods

Taxon sampling

Genomes of selected gastropods, including the aplysiid *Aplysia californica* (base genome), the sacoglossan *Elysia chlorotica* , the hygrophilid *Radix auricularia* , and the caenogastropod*Pomacea canaliculata* (used as outgroup), and transcriptomes of the cephalaspidean *Haminoea antillarum* and the umbraculid*Tylodina fungina* were downloaded from NCBI. Additional transcriptomes across Heterobranchia, including Acochlidia, Acteonida, Architectonicoidea, Ellobiida, Nudibranchia, Pleurobranchida, Pteropoda, Pylopulmonata, and Rissoelloidea, were downloaded for the *in silico* test of the bait set. Finally, two Caenogastropoda were downloaded for matrix construction and for rooting the phylogenetic trees (Table 1, S1).

Specimens for the *in vitro* test of the bait set were obtained from the Museum of Comparative Zoology (MCZ) and the University Museum of Bergen (Department of Natural History; ZMBN, Norway) (Table 2). Most of the material was freshly collected from the Maldives in 2019, preserved at 95% EtOH and kept at –20 °C. The remaining samples were gathered from museum collections, mostly preserved in 70% EtOH and kept at room temperature, sometimes for several years, and are thus considered 'degraded' samples.

Bait set design and synthesis

All downloaded transcriptomes were assembled *de novo* using the pipeline from Cunha & Giribet (2019). Briefly, quality threshold filtering was conducted with Rcorrector v. 3.0 (Song & Florea, 2015) and Trim Galore! V. 3 (*http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/*). rRNA and mitochondrial unwanted sequences for molluscs were filtered out using Bowtie2 v. 2.3.2 (Langmead & Salzberg, 2012). Paired-end reads were *de novo* assembled into transcripts with Trinity v. 2.4 (Grabherr et al., 2010; Haas et al., 2013). A second Bowtie2 round and CD-HIT-EST v. 4.6.4 were used to reduce sequence redundancy (Fu, Niu, Zhu, Wu, & Li, 2012). The software PHYLUCE (Faircloth, 2016) was then used to identify UCE loci and design baits to target them using the online tutorial (*https://phyluce.readthedocs.io/en/latest/tutorial-four.html*). Downloaded FASTA files for the selected genomes/transcriptomes were reformatted into 2bit using faToTwoBit and headers were modified using Bio.SeqIO (Grüning et al., 2018) for compatibility with PHYLUCE.ART (Huang, Li, Myers, & Marth, 2012), which was used to simulate reads of 100 bp in length, covering the genome randomly to roughly 2X, and having an insert size of 200 bp (150 SD), for each species. These were individually mapped into putatively orthologous loci with a sequence divergence of $< 5\%$ from our base genome (*A. californica*) using stampy v. 1 (Lunter & Goodson, 2011) and unmapped reads were removed with SAMtools v. 1.5 (Li et al., 2009). BEDTools (Quinlan & Hall, 2010) was used to convert BAM files, sort the contigs by scaffold and position, and merge them in putative conserved regions. Intervals where the base genome was shorter than 80 bp and where > 25 % of the base genome was masked (i.e., repetitive regions) were deleted in PHYLUCE.

An SQLite table was created to query for conserved loci across taxa with an optimal number of four out of five taxa, resulting in a total of 7,222 shared loci. Temporary baits were designed to capture loci shared among the base genome and the exemplar taxa, buffering to 160 bp to ensure designing two 120mers per locus with 3x tiling density, removing potentially problematic baits with >25% masking and GC content outside of a 30–70% range. Finally, potential duplicates of >50% identity and coverage were parsed and

removed. In order to include baits designed from the base genome and the exemplar taxa, the temporary baits were also aligned against all five exemplar taxa and conserved loci were extracted as FASTA files. An additional SQLite table was created to check for the loci found consistently across taxa.

We finally decided to target loci that were shared among five out of the six taxa, totalling 2,320. Final bait design was performed using the abovementioned steps but using both the base genome and the rest of the exemplar taxa. A subset of locus bait set targeting only the specific heterobranch species (excluding the caenogastropod *P. canaliculata*) was designed using phyluce_probe_get_subsets_of_tiled_probes. The final set contained 19,333 baits and targeted 2,259 loci across Tectipleura. *In silico* tests of the UCE bait set against *de novo* assembled transcriptomes were performed against a wide range of taxa belonging to Heterobranchia and two Caenogastropoda outgroups using*phyluce_assembly_match_contigs_to_probes* (see Table 1).

In order to synthesize the designed bait set, each bait candidate was BLASTed against the base genome in order to filter non-specific or over-represented regions and a hybridization melting temperature (defined as the temperature at which 50% of molecules are hybridized) was estimated for each hit assuming standard myBaits® (Arbor Biosciences, MI, USA) buffers and conditions. There were 129 baits that matched a portion of the genome that was >25% soft masked for repeats and 5 baits failed out Moderate BLAST analysis (candidates pass if they have at most 10 hits at 62.5–65 °C and 2 hits above 65 °C, and fewer than 2 passing baits on each flank), indicating they had multiple hits to the genome at the hybridization temperature and, thus, were removed. Due to technical difficulties in synthesizing the 120mer set, three overlapping 70mers for each 120mer were designed (1 bait every 25 nt), both providing the same coverage of the original design targets with the same capture efficiency as the 120mer design. Interestingly, shorter fragments have the ability to be used at a range of hybridization temperatures, thus, being more effective on degraded museum samples. The total design had 57,606 baits (out of the original 19,202 120mer set).

Molecular data collection

Genomic DNA from freshly preserved material was extracted from the parapodia and/or foot using the DNeasy Blood and Tissue kit (Qiagen Inc., Valencia, CA, USA) following the manufacturer's protocol. In order to obtain a higher DNA yield for millimetric specimens and a non-destructive DNA extraction of mostly 'degraded' whole museum individuals we used the protocol described in Tin, Economo & Mikheyev (2014) using silica-based beads, with some in-lab modifications (Derkarabetian et al., 2019). Extractions were quantified using the Qubit 2.0 fluorometer (Life Technologies, Inc.) dsDNA High Sensitivity kit and visualized on a 2200 Tape Station (Agilent) to assess DNA degradation. Up to 500 ng of DNA in 130 µL for all fresh specimens was sonicated for 80 s with a Covaris S220 Focused-ultrasonicator for a target peak of 500–600 bp, with a Peak Incidence Power of 50, Duty Factor of 10%, and 200 cycles per burst. Sonication time was adjusted depending on the Tape Station results for each sample since museum samples were naturally degraded to the appropriate size for sequence capture library preparation (see Table 2).

Library preparation followed the general protocol on the UCE website (https://www.ultraconserved.org/) and some in-lab modifications (Derkarabetian et al., 2018). Libraries were prepared using the KAPA Hyper Prep kit following the manufacturer's protocol and using up to 250 ng of DNA as starting material. For samples with lower DNA yield, we used it all (down to 5 ng). We used fresh Serapure Speed-beads for all clean-up steps (Rohland & Reich, 2012) washing with freshly prepared 80% EtOH. After the first clean-up, 25 µL of fragmented, double-stranded DNA was assembled for end repair and A-tailing, 20 °C for 30 min and 65 °C for 30 min, respectively. Immediately afterward we proceeded to adapter ligation, samples with >200 ng of DNA were run at 20 °C for 30 min using the universal iTru Stubs at 10 µM, while samples <200 ng were run up to 1h using 5 µM iTru Stubs. An immediate post-ligation clean-up step was carried at 0.8X for high-yield, fresh samples, and at up to 3X for degraded, low-yield samples. Library amplification was conducted using 25 µL of post-ligated libraries, using individual iTrue i5/i7 dual index adaptors (8 bp long ; Glenn et al., 2016), with an initial denaturation step at 98 °C for 45 s, then 6–14 PCR cycles (adjusted to post-ligated Qubit concentrations) of 98 °C for 15 s, 60 °C for 30 s and 72 °C for 1 min, and a final extension step of 72 °C for 5 min. Libraries were then washed, quantified, and 125 ng of each were pooled in batches of eight samples for a total of 1000 ng. Pools were then speed-vacuumed if necessary, to a final volume of 14

μL.

Hybridization capture for targeted UCEs was carried following the myBaits® v.4 user's manual and the target enrichment for Illumina standard workflow protocols (*https://www.ultraconserved.org/#protocols*). Pooled libraries were hybridized with the synthesized bait set at 60 °C for 24 h. Hybrid pools were then bound to streptavidin beads (Dynabeads MyOne C1, Invitrogen), washed, and eluted in 31 μL of NF Water. Post-hybridization amplification was carried in a 50 μL reaction using 15 μL of hybridized pools, with same post-ligation PCR conditions but for a total of 16 cycles. Immediately followed a bead clean up with 70% EtOH, Qubit 2.0 quantification, and visualization and molarity calculation using a 2200 Tape Station. Post-hybridized pools were pooled in equimolar amounts and sequenced in the Illumina NovaSeq 6000 SP platform (paired-end, 150 bp) at the Bauer Core Facility, Harvard University. New sequences were deposited in the NCBI Sequence Read Archive (BioProject PRJNA612319); voucher information and assembly statistics are available in Table 2 and S2.

Species and matrix assembly

Raw reads were demultiplexed per individual and the software PHYLUCE v 1.6.8 (Faircloth, 2016) was used for each species assembly. Raw reads were trimmed for adapter contamination and low-quality bases with trimmomatic v 0.39 (Bolger, Lohse, & Usadel, 2014) implemented in illumiprocessor v 2.0.9 (Faircloth, 2013). We then run cd-hit-dup v 4.6.4 for removing duplicates from sequencing reads (Fu et al., 2012). We used the three assemblers: ABySS v 2.0 (Jackman et al., 2017), Trinity v. 2.1.1 (Haas et al., 2013), and Velvet v 1.2.10 (Zerbino & Birney, 2008) for comparative purposes (see Table S2). In order to recover a higher number of contigs, we combined both Velvet and Trinity assemblies and matched the contigs to the probe set with the modified version of the original script by B. Faircloth*phyluce_assembly_match_contigs_to_probes_duplicates* with*–min-coverage* and *–min-identity* of 80.

Assembly QC statistics are shown in Table 3. Contigs that represent the targeted UCE loci were captured and duplicated loci –either different probes hitting the same locus or a probe hitting multiple loci– were removed. A list of the enriched UCE loci in each taxon, including incomplete loci (not found in all the taxon set), was generated and individual FASTA files for these were extracted (Table 2; see Table S2 for summary statistics for each species using each of the three assemblers). In this step, we also included the UCE loci captured for all genomes and transcriptome assemblies used for the probe set design and two transcriptomes from caenogastropods. Selected UCE loci were aligned in MAFFT v 7.455 (Katoh & Standley, 2013). Edge and internal trimming of the resulting alignments was conducted using Gblocks v 0.91 (Castresana, 2000; Talavera & Castresana, 2007), specifying mid-level arguments, ideal for higher-level taxonomic ranking phylogenies, i.e.*–b1* 0.5, *–b2* 0.5, *–b3* 6, *–b4* 4. A matrix was then built using a percentage of completeness of 50% (see Table 3 for the number of genes per species in the final datasets).

Phylogenetic analyses

Phylogenetic trees were reconstructed using maximum likelihood (ML) implemented in IQ-TREE v. 1.5.5 (Nguyen, Schmidt, Von Haeseler, & Minh, 2015) under the general time-reversible (Tavaré, 1986), four class mixture model (GTR+FO*H4) on the unpartitioned matrix. This is a General Heterogeneous evolution On a Single Topology model (GHOST) that infers separate base frequencies per class and accounts for heterotachy or rate variation across sites and lineages (Crotty et al., 2020). Bootstrap support values (bs) were estimated via the ultrafast bootstrap algorithm with 1500 replicates (Minh, Nguyen, & Haeseler, 2013). Bayesian inference (BI) was assessed in ExaBayes v 1.5 (Aberer, Kobert, & Stamatakis, 2014) with the implemented GTRGAMMA model with four coupled Markov chain Monte Carlo (MCMC) runs, each with 10 million generations, and sampling every 500 generations. Convergence was checked based on the average standard deviation of split frequencies (ASDSF <0.2%). The first 25% of the trees were discarded as burn-in for each MCMC run prior to convergence. Topological robustness was assessed using posterior probabilities (pp). Trees were visualized in FigTree v. 1.4.4 (Rambaut, 2014) and edited in Adobe Illustrator CC 2018 (Fig. 1).

Results & Discussion

Here we provide a relatively inexpensive and universal probe set for sequencing ca. two thousand loci across Heterobranchia, using a particular subset of genomes and transcriptomes of relevant clades within Tectipleura (Euopisthobranchia and Panpulmonata). The final bait set contains 57,606 70mer baits, overlapping threefold with the originally designed 19,333 120mer baits, with the advantage of better capturing degraded DNA from museum material. This was designed to target a total of 2,259 loci across some of the major clades of Tectipleura, i.e. a sea hare (Aplysiida), a bubble snail (Cephalaspidea), a false limpet (Umbraculida), a hygrophil limnic snail (Hygrophila), and a solar-powered slug (Sacoglossa). Prior to synthesis, the bait set was tested *in silico* against 30 Trinity *de novo* transcriptome assemblies of multiple lineages across Heterobranchia, with a number of captured UCEs of 136 to 966 (6–43%). Congruently, the number of loci captured increased when approaching our core taxon Tectipleura (Table 1).

For testing the efficiency of the designed probe set, 31 taxa across the major groups of marine Heterobranchia were sequenced. Samples were collected during 2001–2019 and were either preserved in 70% or 95% EtOH. Regular Qiagen or the 'degraded' museum specimens' extractions for millimetric specimens were carried out. DNA concentrations ranged from undetectable (<0.01 μg/mL) to very high (>600 μg/mL) and the Covaris sonication shearing time was adjusted on samples with degraded DNA (Table 2). Also, additional museum samples older than 40 years, preserved in 70% EtOH, and kept at room temperature, yielded >1,000 UCEs (auth. unpubl. data). The total number of Illumina raw reads obtained ranged from 4 to 32 million but in the species *Creseis acicula* and *Pontohedyle milaschewitchii* the number of reads decreased substantially after trimming. These were excluded from the final matrix construction. In order to increase the yield of captured UCEs, both Trinity and Velvet assemblies were combined since they recovered larger contig lengths than the ABySS assembly (Table S2). The genomes and transcriptomes of*Aplysia californica* , *Elysia chlorotica* , *Radix auricularia* , *Haminoea antillarum* , and *Tylodina fungina*used for the bait set design were included in the trees. Additionally,*Chraronia tritonis* and *Crepidula navicella* were used as outgroups. The final matrix contained 36 taxa and captured all our 2,259 originally targeted UCEs, with species of Aplysiida displaying a 92–98% efficiency in captured UCEs, 67–95% in Cephalaspidea, 93–95% in Umbraculida, 83–92% in Sacoglossa, and 35–72% in Runcinida (Table 3). Samples from the most distantly related groups to our core taxon captured 73–86% UCEs in Nudibranchia, up to 81% in Pleurobranchida, and 76–87% in Acteonoidea (Table 3). The final alignment of the concatenated data set contained 525,599 bp with a mean contig size per UCE of 233 ± 5 bp with 273,694 informative sites (52%) across all loci and less than 21% missing data (mostly found in the downloaded transcriptomes; Table 4). The final matrix contained 12,557,760 nucleotides out of 14,352,347 possible characters, not accounting for the sequences not represented in each alignment (less than 13% of missing data). The 50% occupancy matrix contained 2,156 loci. Overall, our probe set opens up to the possibility to access old museum collections of usually not-well preserved specimens for molecular analysis to be used in future phylogenetic assessments across Heterobranchia (see Derkarabetian et al., 2019). Type taxa and obscure lineages, sometimes seldom recorded again since their original description, could potentially be available for genomic studies henceforth.

Both BI and ML analyses recovered the monophyly (bs = 100, pp = 1.00) of the orders and superorders Pleurobranchida, Nudibranchia, Sacoglossa (bs = 100, pp = 0.94), Acteonoidea, Umbraculida, Aplysiida, Runcinida, and Cephalaspidea with maximum support (Fig. 1). The monophyly of Euopisthobranchia (Aplysiida + Runcinida + Umbraculida; bb = 100, pp = 1.00), Panpulmonata (Sacoglossa + Hygrophila; bb = 100, pp = 1.00), and Nudipleura (Nudibranchia + Pleurobranchida; bb = 100, pp = 1.00), as well as for our core group of study Tectipleura only in the BI (Euopisthobranchia + Panpulmonata; pp = 1.00). The phylogenetic position of Acteonoidea was ambiguous, placed as the sister group to Tectipleura in the ML analysis (bs = 98) or as the sister group to Euthyneura in the BI (pp = 1.00). Our results are congruent with previous studies using multilocus Sanger sequencing and transcriptomic assessments (Jörger et al., 2010; Kano et al., 2016; Schrödl et al., 2011; Zapata et al., 2014). We have proven the versatility of UCEs combined with transcriptomic and genomic available data and it is a matter of time that a comprehensive and thorough genomic dataset is amassed to better establish the evolutionary history of Heterobranchia and its interrelationships. Big questions remain to be answered, for instance, the position of the non-Euthyneura groups (Brenzinger et al., 2013), such as the controversies on the Acteonoidea. Sanger sequencing assessments have placed them as

the sister group to Nudipleura (i.e., Acteopleura; Medina et al., 2011), or in a stemward position the sister group to Euthyneura (Kano et al., 2016), the latter result also supported by anatomical evidence and our BI analysis. Transcriptomic data and our ML results recover Acteonoidea as the sister group to Tectipleura (Pabst & Kocot, 2018; Zapata et al., 2014), although sometimes with little or no support. The latter scenario could imply acteonoids having retained plesiomorphic anatomical characters such as the torted nervous system (i.e. streptoneury). Evidently, a comprehensive taxon sampling, including Rissoellida (+ Acteonidae =Acteonimorpha) among other taxa, will help clarify the disagreeing topologies recovered. The position of Runcinida as the sister group to Cephalaspidea has also been recently supported (Malaquias, Mackenzie-Dodds, Bouchet, Gosliner, & Reid, 2009; Oskars, Bouchet, & Malaquias, 2015) and here we recovered that same topology. Additional taxa remain to be included in order to resolve deeper relationships, such as Pteropoda and major Panpulmonata groups, i.e. Acochlidia, Amphiboloidea, Eupulmonata (Stylommatophora, Systellommatophora, Ellobioidea), Glacidorboidea, Hygrophila, Pyramidelloidea, and Siphonarioidea.

Due to the polyvalent nature of UCEs and the universality of the probe set designed we believe this study will lead to multiple, interesting lines of research in gastropod molluscs, not only resolving the phylogenetic conundrum between many large clades but also aiming at establishing the systematics within major subgroups, which still present many unresolved relationships at species, genus or family level (Carmona, Pola, Gosliner, & Cervera, 2013; Epstein, Hallas, Johnson, Lopez, & Gosliner, 2018; Goodheart, 2017; Korshunova et al., 2020; Krug, Vendetti, & Valdés, 2016; Moles, Avila, & Malaquias, 2019; Oskars et al., 2019; Padula et al., 2016).

Acknowledgements

We are indebted to the members of the Giribet lab, especially to Shahan Derkarabetian for his advice during lab work and UCEs assembly. We are also grateful to the Bauer Core and the RC Computing team. We are especially thankful to Tauana J. Cunha and the MaRHE Center staff for their help while collecting in the Maldives and for providing photographs of live specimens. We appreciate the samples donated from the GROC Association (*https://opistobranquis.org/en/home*) and M. A. E. Malaquias (University Museum of Bergen). Specimen collection was funded by a Putnam Expedition Grant from the Museum of Comparative Zoology to JM. Lab and analytical work were supported by internal funds from the Faculty of Arts and Sciences, Harvard University, to GG. JM postdoctoral fellowships were supported by the Ramón Areces (Spain) and the Alexander von Humboldt (Germany) Foundations.

Bibliography

Abdelkrim, J., Fedosov, A., Kantor, Y., Lozouet, P., Phuong, M., 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. doi: 10.1093/molbev/msy144/5056056

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. doi: 10.1093/molbev/msu236

Avila, C., Núñez-Pons, L., & Moles, J. (2018). From the tropics to the poles: chemical defensive strategies in sea slugs (Mollusca: Heterobranchia). In M. P. Puglisi & M. A. Becerro (Eds.),*Chemical Ecology: The Ecological Impacts of Marine Natural Products* (pp. 71–163). Boca Raton: CRC Press.

Barker, G. M. (2001). Gastropods on land: phylogeny, diversity, and adaptive morphology. In G. M. Barker (Ed.), *The Biology of Terrestrial Molluscs* (pp. 1–146). doi: 10.1079/9780851993188.0001

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* ,*30* (15), 2114–2120. doi: 10.1093/bioinformatics/btu170

Brenzinger, B., Haszprunar, G., & Schrödl, M. (2013). At the limits of a successful body plan - 3D microanatomy, histology and evolution of*Helminthope* (Mollusca: Heterobranchia: Rhodopemorpha), the most worm-like gastropod. *Frontiers in Zoology* , *10* (1), 1–26. doi: 10.1186/1742-9994-10-37

Carmona, L., Pola, M., Gosliner, T. M., & Cervera, J. L. (2013). A tale that morphology fails to tell: A molecular phylogeny of Aeolidiidae (Aeolidida, Nudibranchia, Gastropoda). *PLoS ONE* , *8* (5). doi: 10.1371/journal.pone.0063000

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

Chaban, E. M., Ekimova, I. A., Schepetov, D. M., Kohnert, P. C., Schrödl, M., & Chernyshev, A. V. (2019). Euopisthobranch mollusks of the order Cephalaspidea (Gastropoda: Heterobranchia) of the Kuril-Kamchatka Trench and the adjacent Pacific abyssal plain with descriptions of three new species of the genus *Spiraphiline*(Philinidae). *Progress in Oceanography* , *178* , 102185. doi: 10.1016/j.pocean.2019.102185

Choo, L. Q., Bal, T. M. P., Choquet, M., Smolina, I., Ramos-Silva, P., Marlétaz, F., ... Peijnenburg, K. T. C. A. (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

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. doi: 10.1016/j.ympev.2006.10.009

Colgan, D. J., Ponder, W. F., Beacham, E., & Macaranas, J. M. (2003). Gastropod phylogeny based on six segments from four genes representing coding or non-coding and mitochondrial or nuclear DNA. *Molluscan Research* , *23* (2), 123–148. doi: 10.1071/MR03002

Colgan, D. J., Ponder, W. F., & Eggler, P. E. (2000). Gastropod evolutionary rates and phylogenetic relationships assessed using partial 28s rDNA and histone H3 sequences. *Zoologica Scripta* ,*29* (1), 29–63. doi: 10.1046/j.1463-6409.2000.00021.x

Crawford, N. G., Faircloth, B. C., McCormack, J. E., Brumfield, R. T., Winker, K., & Glenn, T. C. (2012). More than 1000 ultraconserved elements provide evidence that turtles are the sister group of archosaurs. *Biology Letters* , *8* (5), 783–786. doi: 10.1098/rsbl.2012.0331

Crotty, S. M., Minh, B. Q., Bean, N. G., Holland, B. R., Tuke, J., Jermiin, L. S., & Haeseler, A. Von. (2020). GHOST: Recovering Historical Signal from Heterotachously Evolved Sequence Alignments.*Systematic Biology* , *69* (2), 249–264. doi: 10.1093/sysbio/syz051

Cunha, T. J., & Giribet, G. (2019). Deep gastropod relationships resolved. *Proceedings of the Royal Society B: Biological Sciences* , *286* , 20182776.

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* , 1755-0998.13072. doi: 10.1111/1755-0998.13072

Derkarabetian, S., Starrett, J., Tsurusaki, N., Ubick, D., Castillo, S., & Hedin, M. (2018). A stable phylogenomic classification of Travunioidea (Arachnida, Opiliones, Laniatores) based on sequence capture of ultraconserved elements. *ZooKeys* , *2018* (760), 1–36. doi: 10.3897/zookeys.760.24937

Dimitrov, D., Benavides, L. R., Arnedo, M. A., Giribet, G., Griswold, C. E., Scharff, N., & Hormiga, G. (2017). Rounding up the usual suspects: a standard target-gene approach for resolving the interfamilial phylogenetic relationships of ecribellate orb-weaving spiders with a new family-rank classification (Araneae, Araneoidea). *Cladistics* ,*33* (3), 221–250. doi: 10.1111/cla.12165

Dinapoli, A., Zinssmeister, C., & Klussmann-Kolb, A. (2011). New insights into the phylogeny of the Pyramidellidae (Gastropoda).*Journal of Molluscan Studies* , *77* (1), 1–7. doi: 10.1093/mollus/eyq027

Dunn, C. W., Hejnol, A., Matus, D. Q., Pang, K., Browne, W. E., Smith, S. A., ... Giribet, G. (2008). Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* , *452* (7188), 745–749. doi: 10.1038/nature06614

Epstein, H. E., Hallas, J. M., Johnson, R. F., Lopez, A., & Gosliner, T. M. (2018). Reading between the lines: revealing cryptic species diversity and colour patterns in *Hypselodoris* nudibranchs (Mollusca: Heterobranchia: Chromodorididae). *Zoological Journal of the Linnean Society* , *186* (1), 116–189.

Faircloth, B. C. (2013). Illumiprocessor: a trimmomatic wrapper for parallel adapter and quality trimming. Retrieved from http://dx.doi.org/10.6079/J9ILL

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

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

Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: Accelerated for clustering the next-generation sequencing data.*Bioinformatics* , *28* (23), 3150–3152. doi: 10.1093/bioinformatics/bts565

Giribet, G., Okusu, A., Lindgren, A. R., Huff, S. W., Schrodl, M., & Nishiguchi, M. K. (2006). Evidence for a clade composed of molluscs with serially repeated structures: Monoplacophorans are related to chitons.*Proceedings of the National Academy of Sciences of the United States of America* , *103* (20), 7723–7728. doi: 10.1073/pnas.0602578103

Glenn, T., Nilsen, R., Kieran, T., Finger, J., Pierson, T., Bentley, K., ... Faircloth, B. (2016). Adapterama I: universal stubs and primers for thousands of dual-indexed Illumina libraries (iTru & iNext).*BioRxiv* , *049114* . doi: https://doi.org/10.1101/049114

Gonzalez, V. L., Andrade, S. C. S., Bieler, R., Collins, T. M., Dunn, C. W., Mikkelsen, P. M., ... Giribet, G. (2015). A phylogenetic backbone for Bivalvia: an RNA-seq approach. *Proceedings of the Royal Society B: Biological Sciences* , *282* (1801), 20142332–20142332. doi: 10.1098/rspb.2014.2332

Goodheart, J. A. (2017). Insights into the systematics, phylogeny, and evolution of Cladobranchia (Gastropoda: Heterobranchia). *American Malacological Bulletin* , *35* (1), 73–81.

Goodheart, J. A., Bazinet, A. L., Collins, A. G., & Cummings, M. P. (2015). Relationships within Cladobranchia (Gastropoda: Nudibranchia) based on RNA-Seq data: an initial investigation. *Royal Society Open Science* , *2* (9), 150196. doi: 10.1098/rsos.150196

Goodheart, J. A., Bazinet, A. L., Valdes, A., Collins, A. G., & Cummings, M. P. (2017). Prey preference follows phylogeny: Evolutionary dietary patterns within the marine gastropod group Cladobranchia (Gastropoda: Heterobranchia: Nudibranchia). *BMC Evolutionary Biology* , *17* , 221. doi: 10.1186/s12862-017-1066-0

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., ... Regev, A. (2010). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* , *29* , 644–652. doi: 10.1038/nbt.1883

Gruning, B., Dale, R., Sjodin, A., Chapman, B. A., Rowe, J., Tomkins-Tinch, C. H., ... Team, T. B. (2018). Bioconda: Sustainable and comprehensive software distribution for the life sciences.*Nature Methods* , *15* (7), 475–476. doi: 10.1038/s41592-018-0046-7

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., ... Regev, A. (2013). *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols* , *8* (8), 1494–1512. doi: 10.1038/nprot.2013.084

Harasewych, M. G., Adamkewicz, S. L., Blake, J. A., Saudek, D., Spriggs, T., & Bult, C. J. (1997). Neogastropod phylogeny: A molecular perspective. *Journal of Molluscan Studies* , *63* (3), 327–351. doi: 10.1093/mollus/63.3.327

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. doi: 10.1111/j.1463-6409.1998.tb00467.x

Harasewych, M. G., & McArthur, A. G. (2000). A molecular phylogeny of the Patellogastropoda (Mollusca: Gastropoda). *Marine Biology* ,*137* (2), 183–194. doi: 10.1007/s002270000332

Haszprunar, G. (1985). The Heterobranchia: a new concept of the phylogeny and evolution of the higher Gastropoda. *Journal of Zoological Systematics and Evolutionary Research* , *23* , 15–37. doi: 10.1111/j.1439-0469.1985.tb00567.x

Huang, W., Li, L., Myers, J. R., & Marth, G. T. (2012). ART: A next-generation sequencing read simulator. *Bioinformatics* ,*28* (4), 593–594. doi: 10.1093/bioinformatics/btr708

Jackman, S., Vandervalk, B., Mohamadi, H., Chu, J., Yeo, S., Hammond, S., ... Birol, I. (2017). ABySS 2.0: resource-efficient assembly of large genomes using a Bloom filter. *Genome Research* , *27* , 768–777.

Jorger, K. M., Stoger, I., Kano, Y., Fukuda, H., Knebelsberger, T., & Schrodl, M. (2010). On the origin of Acochlidia and other enigmatic euthyneuran gastropods, with implications for the systematics of Heterobranchia. *BMC Evolutionary Biology* , *10* (1), 323. doi: 10.1186/1471-2148-10-323

Kandel, E. R. (1979). *Behavioral biology of Aplysia. A contribution to the comparative study of opisthobranch molluscs* (WH Freeman). San Francisco: WH Freeman.

Kano, Y., Brenzinger, B., Nutzel, A., Wilson, N. G., & Schrodl, M. (2016). Ringiculid bubble snails recovered as the sister group to sea slugs (Nudipleura). *Scientific Reports* , *6* , 30908. doi: 10.1038/srep30908

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability.*Molecular Biology and Evolution* , *30* (4), 772–780. doi: 10.1093/molbev/mst010

Keul, N., Peijnenburg, K. T. C. A., Andersen, N., Kitidis, V., Goetze, E., & Schneider, R. R. (2017). Pteropods are excellent recorders of surface temperature and carbonate ion concentration. *Scientific Reports* , *7* , 12645. doi: 10.1038/s41598-017-11708-w

Klussmann-Kolb, A., & Dinapoli, A. (2006). Systematic position of the pelagic Thecosomata and Gymnosomata within Opisthobranchia (Mollusca, Gastropoda) - Revival of the Pteropoda. *Journal of Zoological Systematics and Evolutionary Research* , *44* (2), 118–129. doi: 10.1111/j.1439-0469.2006.00351.x

Klussmann-Kolb, A., Dinapoli, A., Kuhn, K., Streit, B., & Albrecht, C. (2008). From sea to land and beyond–new insights into the evolution of euthyneuran Gastropoda (Mollusca). *BMC Evolutionary Biology* ,*8* , 57. doi: 10.1186/1471-2148-8-57

Kocot, K. M., Cannon, J. T., Todt, C., Citarella, M. R., Kohn, A. B., Meyer, A., ... Halanych, K. M. (2011). Phylogenomics reveals deep molluscan relationships. *Nature* , *477* (7365), 452–456. doi: 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. doi: 10.1016/j.ympev.2013.07.001

Kocot, K. M., Poustka, A. J., Stoger, I., Halanych, K. M., & Schrodl, M. (2020). New data from Monoplacophora and a carefully-curated dataset resolve molluscan relationships. *Scientific Reports* ,*10* (1), 1–8. doi: 10.1038/s41598-019-56728-w

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. doi: 10.1098/rspb.2019.0115

Korshunova, T., Fletcher, K., Picton, B., Lundin, K., Kashio, S. H. O., Sanamyan, N., ... Martynov, A. (2020). The Emperor 's*Cadlina* , hidden diversity and gill cavity evolution new insights for the taxonomy and phylogeny of dorid nudibranchs (Mollusca: Gastropoda). *Zoological Journal of the Linnean Society* , 1–66.

Krug, P. J., Vendetti, J. E., & Valdes, A. (2016). Molecular and morphological systematics of *Elysia* Risso, 1818 (Heterobranchia: Sacoglossa) from the Caribbean region. *Zootaxa* , *4148* (1), 1–137.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods* , *9* (4), 357–359. doi: 10.1038/nmeth.1923

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. doi: 10.1098/rspb.2018.2684

Lemer, S., Gonzalez, V. L., Bieler, R., & Giribet, G. (2016). Cementing mussels to oysters in the pteriomorphian tree: A phylogenomic approach.*Proceedings of the Royal Society B: Biological Sciences* ,*283* (1833), 20160857. doi: 10.1098/rspb.2016.0857

Lemmon, A. R., Emme, S. A., & Lemmon, E. M. (2012). Anchored hybrid enrichment for massively high-throughput phylogenomics. *Systematic Biology* , *61* (5), 727–744. doi: 10.1093/sysbio/sys049

Lemmon, A. R., & Lemmon, E. M. (2012). High-throughput identification of informative nuclear loci for shallow-scale phylogenetics and phylogeography. *Systematic Biology* , *61* (5), 745–761. doi: 10.1093/sysbio/sys051

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... 1000 Genome Project Data Processing Subgroup. (2009). The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* ,*25* , 2078–2079.

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. doi: 10.1016/j.ympev.2017.10.004

Lu, X. T., Gu, Q. Y., Limpanont, Y., Song, L. G., Wu, Z. D., Okanurak, K., & Lv, Z. Y. (2018). Snail-borne parasitic diseases: An update on global epidemiological distribution, transmission interruption and control methods. *Infectious Diseases of Poverty* , *7* (1), 1–16. doi: 10.1186/s40249-018-0414-7

Lunter, & Goodson. (2011). Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Research* , *21* , 936–939.

Lydeard, C., & Cummings, K. S. (2019). *Freshwater mollusks of the world* (C. Lydeard & K. S. Cummings, Eds.). Baltimore, MD: Johns Hopkins University Press.

Malaquias, M. A. E., Berecibar, E., & Reid, D. G. (2009). Reassessment of the trophic position of Bullidae (Gastropoda: Cephalaspidea) and the importance of diet in the evolution of cephalaspidean gastropods.*Journal of Zoology* , *277* (2002), 88–97. doi: 10.1111/j.1469-7998.2008.00516.x

Malaquias, M. A. E., Mackenzie-Dodds, J., Bouchet, P., Gosliner, T., & Reid, D. G. (2009). A molecular phylogeny of the Cephalaspidea*sensu lato* (Gastropoda: Euthyneura): Architectibranchia redefined and Runcinacea reinstated. *Zoologica Scripta* , *38* (1), 23–41. doi: 10.1111/j.1463-6409.2008.00354.x

McArthur, A. G., & Harasewych, M. G. (2003). Molecular systematics of the major lineages of the Gastropoda. In C. Lydeard & D. Lindberg (Eds.), *Molecular Systematics and Phylogeography of Mollusks* (pp. 140–160). Washington, D.C.: Smithsonian Books.

McCormack, J. E., Faircloth, B. C., Crawford, N. G., Gowaty, P. A., Brumfield, R. T., & Glenn, T. C. (2012). Ultraconserved elements are novel phylogenomic markers that resolve placental mammal phylogeny when combined with species-tree analysis. *Genome Research* ,*22* (4), 746–754. doi: 10.1101/gr.125864.111.746

Medina, M., Lal, S., Valles, Y., Takaoka, T. L., Dayrat, B. A., Boore, J. L., & Gosliner, T. (2011). Crawling through time: transition of snails to slugs dating back to the Paleozoic, based on mitochondrial phylogenomics. *Marine Genomics* , *4* (1), 51–59. doi: 10.1016/j.margen.2010.12.006

Minh, B. Q., Nguyen, M. A. T., & Haeseler, A. von. (2013). Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution* , *30* , 1188–1195. doi: http://dx.doi.org/10.1093/molbev/mst024

Moles, J., Avila, C., & Malaquias, M. A. E. (2019). Unmasking Antarctic mollusc lineages: novel evidence from philinoid snails (Gastropoda: Cephalaspidea). *Cladistics* , *35* (5), 487–513. doi: 10.1111/cla.12364

Neusser, T. P., Jorger, K. M., Lodde-Bensch, E., Strong, E. E., & Schrodl, M. (2016). The unique deep sea—land connection: interactive 3D visualization and molecular phylogeny of *Bathyhedyle bouchet*in. sp. (Bathyhedylidae n. fam.)—the first panpulmonate slug from bathyal zones. *PeerJ* , *4* , e2738. doi: 10.7717/peerj.2738

Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* ,*32* (1), 268–274. doi: 10.1093/molbev/msu300

Oskars, T. R., Bouchet, P., & Malaquias, M. A. E. (2015). A new phylogeny of the Cephalaspidea (Gastropoda: Heterobranchia) based on expanded taxon sampling and gene markers. *Molecular Phylogenetics and Evolution* , *89* , 130–150. doi: 10.1016/j.ympev.2015.04.011

Oskars, T. R., Too, C. C., Rees, D., Mikkelsen, P. M., Willassen, E., & Malaquias, M. A. E. (2019). A molecular phylogeny of the gastropod family Haminoeidae *sensu lato* (Heterobranchia: Cephalaspidea): a generic revision. *Invertebrate Systematics* , *33* , 426–472.

Pabst, E. A., & Kocot, K. M. (2018). Phylogenomics confirms monophyly of Nudipleura (Gastropoda: Heterobranchia). *Journal of Molluscan Studies* , *84* (3), 259–265. doi: 10.1093/mollus/eyy013

Padula, V., Bahia, J., Stoger, I., Camacho-Garcia, Y., Antonio, M., Malaquias, E., ... Schrodl, M. (2016). A test of color-based taxonomy in nudibranchs: Molecular phylogeny and species delimitation of the Felimida clenchi (Mollusca: Chromodorididae) species complex.*Molecular Phylogenetics and Evolution* , *103* , 215–229. doi: 10.1016/j.ympev.2016.07.019

Passamaneck, Y. J., Schander, C., & Halanych, K. M. (2004). Investigation of molluscan phylogeny using large-subunit and small-subunit nuclear rRNA sequences. *Molecular Phylogenetics and Evolution* , *32* , 25–38. doi: 10.1016/j.ympev.2003.12.016

Peijnenburg, K. T. C. A., Janssen, A. W., Wall-palmer, D., Goetze, E., Maas, A., Todd, J. A., & Marletaz, F. (2019). The origin and diversification of pteropods predate past perturbations in the Earth's carbon cycle. *BioRxiv* . doi: https://doi.org/10.1101/813386

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

Rambaut, A. (2014). *Figtree, a graphical viewer of phylogenetic trees* . Retrieved from https://github.com/rambaut/figtree

Rohland, N., & Reich, D. (2012). Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Research* , *22* (5), 939–946. doi: 10.1101/gr.128124.111

Schrodl, M., Jorger, K. M., Klussmann-Kolb, A., & Wilson, N. G. (2011). Bye bye "Opisthobranchia"! a review on the contribution of mesopsammic sea slugs to euthyneuran systematics. *Thalassas* , *27* (2), 101–112.

Sharma, P., & Giribet, G. (2009). Sandokanid phylogeny based on eight molecular markers-The evolution of a southeast Asian endemic family of Laniatores (Arachnida, Opiliones). *Molecular Phylogenetics and Evolution* , *52* (2), 432–447. doi: 10.1016/j.ympev.2009.03.013

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

Song, L., & Florea, L. (2015). Rcorrector: Efficient and accurate error correction for Illumina RNA-seq reads. *GigaScience* , *4* (1), 1–8. doi: 10.1186/s13742-015-0089-y

Talavera, G., & Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology* , *56* , 564–577.

Tanner, A. R., Fuchs, D., Winkelmann, I. E., Gilbert, M. T. P., Pankey, M. S., Ribeiro, A. M., ... Vinther, J. (2017). Molecular clocks indicate turnover and diversification of modern coleoid cephalopods during the Mesozoic Marine Revolution. *Proceedings of the Royal Society B: Biological Sciences* , *284* (1850), 20162818. doi: 10.1098/rspb.2016.2818

Tavare, S. (1986). Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on Mathematics in the Life Sciences* , *17* , 57–86.

Tin, M. M. Y., Economo, E. P., & Mikheyev, A. S. (2014). Sequencing degraded DNA from non-destructively sampled museum specimens for RAD-tagging and low-coverage shotgun phylogenetics. *PLoS ONE* ,*9* (5), e96793. doi: 10.1371/journal.pone.0096793

Vonnemann, V., Schrodl, M., Klussmann-Kolb, A., & Wagele, H. (2005). Reconstruction of the phylogeny of the Opisthobranchia (Mollusca: Gastropoda) by means of 18S and 28S rRNA gene sequences. *Journal of Molluscan Studies* , *71* (2), 113–125. doi: 10.1093/mollus/eyi014

Wagele, H., Deusch, O., Handeler, K., Martin, R., Schmitt, V., Christa, G., ... Martin, W. (2011). Transcriptomic evidence that longevity of acquired plastids in the photosynthetic slugs *Elysia timida*and *Plakobranchus ocellatus* does not entail lateral transfer of algal nuclear genes. *Molecular Biology and Evolution* ,*28* (1), 699–706. doi: 10.1093/molbev/msq239

Wagele, H., & Klussmann-Kolb, A. (2005). Opisthobranchia (Mollusca, Gastropoda) - more than just slimy slugs. Shell reduction and its implications on defence and foraging. *Frontiers in Zoology* ,*2* (3), 1–18. doi: 10.1186/1742-9994-2-3

Wagele, H., Klussmann-Kolb, A., Verbeek, E., & Schrodl, M. (2014). Flashback and foreshadowing—a review of the taxon Opisthobranchia. *Organisms Diversity & Evolution* , *14* , 133–149. doi: 10.1007/s13127-013-0151-5

Wagele, H., & Willan, R. C. (2000). Phylogeny of the Nudibranchia. *Zoological Journal of the Linnean Society* , *130* , 83–181. doi: 10.1006/zjls

Winnepenninckx, B., Backeljau, T., & De Wachter, R. (1996). Investigation of molluscan phylogeny on the basis of 18S rRNA sequences. *Molecular Biology and Evolution* , *13* (10), 1306–1317. doi: 10.1093/oxfordjournals.molbev.a025577

WoRMS Editorial Board. (2020). World Register of Marine Species. Available from. doi: 10.14284/170

Zaharias, P., Pante, E., Gey, D., Fedosov, A. E., & Puillandre, N. (2020). Data, time and money: evaluating the best compromise for inferring molecular phylogenies of non-model animal taxa. *Molecular Phylogenetics and Evolution* , *142* , 106660. doi: 10.1016/j.ympev.2019.106660

Zapata, F., Wilson, N. G., Howison, M., Andrade, S. C. S., Jorger, K. M., Schrodl, M., ... Dunn, C. W. (2014). Phylogenomic analyses of deep gastropod relationships reject Orthogastropoda. *Proceedings of the Royal Society B: Biological Sciences* , *281* , 20141739.

Zerbino, D. R., & Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research* , 821–829. doi: 10.1101/gr.074492.107

Data accessibility statement

Raw reads for all newly sequenced samples have been deposited in the NCBI Sequence Read Archive (BioProject PRJNA612319). The final concatenated matrices, all individual locus alignments, and resulting phylogenies are available from the Harvard Dataverse https:// doi.org/XXX.

Author contributions

J.M. designed and performed research, analysed data, and wrote the paper. G.G. contributed with laboratory reagents, and financial support. Both authors edited and approved the final manuscript.

Table 1. Number of unique recovered loci from the *in silico* test of the baits after removing duplicates from genomes and transcriptomes, including number of contigs and N50. Species in bold were used for the bait set design.

Clade

Species

Number of

contigs

N50

Number of unique contigs

% UCEs

Acochlidia

Microhedyle glandulifera

131,208

621

520

23

Acochlidia

Turbonilla sp.

260,564

827

406

18

Acteonoidea

Hydatina physis

111,189

1,132

196

8.7

Aplysiida

Aplysia californica

164,545

9,586

2,107

93.3

Caenogastropoda

Pomacea canaliculata

14,655

81,153

-

-

Architectonicoidea

Architectonica perspectiva

31,468

803

259

11.5

Caenogastropoda

Charonia tritonis

142,385

761

554

24.5

Caenogastropoda

Crepidula navicella

160,229

871

348

15.4

Cephalaspidea

Haminoea antillarum

133,818

933

714

31.6

Ellobiida

Ophicardelus sulcatus

127,388

856

618

27.4

Nudibranchia

Antiopella barbarensis

55,493

635

151

6.7

Nudibranchia

Armina californica

112,779

1,572

694

30.7

Nudibranchia

Dendronotus venustus

126,496

923

459

20.3

Nudibranchia

Dirona picta

75,472

944

194

8.6

Nudibranchia

Dondice occidentalis

109,230

704

170

7.5

Nudibranchia

Doris kerguelenensis

68,349

1,010

561

24.8

Nudibranchia

Flabellina iodinea

115,498

978

200

8.9

Nudibranchia

Phylliroe bucephala

69,583

954

393

17.4

Nudibranchia

Prodoris clavigera

99,994

892

482

21.3

Nudibranchia

Tritonia tetraquetra

63,261

465

136

6.0

Pleurobranchida

Bathyberthella antarctica

56,335

880

231

10.2

Pleurobranchida

Pleurobranchaea californica

246,413

1,222

409

18.1

Pteropoda

Clio pyramidata

34,131

744

331

14.7

Pteropoda

Limacina antarctica

86,747

1209

759

33.6

Pteropoda

Peracle reticulata

133,448

882

853

37.8

Pteropoda

Spongiobranchaea australis

100,691

1419

766

33.9

Hygrophila

Radix auricularia

69,921

24,354

2,208

97.7

Pylopulmonata

Phallomedusa solida

141,712

993

628

27.8

Rissoelloidea

Rissoella caribaea

142,330

1,709

421

18.6

Sacoglossa

Elysia chlorotica

41,686

30,474

1,994

88.3

Sacoglossa

Elysia timida

101,458

1,286

728

32.2

Sacoglossa

Oxynoe viridis

61,143

617

204

9

Systellommatophora

Onchidella floridana

145,935

874

573

25.4

Umbraculida

Tylodina fungina

79,344

751

966

42.8

Table 2. List of species used in the *in vitro* test of designed baits with collection and DNA extraction details and UCE assembly summary statistics.

Taxonomy	**Taxonomy**		**Collection**	**Collection**	**Collecti**
Order	**Species**		**Voucher**	**Locality**	**Locality**
Aplysiida	*Akera bullata*		ZMBN 83032	Italy	Italy
Aplysiida	*Aplysia dactylomela*		ZMBN 84917	Venezuela	Venezuela
Pleurobranchida	*Berthella martensi*		MCZ 393762	The Maldives	The Mald
Pleurobranchida	*Berthellina citrina*		MCZ 393775	The Maldives	The Mald
Cephalaspidea	*Bulla vernicosa*		MCZ 393827	The Maldives	The Mald
Cephalaspidea	*Colpodaspis thompsoni*		MCZ 393765	The Maldives	The Mald
Pteropoda	*Creseis acicula*		MCZ 393923	Spain	Spain
Sacoglossa	*Cyerce elegans*		MCZ 393754	The Maldives	The Mald
Aplysiida	*Dolabella auricularia*		MCZ 393823	The Maldives	The Mald
Aplysiida	*Dolabrifera dolabrifera*		MCZ 393712	The Maldives	The Mald
Nudibranchia	*Doto cervicenigra*		MCZ 393347	Spain	Spain
Sacoglossa	*Elysia pusilla*		MCZ 393737	The Maldives	The Mald
Nudibranchia	*Glossodoris acosti*	*Glossodoris acosti*	MCZ 393870	MCZ 393870	The Mald
Nudibranchia	*Godiva quadricolor*		MCZ 393921	France	France
Cephalaspidea	*Lamprohaminoea cymbalum*	*Lamprohaminoea cymbalum*	MCZ 393734	MCZ 393734	The Mald
Acteonida	*Hydatina* sp.		MCZ 382639	Australia	Australia
Acteonida	*Micromelo undatus*		MCZ 378584	Philippines	Philippin
Sacoglossa	*Oxynoe antillarum*		MCZ 393908	Caribbean	Caribbea
Cephalaspidea	*Phanerophthalmus batangas*	*Phanerophthalmus batangas*	MCZ 393707	MCZ 393707	The Mald
Nudibranchia	*Phestilla lugubris*		MCZ 393735	The Maldives	The Mald
Cephalaspidea	*Philine orca*		MCZ 393842	The Maldives	The Mald
Nudibranchia	*Phyllidia elegans*		MCZ 393746	The Maldives	The Mald
Sacoglossa	*Polybranchia jensenae*	*Polybranchia jensenae*	MCZ 393773	MCZ 393773	The Mald

Taxonomy	Taxonomy		Collection	Collection	Collecti
Nudibranchia	*Polycera hedgpethi*		MCZ 393920	France	France
Acochlidia	*Pontohedyle milaschewitchii*	*Pontohedyle milaschewitchii*	MCZ 387577	MCZ 387577	Croatia
Runcinida	*Pseudoilbia avellana*	*Pseudoilbia avellana*	MCZ 393926	MCZ 393926	Spain
Runcinida	*Runcina adriatica*		MCZ 393924	Spain	Spain
Runcinida	*Runcina africana*		MCZ 393925	Spain	Spain
Sacoglossa	*Thuridilla gracilis*		MCZ 393755	The Maldives	The Mald
Umbraculida	*Tylodina perversa*		MCZ 392614	Canary Islands	Canary Is
Umbraculida	*Tylodina rafinesquii*		MCZ 392619	Canary Islands	Canary Is

Table 3. Assembly statistics and loci present in final trimmed dataset and 50% occupancy matrix, with combined Trinity and Velvet assemblies and including the genomes and transcriptomes used for the bait set design and the transcriptomes of two caenogastropod outgroups.

Species name, voucher #	Total # contigs	Total bp	Mean length	95 CI length	Min length
Akera bullata ZMBN83032	2,190	410,372	187.4	2.04	27
Aplysia californica	2,107	478,446	227.1	2.27	102
Aplysia dactylomela ZMBN84917	2,211	469,687	212.4	2.19	33
Berthella martensi MCZ393762	1,820	373,732	205.3	2.33	59
Berthellina citrina MCZ393775	318	41,377	130	2.1	24
Bulla vernicosa MCZ393827	2,040	423,241	207.5	2.23	49
Charonia tritonis	554	106,918	193.0	3.74	38
Colpodaspis thompsoni MCZ393765	1,519	231,709	152.5	1.36	33
Crepidula navicella	348	64,753	186.1	3.66	61
Cyerce elegans MCZ393754	1,986	406,416	204.6	2.34	54
Dolabella auricularia MCZ393823	2,083	467,651	224.5	2.38	59
Dolabrifera dolabrifera MCZ393712	2,166	456,281	210.7	2.29	42
Doto cervicenigra MCZ393347	1,953	382,009	195.6	2.34	53
Elysia chlorotica	1,994	452,430	226.9	2.33	89
Elysia pusilla MCZ393737	1,871	358,391	191.6	1.88	38
Glossodoris acosti MCZ393870	1,649	342,197	207.5	2.61	62
Godiva quadricolor MCZ393921	1,933	368,323	190.5	2.13	55
Lamprohaminoea cymbalum MCZ393734	2,102	444,021	211.2	2.40	29
Haminoea antillarum	714	159,424	223.3	4.09	58
Hydatina sp. MCZ382639	1,963	388,553	197.9	1.74	51
Micromelo undatus MCZ378584	1,717	316,028	184.1	1.41	53
Oxynoe antillarum MCZ393908	2,039	417,540	204.8	2.38	39
Phanerophthalmus batangas MCZ393707	2,154	408,491	189.6	2.23	43
Phestilla lugubris MCZ393735	1,723	347,786	201.8	2.44	36
Philine orca MCZ393842	1,895	307,630	162.3	1.58	33
Phyllidia elegans MCZ393746	1,677	325,661	194.2	2.48	29
Polybranchia jensenae MCZ393773	2,075	460,829	222.1	2.31	65
Polycera hedgpethi MCZ393920	1,978	412,049	208.3	2.35	41
Pseudoilbia avellana MCZ393926	788	110,138	139.8	1.51	25
Radix auricularia	1,965	450,223	229.1	2.40	83
Runcina adriatica MCZ393924	1,618	268,956	166.2	1.64	33
Runcina africana MCZ393925	1,573	252,858	160.7	1.49	32
Thuridilla gracilis MCZ393755	2,050	419,745	204.8	2.23	45
Tylodina fungina	966	196,356	203.3	2.88	59
Tylodina perversa MCZ392614	2,141	439,598	205.3	2.30	8

Species name, voucher #	Total # contigs	Total bp	Mean length	95 CI length	Min length
Tylodina rafinesquii MCZ392619	2,099	469,179	223.5	2.38	73

Table 4. Alignment summary data, including informative sites, taxon representation, missing data, character count, and alignments found in occupancy matrices.

Alignment summary	**Alignment summary**
Loci:	Length:
2,259	525,599
Informative sites summary	**Informative sites summary**
Loci:	Total:
2,259	273,694
Taxon summary	**Taxon summary**
Mean:	95% CI:
27.44	0.2
Missing data from trim summary	**Missing data from trim summary**
Mean:	95% CI:
0	0
Character counts (times present)	**Character counts (times present)**
'-'	'A'
1,794,587	3,320,248
Data matrix completeness summary (# alignments)	**Data matrix completeness summary (# alignments)**
Matrix 50%	Matrix 60%
2,156	1,994

Figure 1. Heterobranchia phylogeny inferred from the 50% occupancy matrix (2,156 UCEs) based on BI using the GTRGAMMA model implemented in ExaBayes and ML using the GHOST model on unpartitioned alignments implemented in IQ-TREE. High-level taxonomic ranking groupings depicted in boxes in the right. The tree is rooted with the caenogastropods *Charonia tritonis* and *Crepidula navicella* . Only support values for nodes without full support are depicted. (*) denotes a conflicting topology between BI and ML, see Results for details.