

Assessment of mitochondrial genomes for heterobranch gastropod
phylogenetics

Rebecca M. Varney et al.

Deposited 2023-09-27

Citation of published version:

Varney, R. M., Brenzinger, B., Malaquias, M. A. E., Meyer, C. P., Schrödl, M., & Kocot, K. M. (2021). Assessment of mitochondrial genomes for heterobranch gastropod phylogenetics. In *BMC Ecology and Evolution* (Vol. 21, Issue 1). Springer Science and Business Media LLC. <https://doi.org/10.1186/s12862-020-01728-y>

©2021

This work is licensed under a Attribution 4.0 International (CC BY 4.0) license.



RESEARCH ARTICLE

Open Access



Assessment of mitochondrial genomes for heterobranch gastropod phylogenetics

Rebecca M. Varney¹, Bastian Brenzinger², Manuel António E. Malaquias³, Christopher P. Meyer⁴, Michael Schrödl^{2,5} and Kevin M. Kocot^{1,6*} 

Abstract

Background: Heterobranchia is a diverse clade of marine, freshwater, and terrestrial gastropod molluscs. It includes such disparate taxa as nudibranchs, sea hares, bubble snails, pulmonate land snails and slugs, and a number of (mostly small-bodied) poorly known snails and slugs collectively referred to as the “lower heterobranchs”. Evolutionary relationships within Heterobranchia have been challenging to resolve and the group has been subject to frequent and significant taxonomic revision. Mitochondrial (mt) genomes can be a useful molecular marker for phylogenetics but, to date, sequences have been available for only a relatively small subset of Heterobranchia.

Results: To assess the utility of mitochondrial genomes for resolving evolutionary relationships within this clade, eleven new mt genomes were sequenced including representatives of several groups of “lower heterobranchs”. Maximum likelihood analyses of concatenated matrices of the thirteen protein coding genes found weak support for most higher-level relationships even after several taxa with extremely high rates of evolution were excluded. Bayesian inference with the CAT + GTR model resulted in a reconstruction that is much more consistent with the current understanding of heterobranch phylogeny. Notably, this analysis recovered Valvatoidea and Orbitelloidea in a polytomy with a clade including all other heterobranchs, highlighting these taxa as important to understanding early heterobranch evolution. Also, dramatic gene rearrangements were detected within and between multiple clades. However, a single gene order is conserved across the majority of heterobranch clades.

Conclusions: Analysis of mitochondrial genomes in a Bayesian framework with the site heterogeneous CAT + GTR model resulted in a topology largely consistent with the current understanding of heterobranch phylogeny. However, mitochondrial genomes appear to be too variable to serve as good phylogenetic markers for robustly resolving a number of deeper splits within this clade.

Keywords: Heterobranchia, Gastropoda, Mitochondrial genome, Mitogenomic

Background

Mitochondrial genomes are popular molecular markers for animal phylogenetics: they are relatively easy to sequence, assemble, and annotate, typically have a moderate level of sequence conservation that facilitates phylogenetic comparisons among relatively distantly related

taxa, and can have gene order rearrangements that are potentially phylogenetically informative. Phylogenetic analyses of mitochondrial genomes have clarified relationships within diverse groups of invertebrates such as crustaceans, echinoderms, sponges, hemichordates, and annelids, just to name a few [1–5]. However, the application of mitochondrial genomes to phylogenetics can be limited by differences in evolutionary rates (which can lead to long-branch attraction (LBA) artifacts and incomplete lineage sorting [6]. Mitochondrial genome-based studies of molluscan evolutionary relationships

*Correspondence: kmkocot@ua.edu

¹ Department of Biological Sciences, The University of Alabama, Campus Box 870344, Tuscaloosa, AL 35487, USA

Full list of author information is available at the end of the article



© The Author(s) 2021. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

have been variable in success. Molluscan mitochondrial genomes exhibit wide variation in size, organization, and rate of evolution [7–9]. Mitochondrial genomes have substantially aided in clarification of relationships within clades such as Caudofoveata [10], Cephalopoda [11, 12], and some gastropod clades (e.g., [13–15]), but have had limited success at resolving relationships within other molluscan clades (e.g., [16]) and resolving deep molluscan phylogeny [17, 18].

Heterobranchia is a species-rich clade of gastropod molluscs that encompasses a wide diversity of snails and slugs that occupy marine, freshwater, and terrestrial habitats [19, 20]. Heterobranchs are thought to have diverged from other gastropods approximately 380 million years ago (mya; [21, 22]). Almost every clade within Heterobranchia has been subject to significant and continued taxonomic revision. The name Heterobranchia was coined by Burmeister (1837), but it is most commonly attributed to Gray (1840) who used it to unite Opisthobranchia (e.g., sea slugs) and Pulmonata (e.g., land snails). This group was later renamed Euthyneura to reflect the secondarily detorted arrangement of the cerebrovisceral commissures [23], but Heterobranchia was redefined to include Euthyneura and a grouping of taxa that are generally referred to as the “lower Heterobranchia” or Allogastropoda [24, 25] including, at times, Pyramidelloidea, Architectonicoidea, Valvatoidea, Orbitestelloidea, Omalogyridae, Rissoelloidea, Glacidorbidae, Tjaerhoevidae, Cimidae, Rhodopemorpha, and Murchisonellidae [21, 22, 24–30]. Opisthobranchia has since been demonstrated to be a non-monophyletic group as sea slug clades such as Sacoglossa and Acochlidia share a more recent common ancestor with the pulmonates than other sea slugs as do some “lower” heterobranchs like Pyramidelloidea and Glacidorbidae [8, 13, 22, 31, 32], reviewed by [19].

Phylogenetic analyses to date have been unable to robustly resolve most relationships among major heterobranch clades. However, most of these studies have been limited by taxon sampling. In particular, many of the “lower heterobranchs” were missing from most investigations of heterobranch phylogeny to date. These snails and slugs are thought to represent a critical group to understanding heterobranch evolution as it has been debated whether they form a clade that is sister to all remaining heterobranchs or a “basal” paraphyletic grade. Here, we sequenced mitochondrial genomes from 11 heterobranch taxa, including several so-called lower heterobranchs and select other understudied clades. These new data were analyzed in combination with publicly available heterobranch and outgroup mitochondrial genomes to investigate the utility of mitochondrial genomes for resolving higher-level heterobranch phylogeny, placement of the

lower heterobranchs, and the evolution of heterobranch mitochondrial genome organization.

Results

Genome assemblies and data matrix

We sequenced genomic DNA libraries from eleven species of heterobranch gastropods and extracted mitochondrial sequences (Table 1, Additional file 1: Table S1). Despite high sequencing coverage, a single contiguous mitochondrial genome was recovered for only two of the eleven taxa. All of the newly sequenced mt genomes were incomplete to some degree, possibly due to difficulties in sequencing through secondary structures associated with the 16S rDNA (which was absent from or incomplete in several of our assemblies) and the control region, but most of the mitochondrial protein-coding genes were obtained for all species. Alignments of amino acid sequences were produced for the thirteen protein-coding genes obtained from the newly sequenced taxa and publicly available heterobranch mt genomes on NCBI (Additional file 1: Table S1). After trimming each alignment to remove ambiguously aligned positions, the concatenated data matrix totaled 4735 amino acid sites with 31.3% gaps across 104 taxa (17 outgroups and 87 heterobranchs).

Maximum likelihood analyses

A partitioned maximum likelihood (ML) analysis of this data matrix using the best-fitting model for each gene (Additional file 2: Figure S1; see additional data on FigShare for more information) resulted in a tree with *Valvata cristata* (Valvatoidea) recovered as the sister taxon to a clade containing all other heterobranchs with successive branching of *Microdiscula charopa* (Orbitestelloidea), a clade composed of *Clione limacina* (Gymnosomata), *Psilaxis radiatus* (Architectonicoidea), *Omalogyra atomus* (Omalogyroidea), and *Rissoella morrocayensis* (Rissoelloidea), and then *Rhopalocaulis grandidieri* (Veronicelloidea), which was the sister taxon of all remaining Heterobranchia. All members of the clade containing *C. limacina*, *P. radiatus*, *O. atomus*, and *R. morrocayensis* exhibited extremely long branches relative to the other heterobranchs and it is well-established that Gymnosomata is nested within Euopisthobranchia. Thus, we strongly suspected that this clade was the result of long-branch attraction. This, combined with very low backbone support values, led us to re-run the analysis with the following unstable and long-branched taxa removed: *C. limacina*, *P. radiatus*, *O. atomus*, and *R. morrocayensis* (see Additional file 10: Table S3).

The matrix with unstable and long-branched taxa removed totaled 4447 amino acid sites with 28.7% gaps across 99 taxa. In the resulting partitioned ML analysis using the best-fitting model for each gene (Fig. 1),

Table 1 Mitochondrial genomes sequenced in the present study and associated sources of samples

Taxon	mt contig length (bp)	Protein coding genes (13)	tRNAs (22)	rRNAs (2)	Complete	Missing genes	Sample accession	Provenance
<i>Acochlidium fijiense</i>	14,098	13	20	2	Nearly	trnS2, trnK	ZSM Mol-20130988	Fiji, Viti Levu, Lami River, collected by M Schrödl & E Schwabe, August 2006
<i>Clione limacina</i>	14,599	12	17	1	Nearly	trnI, atp8, trnS1, trnD, rrnL, trnY, trnR	ZSM Mol-20081086	Antarctica, ANDEEP SYSTCO Expedition on R/V Polarstern, ANTXXIV/2 cruise station PS 71/038-04, collected by E Schwabe & H Flores, January 2008
<i>Ilbia ilbi</i>	13,944	12	21	2	Nearly	atp8	MRG10019-03	Shoreham Beach 18 March 2017, collected by A Falconer
<i>Microdiscula charopa</i>	12,965	13	21	2	Yes		MRG828-06	Dutton Way, Portland 2 March 2017, approx 2 m deep, collected by A Falconer
<i>Omalogyra atomus</i>	12,413	10	20	1	No	atp8, rrnS, trnY, nad4I, trnV	ZSM Mol-20142011	France, Pyrénées-Orientales, Banyuls-sur-Mer, from red algae in upper intertidal, collected by B Brenzinger & TP Neusser, July 2014
<i>Psilaxis radiatus</i>	12,154	12	15	1	No	atp8, trnN, trnS1, tncC, trnF, trnY, rrnL, trnQ, trnV	ZMBN 94175	Museum of Zoology at the University of Bergen
<i>Ringicula conformis</i>	14,017	12	22	2	Nearly	nad4I	None	Malta, off Ġnejna Bay, 31 May 2017
<i>Rissoella morroccayensis</i>	11,085	12	5	1	No	trnS1, rrnL, trnI, trnN, trnC, trnF, trnY, trnQ,	ZMBN 99933	Museum of Zoology at the University of Bergen
<i>Runcina ornata</i>	13,862	13	22	2	Yes		ZMBN 87949	Museum of Zoology at the University of Bergen
<i>Tylodina cf. corticalis</i>	14,614	13	21	1	Yes	trnK, rrnL	USNM 1442311	French Polynesia, Moorea, NW side Cook's Bay, collected by C McKeon, G Paulay, J O'Donnell and C Meyer, 12 June 2006
<i>Valvata cf. cristata</i>	14,495	13	21	1	Nearly	trnR, rrnL	ZSM Mol-20170210	Germany, Munich, pond on ZSM grounds, collected by B Brenzinger, March 2017

Yes: genome circularizes via overlapping ends, missing genes likely missed by annotator

Nearly: all genes listed on a single scaffold, but scaffold does not contain a full mitochondrion

No: mitochondrion either exists on two discontinuous scaffolds or is missing multiple protein coding genes

(See figure on next page.)

Fig. 1 Maximum likelihood phylogeny of heterobranch gastropods based on the reduced set of taxa following removal of both unstable leaves flagged by RogueNaRok and the four longest-branched taxa. Taxa for which new sequences were collected are shown in bold. The data set was trimmed with TrimAL with default settings, partitioned by gene in RAxML, and the PROTGAMMAAUTO was used to select the best-fitting model for each partition. Bootstrap support values are presented at each node

Valvata was again recovered as the sister taxon to a clade composed of all other heterobranchs (bootstrap support, $bs = 100$) followed by the successive branching of *Microdiscula* ($bs = 62$) and *Rhopalocaulis* ($bs = 100$). Most major clades of Heterobranchia (Euthyneura sensu lato) were recovered with high support: Acteonoidea, Nudipleura, Cephalaspidea, Runcinida, Aplysiida, Siphonariida, Sacoglossa, and Stylommatophora were all recovered with 100% bootstrap support, and Systellommatophora with 99% bootstrap support. Of the family- and order-level taxa, Ellobioidea is the only one that was recovered as non-monophyletic, with *Pedipes pedipes* and *Myosotella myosotis* falling well outside of the clade containing the rest of Ellobioidea, albeit with very low support. Support for relationships among most higher-level heterobranch clades was generally weak and a number of higher-level groupings within Heterobranchia including Tectipleura, Euopisthobranchia, Panpulmonata, Eupulmonata, Systellommatophora, and Amphipulmonata (sensu [20]) were not monophyletic. However, Nudipleura (Nudibranchia + Pleurobranchomorpha) and a clade composed of Aplysiida + Umbraculoidea were recovered monophyletic with maximal support.

To explore the impact of different partitioning schemes on tree topology, and to determine whether other models better mitigated the long-branch attraction of *C. limacina*, *P. radiatus*, *O. atomus*, and *R. morrocaeyensis*, we ran several additional analyses. A partitioned analysis with a mixed model (LG + C60 + G + F) yielded a tree (Additional file 3: Figure S2) with the same clade of long-branched taxa as sister to all remaining heterobranchs except *Valvata* and *Microdiscula* and did not vary significantly in any other way from the original ML tree (Additional file 2: Figure S1). A ML analysis with Lanfear clustering (Additional file 4: Figure S3) and a fully partitioned ML analysis with resampling within partitions (Additional file 5: Figure S4) both produced similar trees to the initial partitioned ML analysis (Additional file 2: Figure S1), but with the two members of Ellobioidea that did not form a clade with the rest of Ellobioidea (*Myosotella* and *Pedipes*) falling outside Hygrophila and thus even farther from the remaining ellobioids. An analysis with an edge-unlinked model altered the relationships within the long-branched clade, with *O. atomus* and *P. radiatus* as sister to *R. morrocaeyensis* and *C. limacina*,

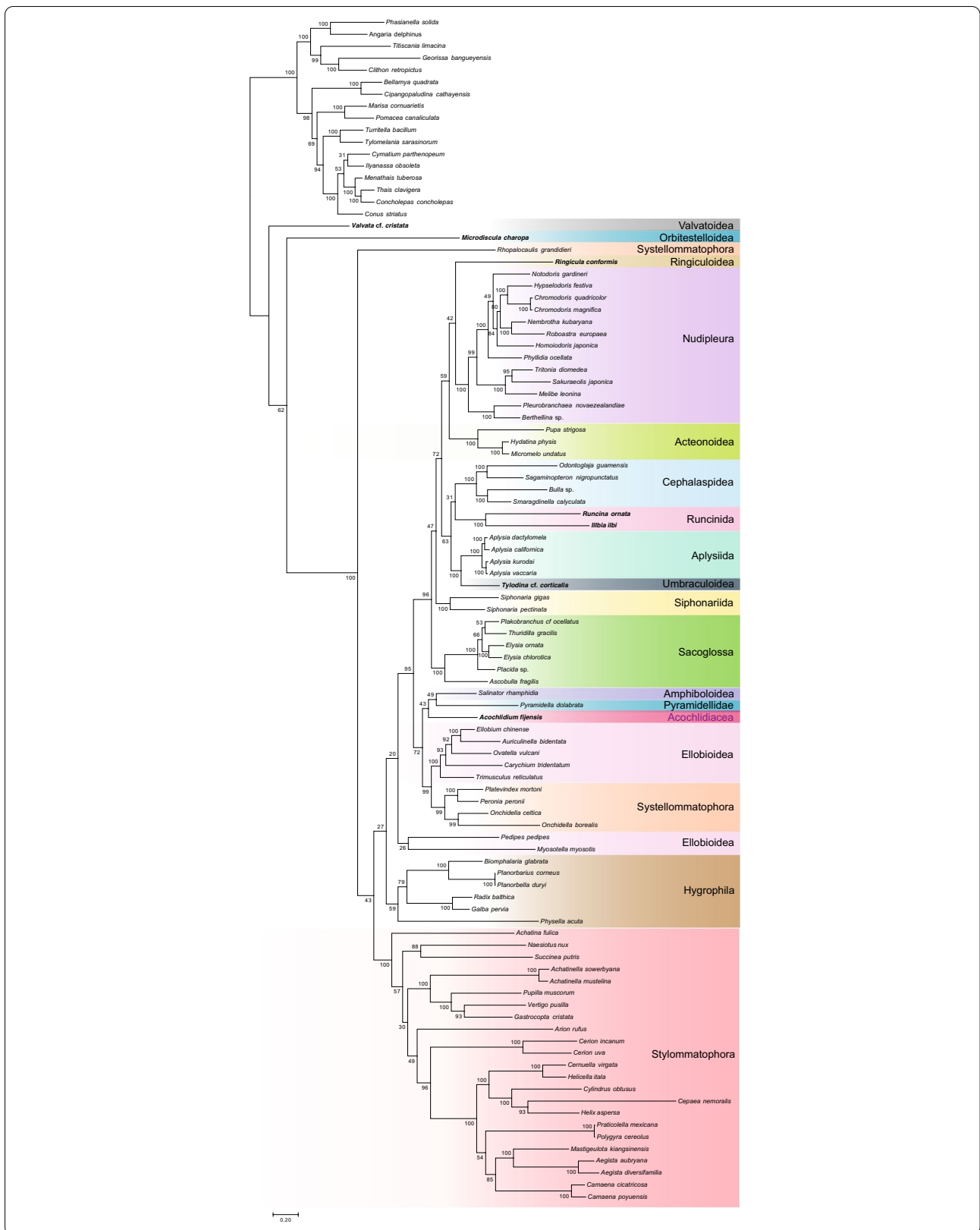
and, while in previous analyses *R. morrocaeyensis* had a much longer branch than the other taxa in this clade, in the edge-unlinked tree, all four taxa had similarly elongated branches (Additional file 6: Figure S5). In this edge-unlinked analysis, the positions of Hygrophila and the pair of ellobioids were recovered with relationships similar to those of the initial partitioned ML tree (Additional file 2: Figure S1). We also analyzed the set of all taxa except *C. limacina* to assess whether removal of this single rapidly-evolving taxon would ‘release’ the other long-branched taxa, which are traditionally considered to be lower heterobranchs, from this putative LBA artifact. The other three long-branched taxa remained in the same location with long branches (Additional file 7: Figure S6).

Bayesian inference

Because of poor support for most nodes of interest in our maximum likelihood analyses, we also performed a Bayesian inference with the CAT + GTR + G4 model on the same datasets, but only the analysis of the dataset with unstable and long-branched taxa removed reached convergence. Of the six chains that were run for over 60,000 generations, four converged according to the PhyloBayes bpcomp maxdiff criterion (maxdiff value = 0.29), yielding a tree with a topology that is much more consistent with the current understanding of heterobranch relationships (Fig. 2). *Valvata* and *Microdiscula* were recovered in a polytomy with a clade that comprised all other heterobranchs, which received maximal support. This clade consisted of a polytomy of Nudipleura + Acteonoidea, *Ringicula*, and the remaining heterobranchs. Nudipleura + Acteonoidea was weakly supported but Nudipleura again received maximal support.

The largest recovered subclade within Heterobranchia, Tectipleura, consisted of Euopisthobranchia (Cephalaspidea, Runcinida, Aplysiida, and Umbraculoidea) and Panpulmonata (Siphonariida, Sacoglossa, Hygrophila, Ellobioidea, Amphiboloidea, Systellommatophora, and Stylommatophora), which were recovered reciprocally monophyletic and both clades received maximal support. Within Euopisthobranchia, Cephalaspidea and Runcinida formed a (weakly supported) clade sister to a clade of Aplysiida + Umbraculoidea, the latter of which was strongly supported (posterior probability, $pp = 0.98$).

Within Panpulmonata, Siphonariida was recovered as the sister taxon to the rest of the clade ($pp = 1$) with



(See figure on next page.)

Fig. 2 Bayesian inference phylogeny of Heterobranch molluscs based on the reduced set of taxa following removal of both unstable leaves flagged by RogueNaRok and the four longest-branched taxa. Taxa for which new sequences were collected in the present study are shown in bold. The data set was trimmed with BMGE and trees were generated in PhyloBayes with four chains using the CAT + GTR + Γ 4 substitution model. Tree shown is the majority rule posterior consensus tree. Posterior probabilities are presented at each node

Sacoglossa sister to all other taxa within that clade (pp=0.96). The remaining panpulmonates formed two clades. One consisted of Stylommatophora, Systellommatophora, and Ellobioidea, although neither Systellommatophora nor Ellobioidea were recovered monophyletic. *Rhopalocaulis* (Systellommatophora) was recovered as the sister taxon of Stylommatophora (pp=0.71) and this clade was recovered in a polytomy with the ellobioids *Pedipes* and *Myosotella* that was maximally supported. Sister to this polytomy was a moderately well-supported clade (pp=0.98) in which the remainder of Systellommatophora (Onchidiidae) was sister to the remaining Ellobioidea. Sister to the Stylommatophora-Systellommatophora-Ellobioidea clade was a clade comprising Hygrophila (pp=1.00), *Pyramidella dolabrata* (Pyramidelloidea), *Salinator rhamphidia* (Amphiboloidea), and *Acochlidium fijiense* (Acochlidia). *Salinator* and *Pyramidella* formed a well-supported clade (pp=0.99) but otherwise, higher-level relationships in the Hygrophila-Pyramidelloidea-Amphiboloidea-Acochlidia clade were weakly supported.

Mitochondrial gene order evolution

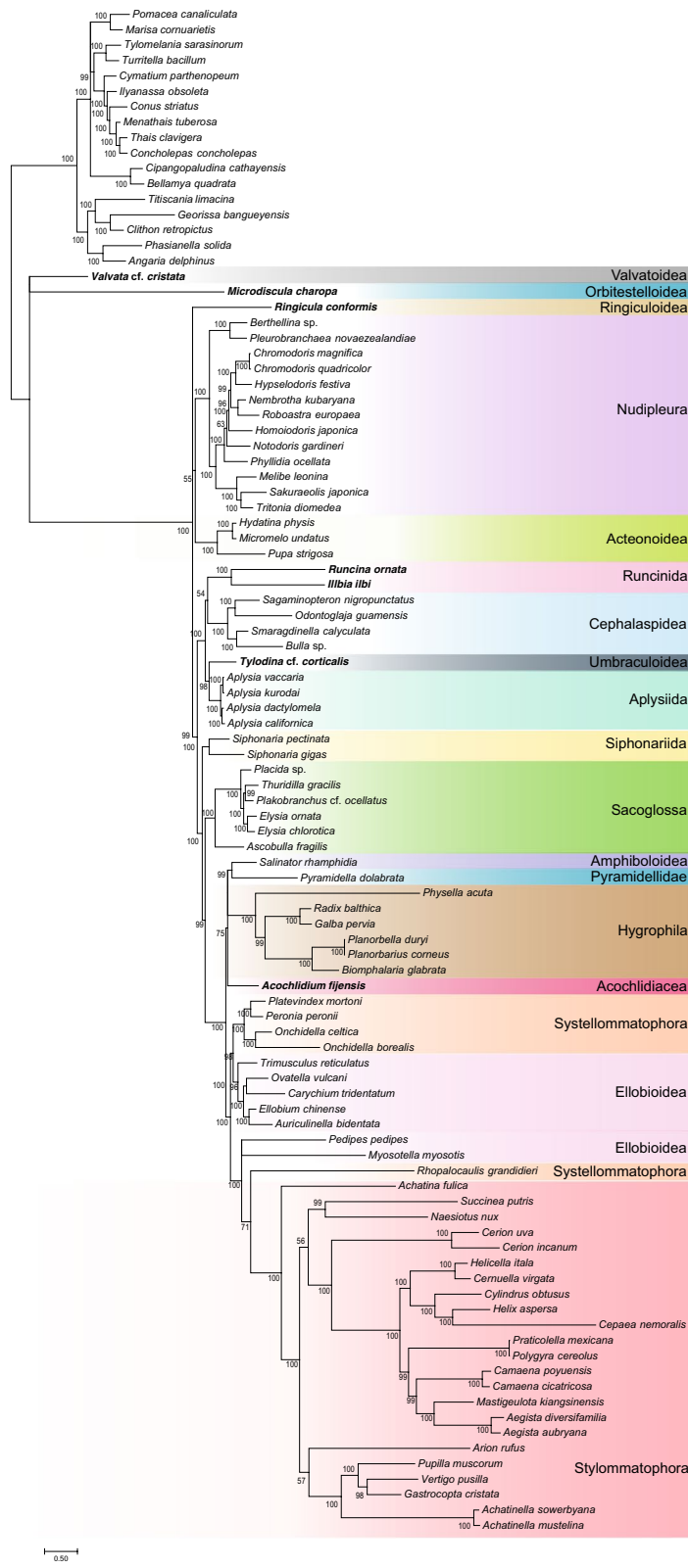
A somewhat diagnostic gene arrangement exists for heterobranchs relative to other gastropod clades, but many heterobranch taxa and subclades have differences in both gene organization and orientation in their mitochondrial genomes (Fig. 3). Caenogastropods encode all mitochondrial genes in the same orientation, while all members of the clade comprising Neritimorpha and Vetigastropoda share a single inversion of [12S rRNA, 16S rRNA, nad1, nad6, cytB, nad4L, nad4, nad5]. Across the diverse taxa used as outgroups in this study, no individual deviations in gene arrangement were found.

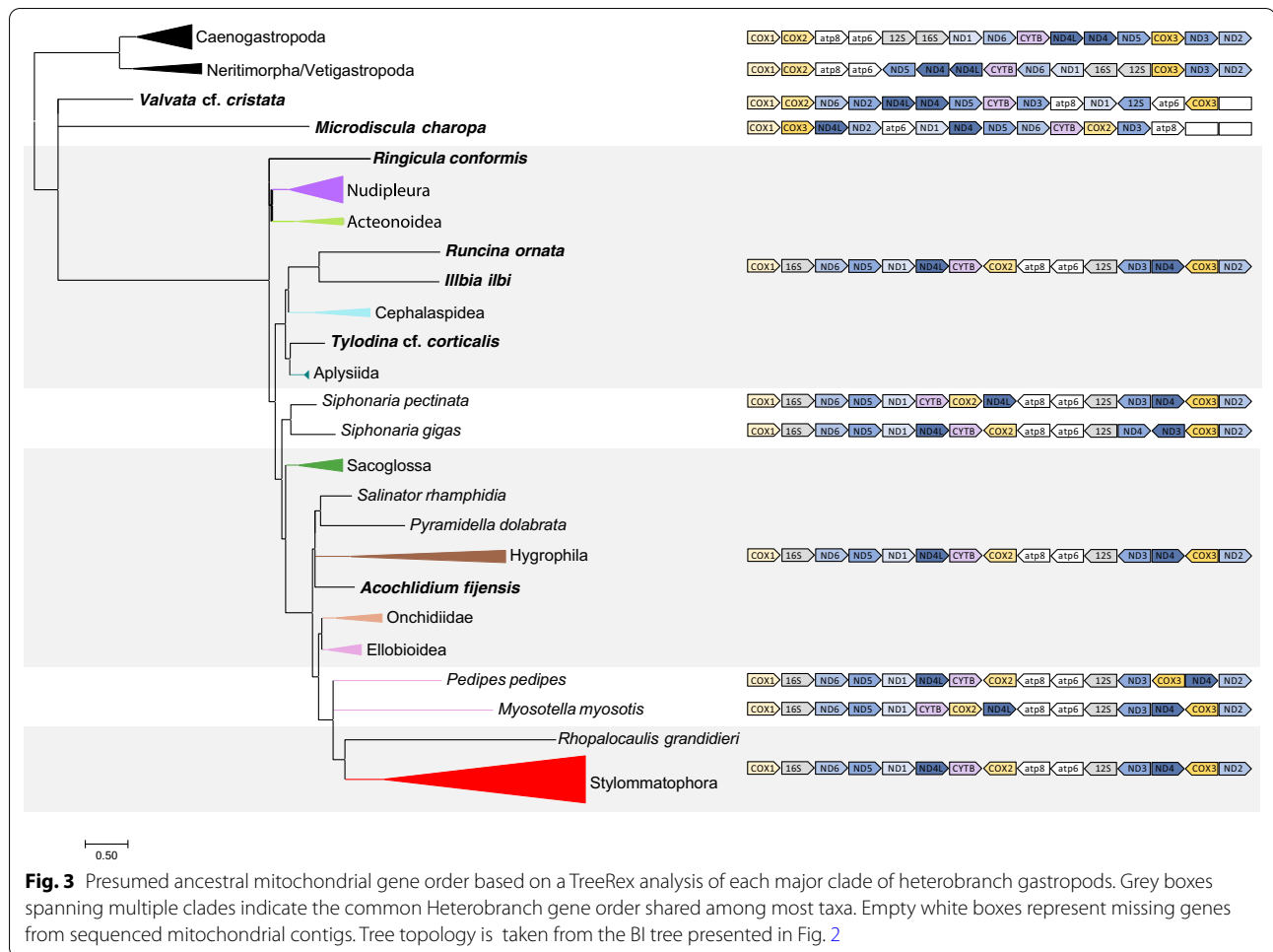
In contrast to this consistency, the taxa at the base of our heterobranch tree all have remarkably different mitochondrial gene arrangements from one another. Mitochondrial gene order within most of the “lower Heterobranchia” is variable: *Psilaxis radiatus* (Architectonicoidea), *Omalogyra atomus* (Omalogyridae), *Rissoella morrocayensis* (Rissoellidae), and *Valvata cf. cristata* (Valvatidae) all have distinct gene orders from one another and from remaining heterobranchs, including changes in both order and orientation. *Microdiscula charopa* also has an entirely unique gene order.

In the remaining heterobranchs, the clade spanning Acteonoidea, Nudipleura, and the subclade including Runcinida, Cephalaspeida, Umbraculoidea, Aplysiida, Siphonariida, Sacoglossa, Amphiboloidea, Pyramidellidae, Hygrophila, Acochlidia, Systellommatophora, Ellobioidea, and Stylommatophora, a relatively stable mitochondrial gene order and orientation exists, referred to here as the common heterobranch gene order [cox1-16SrRNA-nad6-nad5-nad1-nad4L-cytB-cox2-atp8-atp6-12SrRNA-nad3-nad4-cox3-nad2, with atp8-nad3 and cox3 both reversed in direction from cox1]. All members of Nudipleura examined adhere to this common gene order except *Hypselodoris festiva*, in which a single gene (nad4) changed position, and all members of Acteonoidea adhere to the same order as well. Variation exists within the Cephalaspeida, with a shared rearrangement of cytB, nad1, nad4L, and cox2 shared among three-fourths of its members, and *Sagaminopteron nigropunctatum* containing further rearrangements. Aplysiida adheres to the stable arrangement with the exception of *Aplysia kurodai*, in which the orientation of the 12S rRNA gene is reversed though its position remains the same.

Both representatives of Siphonariida have different internal mitochondrial gene rearrangements: *Siphonaria gigas* reversed the positions of nad4 and nad3, while *Siphonaria pectinata* inserts cox2 between nad4L and cytB. All sacoglossans shared a common gene order, as do *Pyramidella dolabrata* and *Acochlidium fijiense*. The majority of Ellobioidea are consistent, excepting *Myosotella myosotis* and *Pedipes pedipes*, which have different single-gene transpositions than one another. Additionally, the mt genome of *P. pedipes* is expanded, with more intergenic space than other closely related taxa. Interestingly, these two taxa are those that fall together in a different part of the phylogeny than the remaining members of Ellobioidea, making this group paraphyletic. The clade comprising Hygrophila was strongly supported, and all members within it share the common heterobranch gene order except *Physella acuta*, which has a completely unique gene arrangement.

Within Stylommatophora, both members of genus *Achatinella* shared a single gene (cox2) moved to a different position relative to other members of the clade. Likewise, the taxa *Cylindrus obtusus*, *Cepaea nemoralis*, and *Cornu aspersum* (syn. *Helix aspersa*) all share a single gene (nad4) inserted at a different location in the





mitochondrial genome. *Arion rufus* has the 12S rRNA placed prior to atp8-atp6 instead of after it, but all other members of Stylommatophora shared the common heterobranch gene order.

Discussion
Heterobranch phylogeny

We sequenced mitochondrial genomes from eleven heterobranch gastropods and investigated heterobranch evolutionary relationships using amino acid sequences from the thirteen mitochondrial protein-coding genes as well as the evolution of heterobranch mitochondrial genome organization. Mitochondrial genomes can be useful in molecular phylogenetics because of the functional constraint that should, in theory, lead to a relatively high degree of conservation across evolutionary time. This has been demonstrated in diverse groups of animals where mitochondrial genomes have served as useful markers for molecular phylogenetics [1–3, 5]. However, in other animal lineages, it has been demonstrated that

mitochondrial genome evolutionary rate is too rapid to recover ancient radiations (e.g., [8, 17, 33]).

Our maximum likelihood-based analysis including all taxa failed to recover most recognized higher-level heterobranch clades but did recover a maximally supported clade of taxa with extremely long branches near the base of Heterobranchia. This clade includes taxa known to have brief lifespans, some of only a few months, which may correlate with a more rapid accumulation of genetic changes [34]. To combat this putative artifact of long-branch attraction, ML analyses of a dataset with long-branched and unstable taxa excluded were performed. Excluding these taxa resulted in a tree that exhibited an apparent mis-rooting within Heterobranchia relative to other studies (e.g., [22, 31, 32], reviewed by [19]) with Panpulmonata paraphyletic with respect to a clade of opisthobranchs. Support for most higher-level heterobranch clades was weak in both maximum likelihood analyses, although most order-level taxa were recovered monophyletic with strong support.

Clear long-branch attraction and weak support for deep relationships within Heterobranchia initially led us to conclude that mitochondrial genomes have little to no phylogenetic signal for deep nodes within Heterobranchia. Additional maximum likelihood analyses that attempted to account for differences in evolutionary rates between genes did not improve resolution of these deeper nodes. However, although mostly weakly supported, a number of previously hypothesized relationships were recovered in all of our maximum likelihood analyses. These include Euthyneura (e.g. all heterobranchs except Valvatoidea and Orbitestelloidea), Pyramidelloidea + Amphiboloidea, Nudipleura + *Ringicula* (Ringipleura), Ringipleura + Acteonoidea (not considering Rissoelloidae), Aplysiida + Umbraculoidea (not considering Gymnosomata and Thecosomata), and Cephalaspidea + Runcinida [22, 31, 32, 35–38].

In order to determine if selecting a model that better accounts for site-specific rate heterogeneity could help improve resolution, we conducted a Bayesian inference using the site heterogeneous CAT + GTR + G4 model. This analysis resulted in a topology that is much more consistent with other studies examining heterobranch evolutionary relationships to date. Again, we recovered Euthyneura to the exclusion of Valvatoidea and Orbitestelloidea with maximal support. Whereas our maximum likelihood analyses recovered Valvatoidea as the sister taxon to all other heterobranchs with moderate to weak support, our Bayesian inference recovered these two “lower heterobranchs” in a polytomy with the rest of Heterobranchia. This is in concordance with previous Sanger-sequencing based approaches [20, 29] but neither clade was so far sampled by phylogenomics [35] or mitogenomics [15]. Valvatoidea (=Ectobranchia) is a group of minute freshwater and marine snails with discoidal to ovoid shells. Haszprunar et al. regarded Valvatoidea as the earliest-branching heterobranch clade based on their broad, rhipidoglossate radula and unusual ectobranch gill [39]. This phylogenetic position was favored by Brenzinger et al. because a clade of all heterobranchs except Valvatoidea is supported by the presence of ciliated strips in the mantle cavity [40]. Sperm ultrastructure is also consistent with their placement among the lower heterobranchs [41]. Orbitestelloidea was considered to belong to Valvatoidea in the past. Our Bayesian analysis produced a polytomy containing these taxa, but all our maximum likelihood analyses separated these taxa from one another with Valvatoidea sister to all other heterobranchs, as consistent with the most recent classification [27]. The fossil record also coincides with a greater age of “lower” heterobranchs (possibly present in mid-Paleozoic) vs. Euthyneura (diversifying in the

Jurassic) [42–44], although unequivocal fossils of Valvatoidea and Orbitestelloidea—with minute, fragile and often inconspicuous shells—are much younger (Cretaceous to Eocene) (see [42, 43]). Architectonicoidea is another candidate for an old group judging from the presence of fossils in the Triassic [45]. Unfortunately, most of the other lower heterobranchs we sampled—*O. atomus* (Omalogyroidea), *Psilaxis radiatus* (Architectonicoidea), and *Rissoella morrocaensis* (Rissoelloidae)—exhibited extremely long branches and the Bayesian inference including these taxa (as well as an analysis including these taxa but excluding *C. limacina*; data not shown) failed to converge.

Our Bayesian inference recovered Pyramidelloidea + Amphiboloidea and Aplysiida + Umbraculoidea with strong support ($pp \geq 0.98$). This analysis also recovered a number of other heterobranch clades that have been identified in other studies but were not recovered in the maximum likelihood analysis of this dataset. Most notably among these is Panpulmonata. We recovered Siphonariida as the sister taxon of the remaining panpulmonates followed by Sacoglossa as the sister taxon to all other panpulmonates after that, all with strong support ($pp \geq 0.99$). Interestingly, support for the relative placement of these two clades has been weak in most studies with the relevant taxon sampling to date (but see [31]). Our results are inconsistent with most studies to date, which have recovered these two taxa as a clade [22, 38] or with Sacoglossa, not Siphonariida, sister to the rest of Panpulmonata [31, 32], reviewed by [20, 46].

Although *Ringicula* was previously recovered as the sister taxon of Nudipleura [32], which we recovered here in our maximum likelihood analyses, this relationship was not supported in our Bayesian inference. Instead, Nudipleura was recovered as the sister taxon of Acteonoidea, but this clade was weakly supported. *Ringicula* was recovered in a polytomy with this weakly supported Nudipleura-Acteonoidea clade (and a strongly supported clade consisting of all other heterobranchs), so although we find no support for the Ringipleura hypothesis in this analysis, our Bayesian inference results are not incompatible with Ringipleura.

All of our analyses failed to recover Ellobioidea as a monophyletic group. A previous analysis that included some of the Ellobioidea mitochondrial genomes analyzed herein, including those of the two taxa that were recovered separately from the rest of Ellobioidea in our analyses (*Pedipes pedipes* and *Myosotella myosotis*), also failed to recover a monophyletic Ellobioidea [47]. However, Dayrat et al. and Romero et al. sampled both of these species and recovered them as nested within Ellobioidea (although Dayrat et al. also recovered *Trimusculus* and *Otininae* within Ellobioidea) [48, 49].

Evolution of heterobranch mitochondrial gene organization

Long-branch attraction, as discussed above, is likely responsible for the recovery of *C. limacina* in a clade with the “lower heterobranchs” *O. atomus*, *P. radia-tus*, and *R. morrocyensis*. Often there is a correlation between a high rate of genome evolution and genome rearrangements [50]; *O. atomus* and members of the genus *Rissoella* are known to have short life cycles [20, 51]. *C. limacina* has a completely unique gene order, so it is possible that some sequence differences may be a result of rearrangement and these in turn contributed to the misplacement of this taxon. Within Ellobioidea, the two members that are consistently recovered apart from the rest (*Myosotella myosotis* and *Pedipes pedipes*) both contain single gene transpositions (though differing from one another).

However, this correlation does not hold for other isolated members of clades with extreme gene order rearrangements. For example, *Physella acuta* has mitochondrial gene reordering so extensive that a minimum of five independent gene rearrangements are necessary to account for the difference between it and the remaining members of Hygrophila [52]. Despite this, *P. acuta* still forms a clade with the rest of Hygrophila with very high support in all analyses. Likewise, *Sagaminopteron nigro-punctatus* forms a clade with the other cephalaspids with very high support despite differing dramatically in gene arrangement from the other three members included in the analysis, and the relationships among these taxa are supported by recent transcriptome-based analyses [53]. The variable relationship of evolutionary rate of gene sequences and mitochondrial gene rearrangement could be interesting to investigate in future studies.

The shared gene arrangement among the majority of heterobranch taxa suggests this common gene order emerged prior to the common ancestor of Nudipleura and remaining taxa. However, most taxa previously identified as “lower heterobranchs,” as well as the additional taxa recovered at the base of the heterobranch tree in our analyses, have unique mitochondrial gene arrangements relative to all other gastropods. The differences among these taxa and between them and the main clade of heterobranchs cannot be explained with stepwise changes, but instead suggest multiple independent inversions and transpositions and may be due to a combination of long evolutionary trajectories (since the mid-Paleozoic) [54] and derived ecologies and lifestyles in many subgroups, including the commonly observed abbreviation and modification of life cycles by multiple evolution of paedomorphic groups. Our results indicate that mitochondrial genome organization started to deviate considerably from the ancestral molluscan arrangement first at the

origin of Heterobranchia and later, even more so, at the origin of Euthyneura.

Conclusions

Here, we sequenced 11 new heterobranch mitochondrial genomes including several “lower heterobranchs”. These new data were analyzed in combination with publicly available heterobranch and outgroup mitochondrial genomes using maximum likelihood and Bayesian inference. Results of maximum likelihood analyses with site-homogeneous models indicated that even with the exclusion of exceptionally rapidly evolving taxa, mitochondrial genomes have limited utility for resolving most higher-level heterobranch relationships. However, Bayesian inference using the site-heterogeneous CAT+GTR+G model recovered most recognized higher-level heterobranch clades including Tectipleura, Euopisthobranchia, and Panpulmonata. Unfortunately, most of the lower heterobranch taxa that we aimed to place in a phylogenetic context exhibited extremely fast rates of evolution. Relationships within most heterobranch order-level clades that were broadly sampled (e.g., Nudipleura, Aplysiida, Sacoglossa, Hygrophila, Stylomatophora) were well-resolved and strongly supported. Despite the relatively rapid rate of nucleotide evolution in heterobranch mitochondrial genomes, gene order was found to be largely conserved across the group. Taken together, these results provide support for several hypothesized heterobranch clades and highlight the non-euthyneuran clades Valvata and Orbitestelloidea as interesting and important taxa to study with respect to understanding early heterobranch evolution. However, a lack of resolution and poor support for a number of deeper nodes within Heterobranchia highlight limitations of mitogenomic data for deep phylogeny, especially for rapidly evolving taxa like the long-branched lower heterobranchs, and reveals the surprising degree of heterogeneity within even closely related molluscan taxa that may in part be responsible for these limitations.

Methods

DNA extraction, library preparation, and sequencing

DNA was extracted from specimens obtained from various sources (Table 1) using the Omega Bio-tek EZNA MicroElute Genomic DNA Kit (Omega Bio-tek, Norcross, GA) or with a MO-BIO Powermax Soil DNA Isolation Kit. As most of the newly sequenced taxa were small-bodied, in most cases entire specimens were placed directly into lysis buffer, and if size permitted, were ground with a sterile pestle prior to digestion to break open shells. DNA concentration was measured using a Qubit 4 Fluorometer (Thermo Fisher Scientific, Waltham, MA) with the ds DNA HS kit. Samples that

yielded too little DNA for library preparation (*Rissoella morrocayensis* and *Omalogyra atomus*) were amplified with multiple strand displacement amplification using the Illustra Single Cell GenomiPhi DNA Amplification Kit (GE Healthcare, Chicago, IL). Dual-indexed sequencing libraries were prepared with the Illumina Nextera Kit (Illumina, San Diego, CA). Library size was assessed via agarose gel following a test PCR (run with provided Illumina primers 1.1 and 2.1, run 95 °C for 10 min followed by 40 cycles of [95° for 10 s, 60° for 30 s]). Pooled libraries were sequenced with a 2 × 100 bp paired-end TruSeq 3000/4000 SBS kit on an Illumina HiSeq4000 (Macrogen, South Korea) using 1/24 lane each.

Assembly and annotation

De novo assemblies of reads were initially carried out with Spades 3.14.0 [55]. In the case of *O. atomus*, the longest mitochondrial genome contig produced by Spades was missing several mitochondrial genes and Ray 2.2.0 [56] was used for assembly. Mitochondrial genomes were identified by creating a BLAST database from each set of assembled scaffolds and querying that database with the complete mitochondrion of *Galba pervia* (NCBI NC_018536.1) via BLASTN with an e-value cut-off of 1e-4. The longest BLAST hits were annotated with the MITOS2 web server with default parameters and the invertebrate mitochondrial genetic code (5) [57].

Data set construction

Coding sequences of the 13 mitochondrial protein-coding genes (*cox1*, *cox2*, *cox3*, *atp6*, *atp8*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *cob*) were extracted from the newly sequenced and annotated mitochondrial genomes, as well as those publicly available on NCBI (see Additional file 1: Table S1). Single-gene alignments were simultaneously produced for DNA and amino acid sequences with MACSE v1.2 [7] using the invertebrate genetic code (5). Alignments were trimmed with trimAL with default settings [58]. Trimmed alignments were checked manually in MEGA 10.0.4 [59] and corrected by hand if translations were initially out of frame. FASconCAT [60] was used to assemble the concatenated amino acid supermatrix file. In response to difficulties with long-branched taxa (Additional file 2: Figure S1) and in keeping with recent recommendations to improve phylogenetic analysis [61], the alignment was also trimmed with BMGE [62] trimming with default settings. The BMGE matrix was used for subsequent analyses. All data matrices are available online via FigShare.

Maximum likelihood analyses

An initial maximum likelihood analysis (Supplemental Figure S1) was conducted on the initial TrimAL-trimmed

(with default settings), partitioned-by-gene supermatrix using RAxML v8.2.4 [63] with the PROTGAMMAUTO model, which automatically selects the best-fitting model for each partition, rapid bootstrapping, and selection of the best-scoring maximum likelihood tree in one run. The number of bootstrap replicates was determined by the majority-rule consensus criterion (autoMRE).

Leaf stability was assessed with RogueNaRok [64] using the majority rule consensus criterion. Four taxa (*R. grandideri*, *P. pedipes*, *P. acuta*, and *M. myosotis*) had a leaf stability difference of <0.75 and were considered to be unstable by RogueNaRok (Additional file 8: Table S2). These taxa, along with the very long-branched taxa *C. limacina*, *P. radiatus*, *O. atomus*, and *R. morrocayensis* were removed and the remaining sequences for each gene were re-aligned, trimmed, and concatenated before reanalysis in RAxML as described above.

To attempt to combat the apparent long branch attraction among *C. limacina*, *P. radiatus*, *O. atomus*, and *R. morrocayensis*, trees were also produced for the BMGE-trimmed matrix with a number of different models and/or analysis settings. We performed a series of ML analyses in IQ-TREE 2 [65] with 1000 rapid bootstraps employing different models and partitioning schemes including (1) the BMGE-trimmed dataset partitioned by gene with a partitioned mixed model (LG + C60 + G + F) and the best tree from RAxML provided as a starting tree (Additional file 3: Figure S2); (2) the same BMGE-trimmed dataset partitioned by gene and using Lanfear clustering to select optimal partitioning [66], resulting in 5 partitions with different models (Additional file 4: Figure S3); (3) a fully partitioned analysis of this matrix where PartitionFinder independently selected the best model for each gene with the –GENESITE correction to resample partitions and then sites within partitions [67] (Additional file 5: Figure S4); (4) an analysis of this matrix with an edge-unlinked model to better account for heterotachy (GHOST) [68] (Additional file 6: Figure S5). We also ran a RAxML analysis on the original TrimAL-trimmed dataset but with *C. limacina* removed (Additional file 7: Figure S6). In all IQ-TREE 2 and RAxML trees, the clade of four (or three, in the last analysis) long-branched taxa persisted, and the overall tree topology did not change.

Bayesian analysis

Bayesian trees were generated with PhyloBayes-MPI v1.6 [69] with four chains and the CAT + GTR + Γ 4 substitution model. Two analyses were attempted based on the BMGE-trimmed matrix: (1) an analysis sampling all taxa; and (2) an analysis excluding the taxa flagged as unstable in the initial maximum likelihood tree (*C. limacina*, *P. radiatus*, *O. atomus*, and *R. morrocayensis*).

Mitochondrial gene order

In light of the apparent heterogeneity in mitochondrial gene sequences within and between clades, we examined gene order across major groups with TreeREx v1.85 [70]. The heterogeneity within several groups made it impossible to visualize all at once (Additional file 8: Figure S7), so nodes of major clades were collapsed and the inferred ancestral gene arrangement for each clade diagrammed again with TreeREx (Additional file 11: Figure S8). Syntenic blocks were visualized with Geneious R11 (Additional file 10: Table S3).

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12862-020-01728-y>.

Additional file 1: Table S1. Data downloaded from NCBI used in the present study.

Additional file 2: Figure S1. Maximum likelihood phylogeny of heterobranched gastropods based on the full set of available heterobranched mitochondrial genomes (including long-branched taxa). The data set was partitioned by gene, trimmed with TrimAL with default settings, and analyzed in RAxML with the PROTGAMMAAUTO setting to select the best-fitting model for each partition. Bootstrap support values are presented at each node.

Additional file 3: Figure S2. Maximum likelihood phylogeny of heterobranched gastropods based on the full set of available heterobranched mitochondrial genomes (including long-branched taxa). The data set was partitioned by gene, trimmed with BMGE, and analyzed in IQ-TREE 2 with the LG + C60 + G + F mixed model. Bootstrap support values are presented at each node.

Additional file 4: Figure S3. Maximum likelihood phylogeny of heterobranched gastropods based on the full set of available heterobranched mitochondrial genomes (including long-branched taxa). The data set was partitioned by gene, trimmed with BMGE, and greedy Lanfean clustering was applied in IQ-TREE 2 to determine the optimal partitioning scheme. Five partitions with independent models were applied. Bootstrap support values are presented at each node.

Additional file 5: Figure S4. Maximum likelihood phylogeny of heterobranched gastropods based on the full set of available heterobranched mitochondrial genomes (including long-branched taxa). The data set was partitioned by gene, trimmed with BMGE, and each partition was allowed to select its own optimal model via ModelFinder implemented in IQ-TREE 2. The analysis was run with the -GENESITE correction to facilitate resampling first within partition and then within sites. Bootstrap support values are presented at each node.

Additional file 6: Figure S5. Maximum likelihood phylogeny of heterobranched gastropods based on the full set of available heterobranched mitochondrial genomes (including long-branched taxa). The data set was trimmed with BMGE, concatenated into a supermatrix, and analyzed with an edge-unlinked model to better account for heterotachy (GHOST). The analysis was run in IQ-TREE 2 with the -GENESITE correction to facilitate resampling first within partition and then within sites. Bootstrap support values are presented at each node.

Additional file 7: Figure S6. Maximum likelihood phylogeny of heterobranched gastropods based on the full set of available heterobranched mitochondrial genomes (including long-branched taxa) except for *C. limacina*. The data set was partitioned, trimmed with TrimAL with default settings, and concatenated into a supermatrix, and run in RAxML with -PROTGAMMAAUTO setting to select the best-fitting model.

Additional file 8: Figure S7. TreeREx output of a rearrangement analysis highlighting the multiple inversions and transpositions across

heterobranched mitochondrial genomes. Rearrangements shown on branches are delineated as T for transposition and TDL for tandem-duplication-random-loss events. Nodes colored green as consistently reconstructed, red reconstructed with the fallback method (where P0 indicates the chosen solution is not better than other possible solutions).

Additional file 9: Table S2. RogueNaRok leaf instability indices, run with the tree from Figure 1. Lsi_42_max represents the maximum leaf instability across four possible quartets, Lsi_42_ent the entropy between the two most different quartets, and Lsi_42_dif the leaf stability differences between the two most common quartets.

Additional file 10: Figure S8: TreeREx output of a subset of representative mitochondrial genomes from heterobranched gastropods to showcase more general patterns in rearrangements between major clades. Rearrangements shown on branches are delineated as T for transposition and TDL for tandem-duplication-random-loss events. Nodes colored green as consistently reconstructed, red reconstructed with the fallback method (where P0 indicates the chosen solution is not better than other possible solutions).

Additional file 11: Table S3. Mitochondrial gene orders in all taxa from the present study, including outgroups, with < and > indicating directionality and orange boxes indicating possible locations of gene rearrangements.

Abbreviation

mt: Mitochondrial.

Acknowledgements

The authors thank Audrey Falconer of the Museum Victoria, Australia for generously providing specimens of *Microdiscula charopa* and *Ilbia ilbi*. We also thank Denise Akob for suggestions for Figure 3. We thank Deb Crocker and Robert Griffin for assistance with the University of Alabama High-Performance Computing cluster.

Authors' contributions

BB, MAEM, CPM, and MS collected and identified specimens. RMV and CP performed DNA extraction. RMV performed sequencing library preparation and quality control. RMV analyzed the data with help from KMK. All authors contributed to the writing of the manuscript. All authors read and approved the final manuscript.

Funding

This work was funded by start-up funds to KMK from The University of Alabama College of Arts and Sciences and Department of Biological Sciences and Deutsche Forschungsgemeinschaft DFG BR5727/1-1 to BB.

Availability of data and materials

The raw Illumina FASTQ reads generated in this study are available via NCBI SRA under BioProject ID PRJNA626646. Mitochondrial sequences and annotations are also available under NCBI BioProject PRJNA626646. Single gene alignments, data matrices, and tree files are available via FigShare: https://figshare.com/projects/Assessment_of_mitochondrial_genomes_for_heterobranched_gastropod_phylogenetics/80765.

Ethics approval and consent to participate

All samples were invertebrate samples from existing museum collections. No additional animals were collected for this study.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹ Department of Biological Sciences, The University of Alabama, Campus Box 870344, Tuscaloosa, AL 35487, USA. ² SNSB-Bavarian State Collection of Zoology, Münchhausenstr. 21, 81247 München, Germany. ³ Department of Natural History, University Museum of Bergen, University of Bergen,

5020 Bergen, Norway. ⁴ National Museum of Natural History, Smithsonian Institution, 10th St. & Constitution Ave. NW, Washington, D.C. 20560, USA. ⁵ BioGeoCenter LMU (Ludwig Maximilian University Munich), University of Munich, Biozentrum, Großhaderner Str. 2, 82152 Planegg-Martinsried, Germany. ⁶ Alabama Museum of Natural History, Campus Box 870344, Tuscaloosa, AL 35487, USA.

Received: 19 May 2020 Accepted: 26 November 2020

Published online: 21 January 2021

References

- Scouras A, Smith MJ. A novel mitochondrial gene order in the Crinoid Echinoderm *Florometra serratissima*. *Mol Biol Evol*. 2001;18:61–73.
- Boore JL, Staton JL. The mitochondrial genome of the Sipunculid *Phascolopsis gouldii* supports its association with Annelida rather than Mollusca. *Mol Biol Evol*. 2002;19:127–37.
- Lavrov DV, Lang BF. Poriferan mtDNA and animal phylogeny based on mitochondrial gene arrangements. *Syst Biol*. 2005;54:651–9.
- Li Y, Sun X, Hu X, Xun X, Zhang J, Guo X, et al. Scallop genome reveals molecular adaptations to semi-sessile life and neurotoxins. *Nat Commun*. 2017;8:1–11.
- Galaska MP, Li Y, Kocot KM, Mahon AR, Halanych KM. Conservation of mitochondrial genome arrangements in brittle stars (Echinodermata, Ophiuroidea). *Mol Phylogenet Evol*. 2019;130:115–20.
- Talavera G, Vila R. What is the phylogenetic signal limit from mitogenomes? The reconciliation between mitochondrial and nuclear data in the Insecta class phylogeny. *BMC Evol Biol*. 2011;11:315.
- Ranwez V, Harispe S, Delsuc F, Douzery EJP. MACSE: multiple alignment of coding sequences accounting for frameshifts and stop codons. *PLoS ONE*. 2011;6:e22594.
- Wägele JW, Bartolomeaus T. Deep metazoan phylogeny: The backbone of the tree of life, new insights from analyses of molecules, morphology, and theory of data analysis. Berlin, Boston: De Gruyter; 2014. <http://doi.org/https://doi.org/10.1515/9783110277524>.
- McCartney MA, Auch B, Kono T, Mallez S, Zhang Y, Obille A, et al. The genome of the zebra mussel, *Dreissena polymorpha*: a resource for invasive species research. *bioRxiv*. 2019. p. 696732.
- Mikkelsen NT. Phylogeny and systematics of Caudofoveata (Mollusca, Aplacophora). The University of Bergen; 2018. <https://bora.uib.no/handle/1956/17813>. Accessed 13 May 2020.
- Allcock AL, Cooke IR, Strugnell JM. What can the mitochondrial genome reveal about higher-level phylogeny of the molluscan class Cephalopoda? *Zool J Linn Soc*. 2011;161:573–86.
- Uribe JE, Zardoya R. Revisiting the phylogeny of Cephalopoda using complete mitochondrial genomes. *J Molluscan Stud*. 2017;83:133–44.
- Gaitán-Espitia JD, Nespola RF, Opazo JC. The complete mitochondrial genome of the land snail *Cornu aspersum* (Helicidae: Mollusca): intra-specific divergence of protein-coding genes and phylogenetic considerations within Euthyneura. *PLoS ONE*. 2013;8:e67299.
- Williams ST, Foster PG, Littlewood DTJ. The complete mitochondrial genome of a turbinid vetigastropod from MiSeq Illumina sequencing of genomic DNA and steps towards a resolved gastropod phylogeny. *Gene*. 2014;533:38–47.
- Uribe JE, Williams ST, Templado J, Abalde S, Zardoya R. Denser mitogenomic sampling improves resolution of the phylogeny of the superfamily Trochoidea (Gastropoda: Vetigastropoda). *J Molluscan Stud*. 2017;83:111–8.
- Grande C, Templado J, Zardoya R. Evolution of gastropod mitochondrial genome arrangements. *BMC Evol Biol*. 2008;8:61.
- Stöger I, Schrödl M. Mitogenomics does not resolve deep molluscan relationships (yet?). *Mol Phylogenet Evol*. 2013;69:376–92.
- Stöger I, Kocot KM, Poustka AJ, Wilson NG, Ivanov D, Halanych KM, et al. Monoplacophoran mitochondrial genomes: convergent gene arrangements and little phylogenetic signal. *BMC Evol Biol*. 2016;16:274.
- Schrödl M, Jörger KM, Klussmann-Kolb A, Wilson NG. Bye bye “Opisthobranchia”! A review on the contribution of mesopsammic sea slugs to euthyneuran systematics. *Thalassas*. 2011;27:101–12.
- Schrödl M, Stöger I. A review on deep molluscan phylogeny: old markers, integrative approaches, persistent problems. *J Nat History*. 2014;48:2773–804.
- Dinapoli A, Klussmann-Kolb A. The long way to diversity – Phylogeny and evolution of the Heterobranchia (Mollusca: Gastropoda). *Mol Phylogenet Evol*. 2010;55:60–76.
- Jörger KM, Stöger I, Kano Y, Fukuda H, Knebelberger T, Schrödl M. On the origin of Acochlidia and other enigmatic euthyneuran gastropods, with implications for the systematics of Heterobranchia. *BMC Evol Biol*. 2010;10:323.
- Spengel J. Die Geruchsorgane und das Nervensystem der Mollusken. *Zeitschrift für Wissenschaftliche Zoologie*. 1881;35:372.
- Haszprunar G. The Heterobranchia—a new concept of the phylogeny of the higher Gastropoda. *J Zool Syst Evol Res*. 1985;23:15–37.
- Ponder WF, Lindberg DR. Towards a phylogeny of gastropod molluscs: an analysis using morphological characters. *Zool J Linn Soc*. 1997;119:83–265.
- Warèn A, Bouchet P. New records, species, genera, and a new family of gastropods from hydrothermal vents and hydrocarbon seeps. *Zoolog Scr*. 1993;22:1–90.
- Bieler R, Ball AD, Mikkelsen PM. Marine Valvatoidea - Comments on anatomy and systematics with description of a new species from Florida (Heterobranchia: Cornirostridae). *Malacologia*. 1998;17:99.
- Bouchet P, Rocroi J-P. Classification and nomenclator of gastropod families. *Malacologia: International Journal of Malacology*. 2005. <http://www.vliz.be/en/imis?refid=78278>. Accessed 29 Nov 2017.
- Brenzinger B, Wilson NG, Schrödl M. Microanatomy of shelled Koloonia cf. minutissima (Laseron, 1951) (Gastropoda: ‘lower’ Heterobranchia: Muriculinellidae) does not contradict a sister-group relationship with enigmatic Rhodopomorphs slugs. *J Molluscan Stud*. 2014;2014(80):518–40.
- Wilson NG, Jörger KM, Brenzinger B, Schrödl M. Phylogenetic placement of the enigmatic worm-like Rhodopomorph slugs as basal Heterobranchia. *J Molluscan Stud*. 2017;83:399–408.
- Kocot KM, Halanych KM, Krug PJ. Phylogenomics supports Panpulmonata: Opisthobranch paraphyly and key evolutionary steps in a major radiation of gastropod molluscs. *Mol Phylogenet Evol*. 2013;69:764–71.
- Kano Y, Brenzinger B, Nützel A, Wilson NG, Schrödl M. Ringicolid bubble snails recovered as the sister group to sea slugs (Nudipleura). *Scientific Reports*. 2016;6:30908.
- Podsiadlowski L, Mwinyi A, Lesný P, Bartolomeaus T. Mitochondrial gene order in Metazoa—theme and variations. In: Deep Metazoan Phylogeny: The Backbone of the Tree of Life, New insights from analyses of molecules, morphology, and theory of data analysis. Berlin, Boston: De Gruyter; 2014. <http://doi.org/https://doi.org/10.1515/9783110277524>.
- Fretter V. The Structure and life history of some minute prosobranchs of rock pools: *Skeneopsis planorbis* (Fabricius), *Omalogyra atomus* (Philippi), *Rissoella diaphana* (Alder) and *Rissoella opalina* (Jeffreys). *J Mar Biol Assoc UK*. 2019;27:597–632.
- Klussmann-Kolb A, Dinapoli A, Kuhn K, Streit B, Albrecht C. From sea to land and beyond—new insights into the evolution of euthyneuran Gastropoda (Mollusca). *BMC Evol Biol*. 2008;8:57.
- Malaquias MAE, Reid DG. Tethyan vicariance, relictualism and speciation: evidence from a global molecular phylogeny of the opisthobranch genus *Bulla*. *J Biogeogr*. 2009;36:1760–77.
- Zapata F, Wilson NG, Howison M, Andrade SCS, Jörger KM, Schrödl M, et al. Phylogenomic analyses of deep gastropod relationships reject Orthogastropoda. *Proc R Soc B*. 2014;281:20141739.
- Ayyagari VS, Sreerama K. Molecular phylogeny and evolution of Pulmonata (Mollusca: Gastropoda) on the basis of mitochondrial (16S, COI) and nuclear markers (18S, 28S): an overview. *J Genet*. 2020;99:17.
- Haszprunar G, Speimann E, Hawe A, Heß M. Interactive 3D anatomy and affinities of the Hyalogyrinidae, basal Heterobranchia (Gastropoda) with a rhipidoglossate radula. *Org Divers Evol*. 2011;11:201–36.
- Brenzinger B, Haszprunar G, Schrödl M. At the limits of a successful body plan - 3D microanatomy, histology and evolution of Helminthope (Mollusca: Heterobranchia: Rhodopomorphs), the most worm-like gastropod. *Front Zool*. 2013;10:37.
- Healy JM. Comparative sperm ultrastructure and spermiogenesis in basal heterobranch gastropods (Valvatoidea, Architectonicoidea, Rissoelloidea, Omalogyroidea, Pyramidelloidea) (Mollusca). *Zoolog Scr*. 1993;22:263–76.

42. Frýda J, Nützel A, Wagner J. Paleozoic gastropod phylogeny. In: Phylogeny and evolution of the Mollusca. New York: University of California Press; 2008. p. 239.
43. Gründel J, Nützel A. On the early evolution (Late Triassic to Late Jurassic) of the Architectibranchia (Gastropoda: Heterobranchia), with a provisional classification. *Neues Jahrbuch für Geologie und Paläontologie*. 2012;264:31–59.
44. Bandel K. Phylogeny of the Caecidae (Caenogastropoda). Institut, Universität Hamburg. 1996;79:53–115.
45. Wägele H, Klussmann-Kolb A, Vonnemann V, Medina M, Heterobranchia I. The Opisthobranchia. In: Phylogeny and evolution of the Mollusca. Berkeley: University of California Press; 2008.
46. Cunha TJ, Giribet G. A congruent topology for deep gastropod relationships. *Proc R Soc B Biol Sci*. 2019;286:20182776.
47. White TR, Conrad MM, Tseng R, Balayan S, Golding R, de Frias Martins AM, et al. Ten new complete mitochondrial genomes of pulmonates (Mollusca: Gastropoda) and their impact on phylogenetic relationships. *BMC Evol Biol*. 2011;11:295.
48. Dayrat B, Conrad M, Balayan S, White TR, Albrecht C, Golding R, et al. Phylogenetic relationships and evolution of pulmonate gastropods (Mollusca): New insights from increased taxon sampling. *Mol Phylogenet Evol*. 2011;59:425–37.
49. Romero PE, Weigand AM, Pfenninger M. Positive selection on panpulmonate mitogenomes provide new clues on adaptations to terrestrial life. *BMC Evol Biol*. 2016;16:164.
50. Shao R, Dowton M, Murrell A, Barker SC. Rates of gene rearrangement and nucleotide substitution are correlated in the mitochondrial genomes of insects. *Mol Biol Evol*. 2003;20:1612–9.
51. Wolstenholme DR. Animal mitochondrial DNA: structure and evolution. *Int Rev Cytol*. 1992;141:173–216.
52. Nolan JR, Bergthorsson U, Adema CM. *Physella acuta*: atypical mitochondrial gene order among panpulmonates (Gastropoda). *J Molluscan Stud*. 2014;80:388–99.
53. Knutson VL, Brenzinger B, Schrödl M, Wilson NG, Giribet G. Most Cephalaspidea have a shell, but transcriptomes can provide them with a backbone (Gastropoda: Heterobranchia). *Mol Phylogenet Evol*. 2020;9:22.
54. Frýda J, Racheboeuf PR, Frýdová B, Ferrová L, Mergl M, Berkyová S. Platyceratiid gastropods - stem group of patellogastropods, neritimorphs or something else? *Bull Geosci*. 2009;1:107–20.
55. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*. 2012;19:455–77.
56. Boisvert S, Laviolette F, Corbeil J. Ray: Simultaneous assembly of reads from a mix of high-throughput sequencing technologies. *J Comput Biol*. 2010;17:1519–33.
57. Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsch G, et al. MITOS: Improved de novo metazoan mitochondrial genome annotation. *Mol Phylogenet Evol*. 2013;69:313–9.
58. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 2009;25:1972–3.
59. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*. 2018;35:1547–9.
60. Kück P, Meusemann K. FASconCAT: Convenient handling of data matrices. *Mol Phylogenet Evol*. 2010;56:1115–8.
61. Anisimova M, Liberles DA, Philippe H, Provan J, Pupko T, von Haeseler A. State-of-the-art methodologies dictate new standards for phylogenetic analysis. *BMC Evol Biol*. 2013;13:161.
62. Criscuolo A, Gribaldo S. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol Biol*. 2010;10:210.
63. Stamatakis A. RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;30:1312–3.
64. Aberer AJ, Krompass D, Stamatakis A. Pruning rogue taxa improves phylogenetic accuracy: an efficient algorithm and webservice. *Syst Biol*. 2013;62:162–6.
65. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams M, von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol*. 2020;37:1530–4.
66. Lanfear R, Calcott B, Kainer D, Mayer C, Stamatakis A. Selecting optimal partitioning schemes for phylogenomic datasets. *BMC Evol Biol*. 2014;14:82.
67. Gadagkar SR, Rosenberg MS, Kumar S. Inferring species phylogenies from multiple genes: concatenated sequence tree versus consensus gene tree. *J Exp Zool*. 2005;304B:64–74.
68. Crotty S, Minh B, Bean N, Holland B, Tuke J, Jeremiin L, et al. GHOST: recovering historical signal from heterotachously evolved sequence alignments. *Syst Biol*. 2020;69:22.
69. Lartillot N, Rodrigue N, Stubbs D, Richer J. PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Syst Biol*. 2013;62:611–5.
70. Bernt M, Merkle D, Middendorf M. An algorithm for inferring mitogenome rearrangements in a phylogenetic tree. In: Nelson CE, Vialette S, editors. *Comparative Genomics*. Springer: Berlin; 2008. p. 143–57.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

