

Investigating Sources of Conflict in Deep Phylogenomics of Vetigastropod Snails

 TAUANA JUNQUEIRA CUNHA^{1,2,*},  JAMES DAVIS REIMER^{3,4}, AND  GONZALO GIRIBET¹

¹*Museum of Comparative Zoology, Department of Organismic and Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge MA 02138, USA*; ²*Smithsonian Tropical Research Institute, Panama City, Panama*; ³*Molecular Invertebrate Systematics and Ecology, University of the Ryukyus, 1 Senbaru, Nishihara, Okinawa 903-0213, Japan*; and ⁴*Tropical Biosphere Research Center, University of the Ryukyus, 1 Senbaru, Nishihara, Okinawa 903-0213, Japan*

*Correspondence to be sent to: *Museum of Comparative Zoology, Department of Organismic and Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge MA 02138, USA*; *E-mail: CunhaT@si.edu*.

Received 24 August 2020; reviews returned 25 August 2021; accepted 27 August 2021

Associate Editor: Danielle Edwards

Abstract.—Phylogenetic analyses may suffer from multiple sources of error leading to conflict between genes and methods of inference. The evolutionary history of the mollusc clade Vetigastropoda makes them susceptible to these conflicts, their higher level phylogeny remaining largely unresolved. Originating over 350 Ma, vetigastropods were the dominant marine snails in the Paleozoic. Multiple extinction events and new radiations have resulted in both very long and very short branches and a large extant diversity of over 4000 species. This is the perfect setting of a hard phylogenetic question in which sources of conflict can be explored. We present 41 new transcriptomes across the diversity of vetigastropods (62 terminals total), and provide the first genomic-scale phylogeny for the group. We find that deep divergences differ from previous studies in which long branch attraction was likely pervasive. Robust results leading to changes in taxonomy include the paraphyly of the order Lepetellida and the family Tegulidae. Tectinae subfam. nov. is designated for the clade comprising *Tectus*, *Cittarium*, and *Rochia*. For two early divergences, topologies disagreed between concatenated analyses using site heterogeneous models versus concatenated partitioned analyses and summary coalescent methods. We investigated rate and composition heterogeneity among genes, as well as missing data by locus and by taxon, none of which had an impact on the inferred topologies. We also found no evidence for ancient introgression throughout the phylogeny. We further tested whether uninformative genes and over-partitioning were responsible for this discordance by evaluating the phylogenetic signal of individual genes using likelihood mapping, and by analyzing the most informative genes with a full multispecies coalescent (MSC) model. We find that most genes are not informative at the two conflicting nodes, but neither this nor gene-wise partitioning are the cause of discordant results. New method implementations that simultaneously integrate amino acid profile mixture models and the MSC might be necessary to resolve these and other recalcitrant nodes in the Tree of Life. [Fissurellidae; Haliotidae; likelihood mapping; multispecies coalescent; phylogenetic signal; phylogenomic conflict; site heterogeneity; Trochoidea.]

Two major goals of systematic biology are to understand the evolutionary relationships of organisms and the sources of discordance when conflicting results are identified. While there are biologically relevant sources of discordance such as introgression and incomplete lineage sorting (ILS), these can also be obscured by various forms of systematic error, such as lack of resolution of individual loci and inference error. Analytical methods have greatly improved in how to deal with sources of error in phylogenomic datasets, for example, more complex models of evolution account for sequence heterogeneity in concatenated matrices (Lartillot and Philippe 2004), and multispecies coalescent (MSC) methods account for ILS and the resulting discordance in gene tree histories (Ogilvie et al. 2017). However, full MSC methods that simultaneously infer gene trees and the species tree are still computationally limited to relatively small datasets (Ogilvie et al. 2017). Summary methods that use gene tree topologies to infer the species tree (Mirarab and Warnow 2015) have thus been predominant in studies with taxon- and gene-rich datasets, potentially carrying artifacts from erroneous gene tree inference (Gatesy and Springer 2014; Meiklejohn et al. 2016). Here, we present the first comprehensive phylogenomic framework for the clade Vetigastropoda, exploring multiple strategies to minimize error and investigating the sources

of phylogenomic conflict at deep nodes in their species tree.

Abalones, turban snails, top shells, keyhole limpets, and slit shells are just some of the diverse Vetigastropoda (Salvini-Plawen 1980). With over 4000 living species (WoRMS 2021) and many thousands more in the fossil record, vetigastropods comprise one of the five major lineages of Gastropoda (Cunha and Giribet 2019). They are all marine and occupy a wide variety of habitats, from shallow hard substrates to deep sea vents and cold seeps. Some shallow water species are high value food items for human populations around the globe (Leiva and Castilla 2002; Ab Lah et al. 2017), while others are natural sources of unique and heavily used proteins (e.g. hemocyanin) in immunological applications (Harris and Markl 1999; Mora Román et al. 2019). Vetigastropods were the dominant clade of gastropods throughout the Paleozoic and most of the Mesozoic eras, with fossils going back at least to the Silurian (Fryda et al. 2008), and the divergence of crown groups being estimated as in the Devonian (Zapata et al. 2014). Despite their diversity, evolutionary importance, and applications for human food and health, the phylogeny of vetigastropods remains contentious, likely due to basal divergences being ancient and the small amount of molecular data that has been available thus far.

TABLE 1. Classification from [Bouchet et al. \(2017\)](#) for orders, superfamilies, and families of the subclass Vetigastropoda, and new proposed classification based on the results of this study.

Bouchet et al.(2017)	Proposed here
Lepetellida	Lepetellida
Lepetelloidea	Lepetelloidea
Lepetellidae	Lepetellidae
Addisoniidae	Addisoniidae
Bathyphytophilidae	Bathyphytophilidae
Caymanabyssiidae	Caymanabyssiidae
Cocculinellidae	Cocculinellidae
Osteopeltidae	Osteopeltidae
*Pseudococculinidae	Pseudococculinidae
*Pyropeltidae	Pyropeltidae
Lepetodriloidea	Lepetodriloidea
*Leptodriliidae	Leptodriliidae
Sutilizonidae	Sutilizonidae
Scissurelloidea	Scissurelloidea
*Scissurellidae	Scissurellidae
Anatomidae	Anatomidae
Depresszonidae	Depresszonidae
Larocheidae	Larocheidae
Fissurelloidea	Fissurellida stat. nov.
*Fissurellidae	Fissurelloidea
Haliotoidea	Haliotida stat. nov.
*Haliotidae	Haliotidae
Trochida	Trochida
Trochoidea	Trochoidea
*Trochidae	Trocidae
*Angariidae	Angariidae
*Areneidae	Areneidae
*Calliostomatidae	Calliostomatidae
*Colloniidae	Colloniidae
Conradiidae	Conradiidae
*Liotiidae	Liotiidae
Margaritidae	Margaritidae
*Phasianellidae	Phasianellidae
*Skeneidae	Skeneidae
Solariellidae	Solariellidae
*Tegulidae	Turbinidae
*Turbinidae	Prisogasterinae
Pleurotomariida	Pleurotomariida
Pleurotomarioidea	Pleurotomarioidea
*Pleurotomariidae	Pleurotomariidae
Seguenziida	Seguenziida
Seguenzioidea	Seguenzioidea
Seguenziidae	Seguenziidae
Cataegidae	Cataegidae
*Chilodontaidae	Chilodontaidae
Choristellidae	Choristellidae
Eucyclidae	Eucyclidae
Eudaroniidae	Eudaroniidae
Pendromidae	Pendromidae
Trochaclididae	Trochaclididae

Note: Changes in the proposed classification are highlighted in bold. The only clades listed below the family level are related to proposed changes, in which taxa that were previously part of the family Tegulidae are here transferred to Turbinidae (see the last part of the Discussion for details). Families marked with an asterisk were sampled in the phylogeny.

Eight superfamilies of vetigastropods are currently accepted, containing 38 extant families ([Bouchet et al. 2017](#); [WoRMS 2021](#)) (Table 1). The first comprehensive

morphological analyses established classifications and identified synapomorphies within the group ([Salvini-Plawen and Haszprunar 1987](#); [Haszprunar 1988](#); [Ponder and Lindberg 1997](#); [Sasaki 1998](#)). On the molecular side, the first studies provided incremental contributions to solving vetigastropod relationships ([Harasewych et al. 1997](#); [Geiger and Thacker 2005](#); [Yoon and Kim 2005](#)), and key publications with densely sampled phylogenies based on a handful of genes helped place not only the most diverse groups but also minute and hard-to-collect taxa ([Kano 2008](#); [Williams et al. 2008](#); [Aktipis and Giribet 2012](#)). The first vetigastropod transcriptomes contributed more broadly to gastropod relationships ([Zapata et al. 2014](#)). Finally, the most recent publications have used mitochondrial genomes to target deep vetigastropod divergences ([Lee et al. 2016](#); [Uribe et al. 2016](#); [Uribe et al. 2017](#); [Wort et al. 2017](#); [Guo et al. 2020](#)). Trochoidea, the most diverse superfamily, has also received considerable attention ([Williams and Ozawa 2006](#); [Williams et al. 2008](#); [Williams et al. 2010](#); [Williams 2012](#); [Uribe et al. 2017](#); [Guo et al. 2020](#)). Despite being numerous, these efforts have led to very few consistent results. Most datasets have nonoverlapping taxon representation and do not have enough sequence data to resolve such ancient divergences. In addition, long branches are often present, leading to long branch attraction (LBA) artifacts ([Uribe et al. 2016](#); [Uribe et al. 2019](#)). Particularly problematic taxa have been Lepetelloidea, Lepetodriloidea, Fissurellidae and Haliotidae ([Williams et al. 2008](#); [Aktipis and Giribet 2012](#); [Lee et al. 2016](#); [Uribe et al. 2016](#); [Uribe et al. 2017](#)). Familial relationships within Trochoidea have been redefined recently with relatively good support ([Uribe et al. 2017](#); [Guo et al. 2020](#)), pending the inclusion of several unsampled families.

With an extensive sampling of 41 new transcriptomes and previously published data, here we targeted deep divergences in the vetigastropod phylogeny. Our sampling covers all superfamilies and about half of the families of vetigastropods (Table 1). We minimized systematic errors by subsampling genes based on evolutionary rates, composition heterogeneity and missing data, and used both concatenation and coalescent-based approaches. Our datasets and analyses were able to resolve the majority of deep relationships with well supported and congruent topologies. Where methods disagree, we tested whether gene-wise partitioning, uninformative genes, or introgression at deep nodes could be the cause of discordance.

MATERIALS AND METHODS

Sampling and Sequencing

We sequenced the transcriptomes of 41 vetigastropod genera and added data from another eight genera with previously published sequences, for a total of 49 ingroup terminals. We further used 12 other gastropods and one bivalve as outgroups, for a total of 62 terminals. This sampling covers all eight superfamilies of vetigastropods, and 18 of the 38 accepted families.

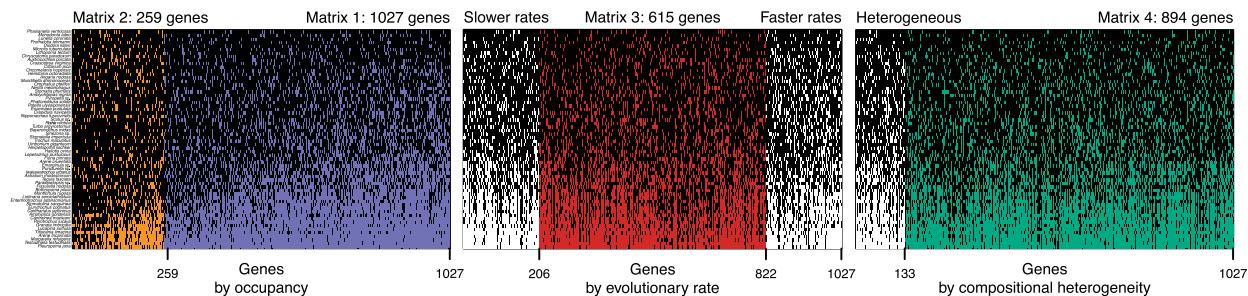


FIGURE 1. Strategies of gene subsampling to infer vetigastropod relationships. Matrix 1: 50% taxon occupancy; all other matrices are subsets of this one. Matrix 2: 70% taxon occupancy. Genes and species are sorted with the best sampling on the upper left. Matrix 3: the 20% slowest and fastest evolving genes are removed. Matrix 4: genes with heterogeneous amino acid composition are removed. Black cells indicate genes present for each species.

All new data and selected published sequences are paired-end Illumina reads. New samples were initially identified based on Okutani (2000), and fixed in RNAlater (Invitrogen). RNA extraction and mRNA isolation were done with the TRIzol Reagent and Dynabeads (Invitrogen). Libraries were prepared with the PrepX RNA-Seq Library kit using the Apollo 324 System (Wafergen). Quality control of mRNA and cDNA was done with a 2100 Bioanalyzer, a 4200 TapeStation (Agilent) and the Kapa Library Quantification kit (Kapa Biosystems). Samples were pooled in equimolar amounts and sequenced in the Illumina HiSeq 2500 platform (paired end, 150 bp) at the Bauer Core Facility at Harvard University. Voucher information, library indexes and assembly statistics are in Supplementary Tables S1 and S2 available on Dryad at <https://doi.org/10.5061/dryad.rxwdbrv64>.

Transcriptome Assembly and Orthology

For transcriptome assembly, we followed the pipeline described in detail in Cunha and Giribet (2019), using the same scripts and software specifications. In summary, we cleaned raw reads with Rcorrector (Song and Florea 2015) and Trim Galore! (Krueger et al. 2018), and removed mitochondrial DNA and ribosomal RNAs with Bowtie2 v2.2.9 (Langmead and Salzberg 2012). Filtered reads were assembled *de novo* with Trinity v2.3.2 (Grabherr et al. 2011; Haas et al. 2013), duplicated transcripts were removed with CD-HIT-EST v4.6.4 (Fu et al. 2012), and assemblies were translated to amino acids with TRANSDECODER v3.0 (Haas et al. 2013), keeping the longest isoform of each gene. The completeness of the assemblies was evaluated with BUSCO v3.0.2 by comparison with the Metazoa database (Simão et al. 2015) and with TransRate (Smith-Unna et al. 2016).

Orthology assignment of the peptide assemblies was done with OMA v2.2.0 (Altenhoff et al. 2018). After orthology, all orthogroups for which at least half of the terminals were represented (50% taxon occupancy) were retained, resulting in a reference matrix 1 with 1027 genes. From this reference matrix, a subset of 259 genes with 70% taxon occupancy constituted matrix 2 (Fig. 1).

Each orthogroup was aligned with MAFFT v7.309 (Katoh and Standley 2013) (–auto –amino), and the ends of the alignments were trimmed to remove positions with more than 80% missing data. Scripts used for selecting orthogroups and trimming the alignments are available in Cunha and Giribet (2019).

Accounting for Sequence Heterogeneity

To avoid possible biases from genes evolving at the ends of the spectrum of substitution rates, matrix 3 was built by removing from matrix 1 the 20% slowest and the 20% fastest evolving genes, as calculated with TrimAI (Capella-Gutiérrez et al. 2009), for a final size of 615 genes (Fig. 1). To avoid model misspecification caused by heterogeneity in amino acid composition, matrix 4 was restricted to the subset of 894 genes from matrix 1 that were homogeneous (Fig. 1). Homogeneity for each gene was determined with a simulation-based test from the python package p4 (Foster 2004) and a conservative *P*-value of 0.1. Besides these four standard amino acid matrices, we further reduced compositional heterogeneity in matrices 1 and 2 by recoding amino acids into the six Dayhoff categories (Dayhoff et al. 1978). Scripts used for the homogeneity test and recoding of the dataset are available in Cunha and Giribet (2019).

Phylogenetic Analyses

For inference methods that require concatenation, genes were concatenated using Phyutility (Smith and Dunn 2008). Amino acid matrices were used for phylogenetic inference with a coalescent-based approach in Astral-II v4.10.12 (Mirarab and Warnow 2015), with maximum likelihood (ML) in IQ-TREE MPI v1.5.5 and v1.6.8 (Nguyen et al. 2015; Chernomor et al. 2016; Kalyaanamoorthy et al. 2017), and with Bayesian inference (BI) in PhyloBayes MPI v1.7a (Lartillot et al. 2013). The two Dayhoff-recoded matrices were analyzed in PhyloBayes. For the coalescent-based method, gene trees were inferred with RAxML v8.2.10 (Stamatakis 2014) (-N 10 -m PROTGAMMALGF) and then used as input for Astral-II for species tree estimation. For

each concatenated matrix, we inferred the best ML tree with two strategies: a partitioned analysis with model search including LG4 mixture models and accounting for heterotachy (-st AA -msub nuclear -ninit 10 -bb 1500 -sp partition_file -m MFP+MERGE -rcluster 10 -madd LG4M,LG4X -mrate G,R,E) (the best-fit partitioning scheme maintained all genes as separate partitions); and a non-partitioned analysis with model search including the LG and WAG rate matrices with a profile mixture model (Le et al. 2008), an ML variant of the Bayesian CAT model (Lartillot and Philippe 2004) (-st AA -msub nuclear -ninit 10 -bb 1500 -m MFP -mset LG,WAG -rcluster 10 -mfreq F+C60 -mrate G,R). Cluster computer memory limited which of the profile models (C10 to C60) could be used depending on the size of the matrix. The following best-fit models were selected for each matrix: WAG+F+C20+R10 (matrices 1 and 4), WAG+F+C60+R7 (matrix 2), WAG+F+C30+R7 (matrix 3). PhyloBayes was run with default priors and the CAT-GTR model on matrices 1 and 2 (both amino acid and Dayhoff-recoded versions), discarding constant sites to speed up computation. Tree figures were made with the R packages *ggtree* (Yu et al. 2017), *treeio* (Wang et al. 2020), and *phytools* (Revell 2012); an R notebook with code for tree figures is available as [Supplementary Code S1](#) available on Dryad.

Accounting for Species with High Missing Data

We tested whether species with more missing data were adversely affecting phylogenetic inference in two ways. In the first test, the 13 taxa with most missing data (bottom rows of Fig. 1, with 60–91% missing genes) were removed from matrix 1, which we analyzed again under the same unpartitioned strategy and best-fit profile mixture model in IQ-TREE. In the second approach, we ran a neighbor-joining (NJ) analysis on a matrix of presence/absence of the 1027 genes in the R package *ape* (Paradis and Schliep 2019). If patterns of missing data were driving phylogenetic inference, we would expect the distance-based tree to match sequence-based results. Code for NJ analysis and tree figures is available in [Supplementary Code S1](#) available on Dryad.

Phylogenetic Signal and Hypothesis Testing

Following the phylogenetic analyses described above, two deep nodes showed conflicting results between analyses based on information from individual genes (Astral and partitioned ML) and analyses of the concatenated datasets using profile mixture models (unpartitioned ML and BI). Gene tree estimation error is a known source of erroneous inference in summary coalescent methods (Molloy and Warnow 2018). RAxML gene trees, for example, will be fully resolved even if there is not enough phylogenetic signal to resolve them. It has also been shown that overpartitioning in concatenated datasets can lead to serious long-branch attraction biases (Wang et al. 2019). We therefore hypothesized that the

discordance between our results could be due to gene-wise partitioning and uninformative genes/gene trees. We predicted that a full Bayesian MSC analysis on a subset of the most informative genes would recover the same topology as the concatenated analyses with site heterogeneous models.

First, we tested whether individual genes contained enough information to resolve between alternative topologies at the two conflicting nodes with likelihood-mapping (LMAP) (Strimmer and von Haeseler 1997) as implemented in IQ-TREE v1.6.8. For each of the two recalcitrant nodes, four clusters of taxa were defined (Pleurotomarioidea, Lepetellida s.s., Trochoidea, and Haliotoidea or Fissurelloidea; other terminals were ignored) (Fig. 2a,b). Topologies 1 and 2 were the ones of interest; the third topology was also possible given the unrooted tree with four clusters but had not been recovered by any of the previous analyses. From the 1027 genes in matrix 1, 596 and 835 genes, respectively, had at least one taxon in each of the four clusters. LMAP was run on each gene alignment using all unique quartets of terminals (-lmap ALL -lmclust clusters.nex -n 0 -st AA -msub nuclear -m LG+G+F) to obtain the quartet support for each area of the likelihood maps. Corners of the triangle maps indicate quartets of terminals that are informative toward either of the three topologies; edges represent quartets that are partly informative between two of the topologies; and the center represents star-like, non-informative quartets (Fig. 2d).

We then built two more matrices in which genes were selected based on their percentage of informative quartets for both recalcitrant nodes, regardless of which topologies those quartets supported. Matrix 5 comprised 80 genes with more than 70% of resolved quartets, and matrix 6 had 44 genes with more than 75% of resolved quartets (Fig. 2c). Because not all taxa were represented in each gene alignment, the total number of quartets varied between genes, but this variation did not affect which genes were more informative and therefore selected for the new matrices ([Supplementary Fig. S1](#) available on Dryad). On both matrices we ran a full MSC model in StarBEAST2 v0.15.13 (Ogilvie et al. 2017) in BEAST v2.6.3 (Bouckaert et al. 2019), inferring gene trees and the species tree simultaneously. Non-default settings were: linked WAG site model with eight gamma categories; unlinked uncorrelated lognormal clock with estimated average rate and standard deviation; birth-death model for the species tree prior; exponential with mean 1 for the diversification rate prior; log normal with mean 0.1 and standard deviation 1 for the average population size prior; outgroups were removed and monophyletic constraints were set on the well-established Pleurotomariidae and its sister clade. Analyses were run for 1–3.5 billion generations sampling every 100,000 generations. The convergence of species tree parameters was confirmed in Tracer v1.6 (Rambaut et al. 2018), and posterior trees were summarized into a maximum clade credibility tree using median heights with TreeAnnotator v2.6 (Drummond et al. 2012). Code

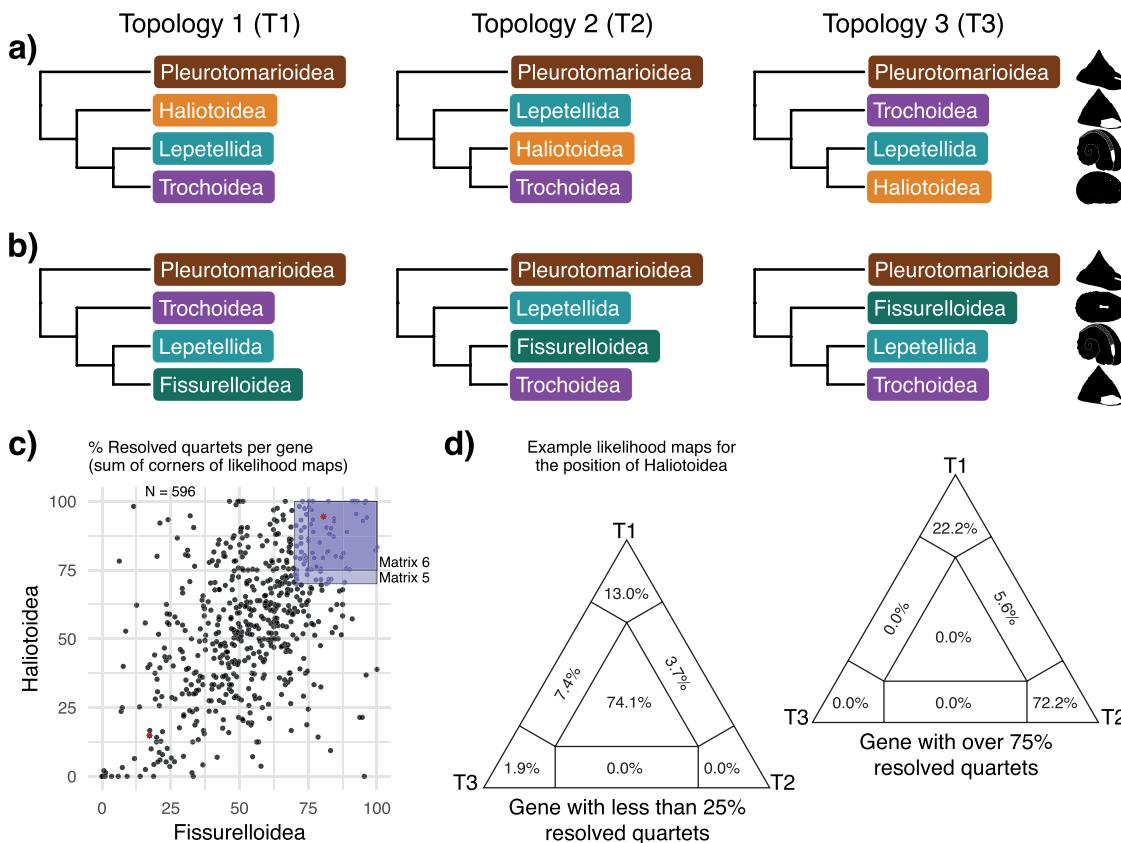


FIGURE 2. Likelihood mapping analyses (LMAP). a, b) Possible topologies for the position of Haliotoidea and Fissurelloidea based on four clusters of taxa. Topologies 1 and 2 are of interest, having been recovered by multiple analyses in this study. LMAP was done on 596 genes for Haliotoidea and 835 for Fissurelloidea, each containing at least one terminal from each cluster. c) Distribution of genes with different amounts of resolved quartets for both recalcitrant nodes. For each gene, this is the sum of the corners of the likelihood maps. Two asterisks correspond to example maps illustrated in (d). Shaded squares comprise genes with the highest number of resolved quartets, composing matrix 5 and 6 (at least 70% and 75%, respectively). d) Example likelihood maps for the position of Haliotoidea from two genes with contrasting distributions of quartets. Silhouettes by Tauana Cunha, available at phylopic.org.

for figures related to LMAP analyses is available in [Supplementary Code S2](#) available on Dryad.

Gene Support for Alternative Topologies

Besides the standard measures of support provided by each inference method, we calculated gene and site concordance factors (gCF, sCF) in IQ-TREE v2.1.2 ([Minh et al. 2020a](#); [Minh et al. 2020b](#)). Concordance factors were calculated on both the partitioned and unpartitioned ML trees from matrix 1 (-t tree_file -gcf-gene_trees -p alignments_folder -scf 100 -seed 13 -cf-quartet). Partitioned coalescence support (PCS) ([Gatesy et al. 2019](#)) was also calculated to evaluate conflicting signal for alternative topologies and to identify potential outlier gene trees that could have a disproportionate effect on summary coalescent analyses. PCS was run on the Astral tree from matrix 1 with the unpartitioned ML tree as the alternative tree. Code for the related figures is available in [Supplementary Code S3](#) available on Dryad.

We further used the results of the LMAP analyses described above to investigate whether genes with

specific properties preferably support either one of the alternative topologies. For each gene in LMAP analyses, we identified the area of the likelihood map with the highest quartet support, then plotted the distribution of genes supporting each topology while discerning groups of genes by their category of evolutionary rate, compositional heterogeneity, and occupancy. Because these were the criteria for building the initial four matrices, this evaluation also showed which sets of genes supporting each topology were retained or excluded in matrices 1–4.

Introgression

We tested whether introgression in ancient lineages could be responsible for the conflict between the two phylogenies obtained in the initial analyses. The procedure described below was applied twice, checking for signs of introgression on either resolution of the vetigastropod species tree ([Supplementary Code S4](#) available on Dryad). Originally described by [Huson et al. \(2005\)](#) and recently adapted by [Vanderpool et al. \(2020\)](#) to detect introgression at deeper timescales than

most available approaches, the method calculates the test statistic Δ . Based on the distribution of gene trees, Δ captures the deviation from the expected equal amounts of the two most frequent alternative topologies for any given branch. Among the 1027 genes in matrix 1, the number of gene trees supporting alternative resolutions for each internal branch of the species tree was taken from the gene concordance factors described in the section above. Δ was calculated as the absolute value of the difference between the number of gene trees supporting alternative topology 1 and alternative topology 2, divided by the sum of the same two numbers. To test whether these observed values of Δ were significantly higher than expected by chance (and therefore indicative of introgression), we generated a null distribution based on 2000 resamplings of the 1027 gene trees with replacement. Gene concordance factors on the original species tree were obtained for each resampling, and Δ was calculated in the same way for all branches of each resampled dataset. The observed Δ were then transformed in standardized Z-scores for each branch as the difference between the observed Δ and the mean of the null distribution, divided by the standard deviation of the null distribution. We tested all branches for which more than 5% of gene trees were discordant with the species tree (43 out of 59 internal branches), including the two target recalcitrant nodes, and at the end calculated the *P*-value of each observed Z-score. For this one-tailed test, evidence of introgression would be indicated by Z-scores of at least 1.65 at a threshold *P*-value of 0.05.

RESULTS

Sources of Phylogenomic Conflict

The 16 topologies initially inferred from matrices 1–4 were congruent and fully supported, regardless of matrix composition and inference method, for all but two of the deeper nodes in the vetigastropod phylogeny (Fig. 3, *Supplementary Figs. S2–S5* available on Dryad). The two recalcitrant nodes concern the position of the superfamilies Haliotoidea and Fissurelloidea, each with a single family (Haliotidae and Fissurellidae, respectively). A range of potential sources of systematic error related to gene content were accounted for, none of which impacted the results, as evidenced by congruence across matrices subsampled by heterogeneity in evolutionary rates, heterogeneity in amino acid composition, and occupancy. Highly incomplete taxa also had no adverse effect on the results, with the same topology being recovered for the remaining taxa after removal of the terminals with most missing data (*Supplementary Fig. S6a* available on Dryad). This was further supported by the fact that the neighbor-joining (NJ) tree on the presence/absence of genes did not resemble any of the results based on sequence data (*Supplementary Fig. S6b* available on Dryad), as would be expected if patterns of missing data instead of phylogenetic signal were driving the results. This is illustrated by terminals from any given

vetigastropod family being spread across the NJ tree, and species with poor gene sampling grouping in a cluster, instead of being recovered with their closest relatives in other parts of the tree (*Supplementary Fig. S6b* available on Dryad).

Results for the two conflicting nodes varied according to the inference method and model, with BI and ML analyses using profile mixture models of amino acid frequencies recovering a different topology compared to the summary coalescent method and ML analyses with gene-wise partitioning (Fig. 3). LMAP tests on individual genes showed that most genes in this transcriptomic dataset were not informative enough to resolve those two internal branches (Fig. 2c), with most genes having small percentages of resolved quartets. Criteria to determine what is an acceptable amount of unresolved quartets are arbitrary, but even about 8% of unresolved quartets can be too much in some cases (Strimmer and von Haeseler 1997), meaning that few genes in this dataset were informative enough to resolve the two conflicting basal splits in the vetigastropod tree. We selected two thresholds (over 70% and over 75% of resolved quartets) to subsample the dataset into further matrices composed only of the most resolved genes. We hypothesized that incongruence of results was due to over-partitioning and uninformative genes, and predicted that a full MSC model on these subsets would recover a topology congruent with analyses based on profile mixture models (Fig. 3a, left). Contrary to our prediction, StarBEAST2 trees did not support topology 1 (*Supplementary Fig. S7* available on Dryad). The position of Haliotidae remained the same in full coalescence trees as it was in Astral analyses based on gene trees and in the partitioned ML analyses, while the position of Fissurellidae in StarBEAST2 trees was resolved as in the concatenated analyses with the best models of site heterogeneity (Fig. 3b, *Supplementary Fig. S7* available on Dryad).

When looking at which topology was preferentially supported by each gene in the LMAP tests, we found that the relative proportion of genes supporting topology 2 was higher in the set of most informative genes than among all genes included in the LMAP analyses (*Supplementary Fig. S8* available on Dryad). Categorizing the genes by their specific properties revealed that each topology, as well as the unresolved areas of likelihood maps, were equally supported by similar proportions of genes with different characteristics (*Supplementary Fig. S8a–c* available on Dryad), such as evolutionary rate and heterogeneity in amino acid composition, further emphasizing that these were not the sources of discordance.

With the partitioned coalescent support, we identified four genes that could have had an especially strong effect on summary coalescent analyses (*Supplementary Fig. S9a* available on Dryad). Nonetheless, removal of such gene trees did not alter the result of Astral (*Supplementary Fig. S9c* available on Dryad). The distribution of concordance factors across branches of topologies 1 and 2 is further evidence that the two recalcitrant nodes are indeed some of the hardest to

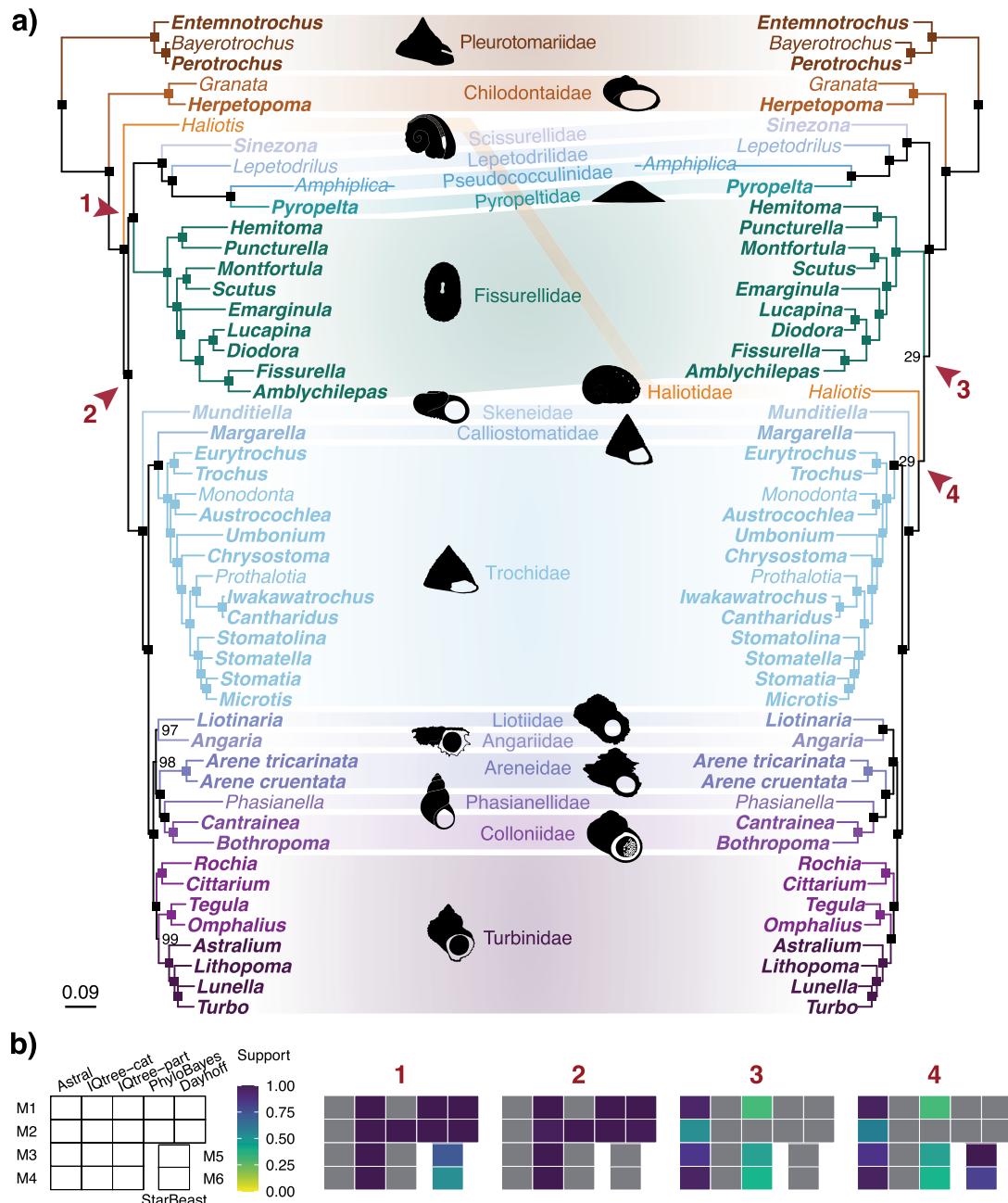


FIGURE 3. Vetigastropod phylogeny and support values across inference methods for key nodes. Family assignment follows the changes in classification proposed in this study, with the two clades that previously comprised Tegulidae marked as Turbinidae in the trees. a) Basal divergences recovered by all analyses are illustrated in these two topologies. Left: maximum likelihood with a profile mixture model (IQtree-cat) on matrix 1. Right: maximum likelihood with gene-wise partitioning (IQtree-part) of matrix 1. Small squares mark branches with full support. Arrows indicate conflicting nodes between topologies. New transcriptomes in bold. b) Grid of matrices and inference methods, colored according to support value (local posterior probability for Astral, bootstrap for IQ-TREE, posterior probability for PhyloBayes). Grids 1–4 correspond to the four nodes indicated in (a). Gray cells represent splits that are absent in a given analysis. M1–M6: matrices 1–6; Dayhoff: PhyloBayes on Dayhoff-recoded matrices. Silhouettes by Tauana Cunha, available at phylopic.org.

resolve in the vetigastropod tree, displaying some of the lowest scores (Supplementary Fig. S9b available on Dryad).

Along with all the potential sources of systematic error above, we further tested whether introgression could be behind conflicting results as a biological source

of discordance. We looked for asymmetric patterns of gene tree discordance throughout the vetigastropod tree, but found no evidence of introgression at any of the internal branches, including those subtending the two recalcitrant nodes. The highest observed Z-score among all branches was 0.057 ($P=0.48$), while only Z-scores of

at least 1.65 would be indicative of possible introgression at a threshold P -value of 0.05.

Vetigastropod Relationships

In all analyses, Pleurotomarioidea was the sister group to all other vetigastropods, and the following divergence separated Seguenzioidea (represented by Chilodontidae) from the remaining groups (Fig. 3a). Scissurelloidea (represented by Scissurellidae), Lepetodriloidea (represented by Lepetodrilidae), and Lepetelloidea (represented by Pseudococculinidae and Pyropeltidae) formed a clade, from here on referred to as Lepetellida *sensu stricto* (Fig. 3a). The position of Haliotoidea and Fissurelloidea varied with inference method: with BI and ML using profile mixture models of amino acid frequencies, Fissurelloidea was sister group to Lepetellida s.s., this clade being sister group to Trochoidea, and all shared a common ancestor with Haliotoidea (Fig. 3a, left). With a summary coalescent method and ML with gene-wise partitioning, Haliotoidea was instead the sister group to Trochoidea, which together were sister group to Fissurelloidea, and this entire clade was sister group to Lepetellida s.s. (Fig. 3a, right). Both of these topologies thus contradict the clade Lepetellida as currently accepted, in which Haliotoidea and Fissurelloidea are also included in the order (Bouchet et al. 2017).

Within Trochoidea, Skeneidae was sister group to all other trochoid families, which split into a clade with Calliostomatidae and Trochidae, and another clade encompassing the rest of the familial diversity (Fig. 3a). Among the latter, coalescent trees had some unresolved nodes, but other analyses recovered Colloniidae and Phasianellidae as sister group to Areneidae with high support, and this clade as the sister group of Angariidae and Liottiidae. Tegulidae is the only family that was not monophyletic, with the clade containing *Tegula* being more closely related to Turbinidae than to other tegulids (Fig. 3a).

In Trochidae, all subfamilies with multiple sampled genera were monophyletic. Trochinae was the sister group to all other trochids, with the following divergences separating, in this order, Monodontinae, Umboniinae, Chrysostomatinae, Stomatellinae, and Cantharidinae (Fig. 3a). In Turbinidae, *Astralium* was recovered as the sister group to other sampled genera, with *Lithopoma* then diverging from *Lunella* and *Turbo*. In Pleurotomariidae, *Perotrochus* and *Bayerotrochus* were more closely related than either is to *Entemnotrochus* (Fig. 3a). Within Fissurellidae, Zeidorinae (*Puncturella* and *Hemitoma*) was sister group to all other fissurellids. Diodorinae and Fissurellinae were sister groups, and Emarginulinae was not monophyletic, with *Emarginula* being more closely related to Diodorinae and Fissurellinae than to *Scutus* and *Montfortula* (Fig. 3a). Emarginulinae was recovered as monophyletic in summary coalescent analyses, but with low support.

DISCUSSION

We present the first comprehensive phylogenomic framework for Vetigastropoda, including all eight superfamilies and 18 of the 38 currently accepted families (Fig. 3a). With the exception of Tegulidae, all families and superfamilies with multiple sampled taxa are monophyletic. Pleurotomarioidea is the sister group to all other vetigastropods, a result that has been consistent since the first molecular phylogenies (Harasewych et al. 1997). Apart from this, previous work on this major gastropod lineage has resulted in many alternative and poorly resolved topologies for deep nodes. Our resulting backbone of the vetigastropod tree is largely concordant among matrices and inference methods, resolving most of the basal divergences in Vetigastropoda. Despite the substantial number of genes and multiple strategies to minimize error, discordance between methods still resulted in conflicting topologies for two basal splits in the tree.

Methodological Discordance

Phylogenetic results were not affected by strategy of gene subsampling or taxa with higher missing data, instead being discordant based on the model of inference that was used. Bayesian and ML analyses that used the more complex CAT/C10-C60 models for site heterogeneity favored one topology for the position of Haliotidae and Fissurellidae, while gene-partitioned ML and a coalescent approach based on gene trees recovered an alternative tree. Each of these methods has important limitations: concatenated analyses do not allow for discordance in gene history, which is a well-known violation of biological sources of discordance such as ILS (Degnan and Rosenberg 2009); for partitioned analyses, it has been demonstrated that small-sample bias from many individual genes can lead to a large accumulated error driving strong LBA artifacts (Wang et al. 2019); and summary coalescent methods are prone to fault when gene tree estimation error is present (Gatesy and Springer 2014; Meiklejohn et al. 2016; Molloy and Warnow 2018), which is likely the case if individual alignments are not informative enough.

Because of the topological agreement between methods that rely on information from individual genes, we hypothesized that systematic error from over-partitioning and/or uninformative genes was behind conflicting results in our phylogenomic analyses. We therefore tackled the three methodological limitations simultaneously by using a full MSC approach, with a linked-sites model, using exclusively the genes most informative for the two target nodes. Contrary to our prediction, the position of Haliotidae reflected the same relationship recovered by partitioned and summary methods, while the position of Fissurellidae supported the reconstruction from concatenation methods using models of site heterogeneity. This result and the low support metrics for these two divergences in different analyses highlight that even the most informative genes

of this transcriptomic dataset are not enough to resolve the two recalcitrant nodes. They also indicate that gene-wise partitioning and uninformative genes are not the cause of conflict. Combined, our results from matrix subsampling, evaluation of missing data, phylogenetic signal, distribution of support metrics, and properties of genes supporting each alternative topology, reveal that none of these systematic sources of error are the cause of discordant results.

Biological sources of conflict can then be considered. Hybridization is one such factor that has been increasingly detected, including among ancestral lineages, with most studies focusing on plants or vertebrates (Lin et al. 2019; Macguigan and Near 2019; Vanderpool et al. 2020; Cai et al. 2021). Here, we found no evidence of introgression in any internal branch of the vetigastropod tree. While these events can be difficult to detect in deeper parts of a phylogeny, our tests did not reveal the slightest signs of asymmetrical patterns of gene tree discordance that would be consistent with ancient gene flow, and therefore the conflict between some of our inference methods cannot be explained by introgression. Another possibility is that the conflict derives from ILS, with MSC approaches more accurately inferring the true history of vetigastropods. Short branches lead to the two uncertain splits in the tree, which indicates that these likely occurred in times of rapid divergence between vetigastropod clades. This is consistent with a scenario of ILS. However, none of the current implementations of the MSC offers site-heterogeneous modeling of amino acid frequencies, which is one of the most important sources of heterogeneity in genomic datasets (the site-homogeneous WAG being the best available model in StarBEAST2 at the moment). Compared to analyses based on site-homogeneous models, site-heterogeneous models have been shown to more accurately infer the true species tree even in the presence of ILS (Wang et al. 2019). These deep vetigastropod divergences are between 200 and 400 Myr old (Zapata et al. 2014), if not older (Fryda et al. 2008), therefore we could alternatively hypothesize high heterogeneity in the evolution of sequences to be the cause of topological discordance. In this case, the more accurate inference could be that from the concatenated methods using profile mixture models.

For another gastropod phylogenomic dataset with even older divergences, we found congruence between methods (Cunha and Giribet 2019), giving high confidence in the results. However, when methods disagree, the question remains about the cause of the discordance. Studies on different groups have tried to tease apart these sources of conflict. The importance of accounting for estimation error and ILS has been shown repeatedly with empirical datasets of many organisms, such as vertebrates, angiosperms, and fungi (Burbrink et al. 2020; Cai et al. 2021; Shen et al. 2021). A major role for ancient gene flow has been identified in divergences of at least 20 Ma in fishes (Macguigan and Near 2019), 10 Ma in primates (Vanderpool et al. 2020), and likely older in angiosperms of the order Malpighiales (Cai et al. 2021), but in this study no introgression was detected

for gastropods. Future work in other invertebrate clades could help clarify whether life history traits of these organisms might be related to the absence of lasting traces of gene flow in the phylogeny. Importantly, being able to detect the multiple issues responsible for lack of resolution at early branches has not necessarily allowed them to be resolved, and many key divergences across taxa remain unsettled (e.g. Cai et al. 2021). Taken together, recent work on phylogenomic conflict shows that better models are needed to accommodate the complexity of the various processes shaping the evolution of organisms. Here, we were able to rule out introgression and an array of possible sources of systematic error for the conflict at two early splits in the vetigastropod phylogeny. However, we find that there is not enough evidence to confidently discern between the two remaining alternative topologies, and therefore the exact placement of *Haliotidae* and *Fissurellidae* remains uncertain. These represent ancient and fast divergences that are hard to resolve. We argue that the development of methods that simultaneously integrate the MSC and the best existing substitution models for site-heterogeneity should bring novel insights to many phylogenetic questions at deep timescales that remain problematic even with genomic data, including these and many other recalcitrant nodes in the Tree of Life.

Specifically for vetigastropod relationships, future phylogenomic work should also increase the sampling of abalones (*Haliotidae*). The family has over 50 extant species, all in the genus *Haliotis*, which is why we originally included a single species in our analyses. Now that we have detected *Haliotis* as a key lineage in the conflicting topologies, better representation of the group should increase the signal-to-noise ratio in resolving its position among other vetigastropods.

Vetigastropod Relationships

Despite two nodes remaining unclear, the other basal divergences in the phylogeny of Vetigastropoda were well resolved and fully supported across our analyses. We recovered a clade composed of Scissurelloidea, Lepetodriloidea, and Lepetelloidea. Because of the minute size of scissurelloids and the deep-sea environments inhabited by the two latter superfamilies, these taxa have been some of the hardest to sample for molecular phylogenetic studies. Scissurelloidea and Lepetodriloidea have been recovered as sister groups in many molecular studies in which Lepetelloidea was absent (Geiger and Thacker 2005; Yoon and Kim 2005; Williams and Ozawa 2006; Kano 2008; Williams et al. 2008; Aktipis and Giribet 2010), which is concordant with our results. Lepetelloidea, on the other hand, has been absent from most studies and was recovered as the sister group to the Patellogastropoda, nested within Vetigastropoda, in a seven-gene phylogeny (Aktipis and Giribet 2012). That was an unexpected position for both Lepetelloidea and Patellogastropoda [another gastropod lineage that is actually sister group

to vetigastropods (Cunha and Giribet 2019)], and their exceptionally long branches in the seven-gene tree indicate LBA as the possible cause of such results. In mitogenome analyses, Scissurelloidea has not yet been sampled, and Lepetodriloidea is sometimes placed with Haliotoidea and Seguenzioidea (Lee et al. 2016; Uribe et al. 2016). These taxa display very long branches in mitochondrial trees, again indicating possible LBA. In our study, we consistently recovered Lepetelloidea and Lepetodriloidea as sister groups, which are in turn the sister clade of Scissurelloidea.

In the current classification of vetigastropods, Lepetelloidea, Lepetodriloidea, Scissurelloidea, Haliotoidea, and Fissurelloidea are part of the order Lepetellida (Bouchet et al. 2017). In our analyses, however, Haliotidae is not closely related to these other superfamilies and, regardless of the topology (Fig. 3a), is an independent lineage not pertaining to any of the four recognized orders. Its position warrants that Haliotoidea be elevated to its own order, here designated as Haliotida Rafinesque, 1815 **status nov.** (Table 1). Fissurellidae was recovered as either sister group to Lepetellida s.s. (Fig. 3a, left), or more distantly related to it (Fig. 3a, right). To keep the order-level classification of vetigastropods organized, we also elevate Fissurelloidea to its own order, Fissurellida J. Fleming, 1822 **status nov.** (Table 1), which is consistent with both reconstructions of the vetigastropod phylogeny (Fig. 3a). Lepetellida, here treated as Lepetellida s.s., is then redefined to comprise Lepetelloidea, Lepetodriloidea, and Scissurelloidea.

Seguenzioidea (represented by Chilodontaidae) is the sister group to all vetigastropods excepting pleurotomariids. This differs from most studies based on mitogenomes or a few markers (Kano 2008; Williams et al. 2008; Aktipis and Giribet 2012; Lee et al. 2016; Uribe et al. 2016; Uribe et al. 2017), but interestingly it is the same position as that of Seguenziidae in early morphological studies (Ponder and Lindberg 1997; Sasaki 1998). The group has had various placements even more broadly in the gastropod phylogeny, sometimes being considered as more closely related to Caenogastropoda, due to similarly complex reproductive anatomy [reviewed in Kano (2008)]. It has been hypothesized that such traits, including sperm storage, evolved independently several times, possibly as a more efficient investment of resources in deep sea environments, where locating partners can be more challenging (Quinn Jr 1983).

Both of the two alternative hypotheses for the position of Fissurelloidea (Fig. 3a) disagree with previous studies, in which the divergence between Fissurellidae and other vetigastropods has been one of the first splits in the tree (Kano 2008; Aktipis and Giribet 2010; Lee et al. 2016; Uribe et al. 2016; Guo et al. 2020). The family has a different gene arrangement in the mitochondria and faster rates of mitochondrial evolution compared to other vetigastropods (Lee et al. 2016; Uribe et al. 2016; Uribe et al. 2017), leading to Fissurellidae always having the longest branch in mitogenomic studies, and indicating an effect of LBA pushing fissurellids to diverge early (Guo et al. 2020). Within the family, relationships in

our analyses were well resolved and congruent with our past inference based on denser taxon sampling and fewer molecular markers (Cunha et al. 2019). Zeidoriniae is sister group to the other sampled fissurellids, and Emarginulinae is not monophyletic, with *Emarginula* being more closely related to Diodorinae and Fissurellinae than to other emarginulines.

Trochoidea is the most diverse superfamily of Vetigastropoda, with over 2300 described species (WoRMS 2021), and our sampling includes ten out of 13 families. Phasianellidae and Angariidae had been elevated to the superfamilies Phasianelloidea and Angarioidea (Williams et al. 2008), based on their recovered position diverging earlier in the vetigastropod tree (Williams et al. 2008; Aktipis and Giribet 2012). Our results do not support this hypothesis, and instead confirm the results from mitogenomic analyses that reinstated these families as members of Trochoidea (Uribe et al. 2017).

Relationships Within Trochoidea

Where sampling overlaps, our results agree with mitogenomic studies for trochoids (Lee et al. 2016; Uribe et al. 2016; Uribe et al. 2017), Calliostomatidae and Trochidae being sister groups, and together as the sister group to a clade of Angariidae, Phasianellidae, Tegulidae (not monophyletic), and Turbinidae. With our extended sampling of families, we can further compare our results to more densely sampled studies of Trochoidea that used a handful of nuclear and mitochondrial genes (Williams et al. 2010; Williams 2012). Skeneidae is the sister group to all other trochoids, which is the same placement found by Williams (2012) with a different representation of genera. While past work has found Liotiidae closely related to Calliostomatidae or Tegulidae (Kano 2008; Williams et al. 2008; Aktipis and Giribet 2012; Williams 2012), we recovered Liotiidae as sister group to Angariidae.

Relationships of subfamilies within Trochidae were well resolved and are fully concordant with the latest and more densely sampled phylogeny based on a handful of markers (Williams 2012), providing additional support for the backbone tree of this diverse family. While Williams et al. (2010) recovered a clade with *Stomatella* and *Stomatolina* with low support, we instead consistently found that *Stomatella* is more closely related to *Stomatia* than to *Stomatolina*.

Tegulidae is paraphyletic and divided in two clades, one of which is more closely related to Turbinidae, confirming results from Williams (2012) and Uribe et al. (2017). The clade including *Tectus*, *Rochia*, and *Cittarium* was treated as unassigned in those studies, with the authors suggesting that a new family designation would be appropriate once results were corroborated with more data. A more conservative alternative, which we propose here, is to transfer both clades to Turbinidae. Before being given familial rank by Williams (2012), Teguliniae was already treated as a subfamily of Turbinidae (Bouchet and Rocroi 2005; Williams et al. 2008), which was also supported recently by morphological cladistic analyses

(Dornellas et al. 2020). Our redefined Turbinidae differs in composition from that of Dornellas et al. (2020) in that we include *Cittarium*, but not *Phasianella*, according to our phylogenomic results (Fig. 3). Of the nine genera presently classified as Tegulidae, seven have been sampled here or in the molecular analyses of Williams et al. (2008), Williams (2012), Uribe et al. (2017), and Guo et al. (2020), with *Omphalius*, *Norrisia*, and *Chlorostoma* being recovered closely to *Tegula*, which places them in Teguliniae. *Tectus*, *Cittarium*, and *Rochia* are consistently recovered as a clade, which we name Tectinae **subfam. nov.**, with *Tectus* Montfort, 1810 as the type genus. Two genera (*Callistele* and *Carolesia*) have not yet been sampled in molecular phylogenies, but morphological analyses place *Carolesia* in Teguliniae (Dornellas et al. 2020). In summary, in our proposed classification Turbinidae consists of subfamilies Turbininae and Prisogasterinae as currently accepted, plus Teguliniae **status nov.** and Tectinae **subfam. nov.**, with *Callistele* provisionally unassigned to any subfamily. The position of Prisogasterinae among other subfamilies remains to be tested with phylogenomic data.

Turbinidae Rafinesque, 1815

Prisogasterinae Hickman & McLean, 1990
Prisogaster Mörch, 1850

Tectinae Cunha & Giribet **subfam. nov.**

<http://zoobank.org/urn:lsid:zoobank.org:act:3C3714FF-2219-4E9F-AEC0-76E9BA958E59>

The least inclusive monophyletic group containing the following genera:
Cittarium Philippi, 1847
Rochia Gray, 1857
Tectus Montfort, 1810

Teguliniae Kuroda, Habe & Oyama, 1971

Carolesia Güller & Zelaya, 2014
Chlorostoma Swainson, 1840
Norrisia Bayle, 1880
Omphalius Philippi, 1847
Tegula Lesson, 1832

Turbininae Rafinesque, 1815

Astraea Röding, 1798
Astralium Link, 1807
Bellastraea Iredale, 1924
Bolma Risso, 1826
Cookia Lesson, 1832
Guildfordia Gray, 1850
Lithopoma Gray, 1850
Lunella Röding, 1798
Megastraea McLean, 1970
Modelia Gray, 1850
Olearia Herrmannsen, 1847
Pomaulax Gray, 1850
Turbo Linnaeus, 1758
Uvanilla Gray, 1850

Incertae sedis

Callistele Cotton & Godfrey, 1935
Tropidomarga Powell, 1951

CONCLUSIONS

With an extensive sample of all vetigastropod superfamilies and about half of the modern familial diversity, we provide the first phylogenomic framework for deep relationships in Vetigastropoda. Our sampling also provides a robust backbone for the most diverse families, Trochidae, Fissurellidae, and Turbinidae (here redefined to include taxa previously assigned to Tegulidae). We explored strategies to minimize systematic error by subsampling genes and comparing inference methods. Divergences at all levels are generally well supported and largely concordant between analyses. Still, two basal nodes involving the position of Haliotidae and Fissurellidae show conflicting topologies across analytical approaches (profile mixture models on concatenated datasets versus gene-wise partitioning and summary coalescent methods). We evaluated the phylogenetic signal of individual genes and found that, even though most genes are unable to resolve the two splits, they are not responsible for the methodological discordance. In addition, no signs of introgression were detected, ruling out ancient gene flow as a source of conflict. Methods not yet available that simultaneously consider the MSC and amino acid profile mixture models may be needed to confidently place these two commercially and evolutionarily important vetigastropod families, as well as to resolve many other ancient and recalcitrant nodes across the Tree of Life.

Our results for the backbone of vetigastropod relationships differ considerably from previous work on group. We are able to resolve many basal relationships that have been greatly affected by LBA artifacts in past molecular studies. Such biases have been widespread, from datasets of a handful of markers to complete mitochondrial genomes. Difficulties in resolving deep relationships in the group are likely due to the ancient and fast divergences between main vetigastropod lineages, and the lack of power in previous datasets from insufficient sequence data. We show that a large gene and taxon sampling and a careful exploration of methods are necessary, allowing many such phylogenetic questions at ancient timescales to be resolved.

SUPPLEMENTARY MATERIAL

Data available from the Dryad Digital Repository: <http://dx.doi.org/10.5061/dryad.rxwdbrv64>. R notebooks are also available from GitHub: https://github.com/tauanajc/Cunha_Reimer_Giribet_2021_SystBio. Raw data for new transcriptomes are deposited in the NCBI Sequence Read Archive (BioProject PRJNA754417).

ACKNOWLEDGMENTS

We are grateful to Vanessa Knutson, Shawn Miller, and the MISE laboratory (University of the Ryukyus) for help in the field in Okinawa. We thank three anonymous

reviewers and the associate editor for suggestions that helped refine the paper. Computations were run on the FASRC Odyssey cluster supported by the FAS Division of Science Research Computing Group at Harvard University.

FUNDING

This work was supported by a Putnam Expedition Grant from the Museum of Comparative Zoology, a Graduate Student Research Award from the Society of Systematic Biologists, and a Faculty for the Future Fellowship from the Schlumberger Foundation to T.J.C., by a Doctoral Dissertation Improvement Grant from NSF (Award #1701648 to T.J.C. and G.G.), and by internal funds from the Faculty of Arts and Sciences at Harvard University to G.G.

AUTHORS CONTRIBUTIONS

T.J.C. and G.G. conceived the study, collected, and identified specimens. T.J.C. carried out lab work, analyzed the data, produced figures and drafted the manuscript. J.D.R. facilitated collection of specimens. All authors contributed to the manuscript and gave final approval for publication.

REFERENCES

Ab Lah R., Smith J., Savins D., Dowell A., Bucher D., Benkendorff K. 2017. Investigation of nutritional properties of three species of marine turban snails for human consumption. *Food Sci. Nutr.* 5:14–30.

Aktipis S.W., Giribet G. 2010. A phylogeny of Vetigastropoda and other “archaeogastropods”: re-organizing old gastropod clades. *Invertebr. Biol.* 129:220–240.

Aktipis S.W., Giribet G. 2012. Testing relationships among the vetigastropod taxa: a molecular approach. *J. Molluscan Stud.* 78:12–27.

Altenhoff A.M., Glover N.M., Train C.-M., Kaleb K., Warwick Vesztrocy A., Dylus D., de Farias T.M., Zile K., Stevenson C., Long J., Redestig H., Gonnet G.H., Dessimoz C. 2018. The OMA orthology database in 2018: retrieving evolutionary relationships among all domains of life through richer web and programmatic interfaces. *Nucleic Acids Res.* 46:D477–D485.

Bouchet P., Rocroi J.-P. 2005. Classification and nomenclator of gastropod families. *Malacologia*. 47:1–397.

Bouchet P., Rocroi J.-P., Hausdorf B., Kaim A., Kano Y., Nützel A., Parkhaev P., Schrödl M., Strong E.E. 2017. Revised classification, nomenclator and typification of gastropod and monoplacophoran families. *Malacologia*. 61:1–526.

Bouckaert R., Vaughan T.G., Barido-Sottani J., Duchêne S., Fourment M., Gavryushkina A., Heled J., Jones G., Kühnert D., De Maio N., Matschiner M., Mendes F.K., Müller N.F., Ogilvie H.A., du Plessis L., Popinga A., Rambaut A., Rasmussen D., Siveroni I., Suchard M.A., Wu C.-H., Xie D., Zhang C., Stadler T., Drummond A.J. 2019. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLOS Comput. Biol.* 15:e1006650.

Burbrink F.T., Grazziotin F.G., Pyron R.A., Cundall D., Donnellan S., Irish F., Keogh J.S., Kraus F., Murphy R.W., Noonan B., Raxworthy C.J., Ruane S., Lemmon A.R., Lemmon E.M., Zaher H. 2020. Interrogating genomic-scale data for Squamata (lizards, snakes, and amphisbaenians) shows no support for key traditional morphological relationships. *Syst. Biol.* 69:502–520.

Cai L., Xi Z., Lemmon E.M., Lemmon A.R., Mast A., Buddenhagen C.E., Liu L., Davis C.C. 2021. The perfect storm: gene tree estimation error, incomplete lineage sorting, and ancient gene flow explain the most recalcitrant ancient angiosperm clade, Malpighiales. *Syst. Biol.* 70:491–507.

Capella-Gutiérrez S., Silla-Martínez J.M., Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 25:1972–1973.

Chernomor O., von Haeseler A., Minh B.Q. 2016. Terrace aware data structure for phylogenomic inference from supermatrices. *Syst. Biol.* 65:997–1008.

Cunha T.J., Giribet G. 2019. A congruent topology for deep gastropod relationships. *Proc. R. Soc. B Biol. Sci.* 286:20182776.

Cunha T.J., Lemer S., Bouchet P., Kano Y., Giribet G. 2019. Putting keyhole limpets on the map: phylogeny and biogeography of the globally distributed marine family Fissurellidae (Vetigastropoda, Mollusca). *Mol. Phylogenet. Evol.* 135:249–269.

Dayhoff M.O., Schwartz R.M., Orcutt B.C. 1978. A model of evolutionary change in proteins. In: Dayhoff M.O., editors. *Atlas of protein sequence and structure*. Washington DC, National Biomedical Research Foundation. p. 345–352.

Degnan J.H., Rosenberg N.A. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol. Evol.* 24:332–340.

Dornellas A.P., Couto D.R., Simone L.R.L. 2020. Morphological phylogeny of the Teguliniae (Mollusca: Vetigastropoda) reinforces a Turbinidae position. *Cladistics*. 36:129–163.

Drummond A.J., Suchard M.A., Xie D., Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29:1969–1973.

Foster P.G. 2004. Modeling compositional heterogeneity. *Syst. Biol.* 53:485–495.

Fryda J., Nützel A., Wagner P.J. 2008. Paleozoic gastropoda. In: Ponder W.F., Lindberg D.R., editors. *Phylogeny and evolution of the mollusca*. Berkeley, CA: University of California Press. p. 239–270.

Fu L., Niu B., Zhu Z., Wu S., Li W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*. 28:3150–3152.

Gatesy J., Sloan D.B., Warren J.M., Baker R.H., Simmons M.P., Springer M.S. 2019. Partitioned coalescence support reveals biases in species-tree methods and detects gene trees that determine phylogenomic conflicts. *Mol. Phylogenet. Evol.* 139:106539.

Gatesy J., Springer M.S. 2014. Phylogenetic analysis at deep timescales: unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. *Mol. Phylogenet. Evol.* 80:231–266.

Geiger D.L., Thacker C.E. 2005. Molecular phylogeny of Vetigastropoda reveals non-monophyletic Scissurellidae, Trochoidea, and Fissurelloidea. *Molluscan Res.* 25:47–55.

Grabherr M.G., Haas B.J., Yassour M., Levin J.Z., Thompson D.A., Amit I., Adiconis X., Fan L., Raychowdhury R., Zeng Q., Chen Z., Mauceli E., Hacohen N., Gnrke A., Rhind N., di Palma F., Birren B.W., Nusbaum C., Lindblad-Toh K., Friedman N., Regev A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29:644–652.

Guo E., Yang Y., Kong L., Yu H., Liu S., Liu Z., Li Q. 2020. Mitogenomic phylogeny of Trochoidea (Gastropoda: Vetigastropoda): new insights from increased complete genomes. *Zool. Scr.* 50:43–57.

Haas B.J., Papanicolaou A., Yassour M., Grabherr M., Blood P.D., Bowden J., Couger M.B., Eccles D., Li B., Lieber M., MacManes M.D., Ott M., Orvis J., Pochet N., Strozzi F., Weeks N., Westerman R., William T., Dewey C.N., Henschel R., LeDuc R.D., Friedman N., Regev A. 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8:1494–1512.

Harasewych M.G., Adamkewicz S.L., Blake J.A., Saudek D., Spriggs T., Bult C.J. 1997. Phylogeny and relationship of pleurotomariid gastropods (Mollusca: Gastropoda): an assessment based on partial 18S rDNA and cytochrome c oxidase I sequences. *Mol. Mar. Biol. Biotechnol.* 6:1–20.

Harris J.R., Markl J. 1999. Keyhole limpet hemocyanin (KLH): a biomedical review. *Micron* 30:597–623.

Haszprunar G. 1988. On the origin and evolution of major gastropods group, with special reference to the Streptoneura. *J. Molluscan Stud.* 54:367–441.

Huson D.H., Klöpper T., Lockhart P.J., Steel M.A. 2005. Reconstruction of reticulate networks from gene trees. In: Miyano S., Mesirov J., Kasif S., Istrail S., Pevzner P.A., Waterman M., editors. *Research in computational molecular biology. RECOMB 2005*. Berlin, Heidelberg: Springer. p. 233–249.

Kalyaanamoorthy S., Minh B.Q., Wong T.K.F., von Haeseler A., Jermiin L.S. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14:587–589.

Kano Y. 2008. Vetigastropod phylogeny and a new concept of Seguenzioidea: independent evolution of copulatory organs in the deep-sea habitats. *Zool. Scr.* 37:1–21.

Katoh K., Standley D.M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30:772–780.

Krueger F., James F., Ewels P., Afyounian E., Schuster-Boeckler B. 2018. TrimGalore - DOI via Zenodo: 10.5281/zenodo.5127899. Available from: <https://github.com/FelixKrueger/TrimGalore>.

Langmead B., Salzberg S.L. 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9:357–359.

Lartillot N., Philippe H. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol. Biol. Evol.* 21:1095–1109.

Lartillot N., Rodrigue N., Stubbs D., Richer J. 2013. PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Syst. Biol.* 62:611–615.

Le S.Q., Gascuel O., Lartillot N. 2008. Empirical profile mixture models for phylogenetic reconstruction. *Bioinformatics*. 24:2317–2323.

Lee H., Samadi S., Puillandre N., Tsai M.H., Dai C.F., Chen W.J. 2016. Eight new mitogenomes for exploring the phylogeny and classification of Vetigastropoda. *J. Molluscan Stud.* 82:534–541.

Leiva G.E., Castilla J.C. 2002. A review of the world marine gastropod fishery: evolution of catches, management and the Chilean experience. *Rev. Fish Biol. Fish.* 11:283–300.

Lin H.-Y., Hao Y.-J., Li J.-H., Fu C.-X., Soltis P.S., Soltis D.E., Zhao Y.-P. 2019. Phylogenomic conflict resulting from ancient introgression following species diversification in *Stewartia* s.l. (Theaceae). *Mol. Phylogenet. Evol.* 135:1–11.

Macguigan D.J., Near T.J. 2019. Phylogenomic signatures of ancient introgression in a rogue lineage of darters (Teleostei: Percidae). *Syst. Biol.* 68:329–346.

Meiklejohn K.A., Faircloth B.C., Glenn T.C., Kimball R.T., Braun E.L. 2016. Analysis of a rapid evolutionary radiation using ultraconserved elements: evidence for a bias in some multispecies coalescent methods. *Syst. Biol.* 65:612–627.

Minh B.Q., Hahn M.W., Lanfear R. 2020a. New methods to calculate concordance factors for phylogenomic datasets. *Mol. Biol. Evol.* 37:2727–2733.

Minh B.Q., Schmidt H.A., Chernomor O., Schrempf D., Woodhams M.D., Von Haeseler A., Lanfear R. 2020b. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37:1530–1534.

Mirarab S., Warnow T. 2015. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics*. 31:i44–i52.

Molloy E.K., Warnow T. 2018. To include or not to include: the impact of gene filtering on species tree estimation methods. *Syst. Biol.* 67:285–303.

Mora Román J.J., Del Campo M., Villar J., Paolini F., Curzio G., Venuti A., Jara L., Ferreira J., Murgas P., Lladser A., Manubens A., Becker M.I. 2019. Immunotherapeutic potential of mollusk hemocyanins in combination with human vaccine adjuvants in murine models of oral cancer. *J. Immunol. Res.* 2019:1–19.

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. *Mol. Biol. Evol.* 32:268–274.

Ogilvie H.A., Bouckaert R.R., Drummond A.J. 2017. StarBEAST2 brings faster species tree inference and accurate estimates of substitution rates. *Mol. Biol. Evol.* 34:2101–2114.

Okutani T. 2000. *Marine Mollusks in Japan*. Japan: Tokai University Press.

Paradis E., Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*. 35:526–528.

Ponder W.F., Lindberg D.R. 1997. Towards a phylogeny of gastropod molluscs: an analysis using morphological characters. *Zool. J. Linn. Soc.* 119:83–265.

Quinn Jr J.F. 1983. A revision of the Seguenziacea Verrill, 1884 (Gastropoda: Prosobranchia). I. Summary and evaluation of the superfamily. *Proc. Biol. Soc. Washington*. 96:725–757.

Rambaut A., Drummond A.J., Xie D., Baele G., Suchard M.A. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* 67:901–904.

Revell L.J. 2012. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* 3:217–223.

Salvini-Plawen L.V. 1980. A reconsideration of systematics in the Mollusca (phylogeny and higher classification). *Malacologia*. 19:249–278.

Salvini-Plawen L.V., Haszprunar G. 1987. The Vetigastropoda and the systematics of streptoneurous Gastropoda (Mollusca). *J. Zool.* 211:747–770.

Sasaki T. 1998. Comparative anatomy and phylogeny of the Recent Archaeogastropoda (Mollusca: Gastropoda). *Univ. Tokyo Bull.* 38:1–223.

Shen X.-X., Steenwyk J.L., Rokas A. 2021. Dissecting incongruence between concatenation- and quartet-based approaches in phylogenomic data. *Syst. Biol.* 70:997–1014.

Simão F.A., Waterhouse R.M., Ioannidis P., Kriventseva E.V., Zdobnov E.M. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 31:3210–3212.

Smith S.A., Dunn C.W. 2008. Phyutility: a phyloinformatics tool for trees, alignments and molecular data. *Bioinformatics*. 24:715–716.

Smith-Unna R., Boursnell C., Patro R., Hibberd J.M., Kelly S. 2016. TransRate: reference-free quality assessment of *de novo* transcriptome assemblies. *Genome Res.* 26:1134–1144.

Song L., Florea L. 2015. Rcorrector: efficient and accurate error correction for Illumina RNA-seq reads. *GigaScience*. 4:48.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 30:1312–1313.

Strimmer K., von Haeseler A. 1997. Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. *Proc. Natl. Acad. Sci. USA*. 94:6815–6819.

Uribe J.E., Irisarri I., Templado J., Zardoya R. 2019. New patellogastropod mitogenomes help counteracting long-branch attraction in the deep phylogeny of gastropod mollusks. *Mol. Phylogenet. Evol.* 133:12–23.

Uribe J.E., Kano Y., Templado J., Zardoya R. 2016. Mitogenomics of Vetigastropoda: insights into the evolution of pallial symmetry. *Zool. Scr.* 45:145–159.

Uribe J.E., Williams S.T., Templado J., Abalde S., Zardoya R. 2017. Denser mitogenomic sampling improves resolution of the phylogeny of the superfamily Trochoidea (Gastropoda: Vetigastropoda). *J. Molluscan Stud.* 83:111–118.

Vanderpool D., Minh B.Q., Lanfear R., Hughes D., Murali S., Harris R.A., Raveendran M., Muzny D.M., Hibbins M.S., Williamson R.J., Gibbs R.A., Worley K.C., Rogers J., Hahn M.W. 2020. Primate phylogenomics uncovers multiple rapid radiations and ancient interspecific introgression. *PLoS Biol.* 18:e3000954.

Wang H.-C., Susko E., Roger A.J. 2019. The relative importance of modeling site pattern heterogeneity versus partition-wise heterotachy in phylogenomic inference. *Syst. Biol.* 68:1003–1019.

Wang L.-G., Lam T.T.-Y., Xu S., Dai Z., Zhou L., Feng T., Guo P., Dunn C.W., Jones B.R., Bradley T., Zhu H., Guan Y., Jiang Y., Yu G. 2020. Treeio: an R package for phylogenetic tree input and output with richly annotated and associated data. *Mol. Biol. Evol.* 37:599–603.

Williams S.T. 2012. Advances in molecular systematics of the vetigastropod superfamily Trochoidea. *Zool. Scr.* 41:571–595.

Williams S.T., Donald K.M., Spencer H.G., Nakano T. 2010. Molecular systematics of the marine gastropod families Trochidae and Calliostomatidae (Mollusca: Superfamily Trochoidea). *Mol. Phylogenet. Evol.* 54:783–809.

Williams S.T., Karube S., Ozawa T. 2008. Molecular systematics of Vetigastropoda: Trochidae, Turbinidae and Trochoidea redefined. *Zool. Scr.* 37:483–506.

Williams S.T., Ozawa T. 2006. Molecular phylogeny suggests polyphyly of both the turban shells (family Turbinidae) and the superfamily Trochoidea (Mollusca: Vetigastropoda). *Mol. Phylogenet. Evol.* 39:33–51.

WoRMS Editorial Board. 2021. World Register of marine species. Available from <http://www.marinespecies.org> doi: 10.14284/170.

Wort E.J.C., Fenberg P.B., Williams S.T. 2017. Testing the contribution of individual genes in mitochondrial genomes for assessing phylogenetic relationships in Vetigastropoda. *J. Molluscan Stud.* 83:123–128.

Yoon S.H., Kim W. 2005. Phylogenetic relationships among six vetigastropod subgroups (Mollusca, Gastropoda) based on 18S rDNA sequences. *Mol. Cells* 19:283–288.

Yu G., Smith D.K., Zhu H., Guan Y., Lam T.T.-Y. 2017. *ggtree*: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* 8:28–36.

Zapata F., Wilson N.G., Howison M., Andrade S.C.S., Jorger K.M., Schrodil M., Goetz F.E., Giribet G., Dunn C.W. 2014. Phylogenomic analyses of deep gastropod relationships reject Orthogastropoda. *Proc. R. Soc. B Biol. Sci.* 281:20141739.