A phylogenomic framework for Vetigastropoda (Mollusca)

Tauana Junqueira Cunha and Gonzalo Giribet

Harvard University

Corresponding author: tauanacunha@g.harvard.edu | orcid.org/0000-0002-8493-2117

Abstract

Abalones, turban snails, top snails, keyhole limpets and slit shells are just some of the diverse marine Vetigastropoda. With major lineages having ancient divergences in the Paleozoic Era, basal nodes in the phylogeny have been largely unresolved. Here we present the first genomic-scale dataset focused on vetigastropods, including a comprehensive sampling of taxa with all superfamilies and about half of the extant families (41 new transcriptomes, 49 total ingroup terminals). Our recovered topology is congruent among inference methods for the majority of nodes, and differs from previous studies, in which long branch attraction artifacts seem to be pervasive. The first three divergences sequentially separate Pleurotomarioidea, Seguenzioidea and Haliotoidea from remaining vetigastropods. Fissurelloidea is sister group to Scissurelloidea, Lepetodriloidea and Lepetelloidea, which all together are the sister group to Trochoidea. A methodological discordance exists in the position of Haliotidae and Fissurellidae, possibly due to a compounded effect of small sample bias derived from gene-wide partitioning. Dense sampling of trochoids and fissurellids also provides a backbone for family and subfamily-level relationships, with Tegulidae not monophyletic.

1.1   Introduction

Abalones, turban snails, top snails, keyhole limpets, slit shells and little slit shells are just some of the diverse marine Vetigastropoda. With about 3,700 living species, and many more thousands of fossil taxa, vetigastropods are one of the five major lineages of gastropods (Cunha and Giribet 2019), and the

dominant gastropod clade throughout the Paleozoic and most of the Mesozoic eras (Fryda et al. 2008). They are distributed at all depths and types of habitats, from shallow hard substrates, to sand, to deep sea vents and cold seeps. Despite many attempts, an understanding of the evolutionary relationships between the main clades of vetigastropods remains a challenge. With such an old divergence of the major lineages, many morphological traits show convergence, and the amount of molecular data collected so far has not been sufficient to resolve short internodes in the tree.

The following eight superfamilies are currently accepted containing 38 extant families (Bouchet et al. 2017; WoRMS 2019): Pleurotomarioidea (Pleurotomariidae: slit shells), Fissurelloidea (Fissurell-idae: keyhole and slit limpets), Haliotoidea (Haliotidae: abalones), Lepetelloidea (eight families: deep sea limpets), Lepetodriloidea (two families: hydrothermal vent limpets), Scissurelloidea (four families: little slit shells), Seguenzioidea (eight families), Trochoidea (13 families: turban and top snails).

Since Vetigastropoda was first recognized in 1980 (Salvini-Plawen 1980), there has been no shortage of phylogenetic studies on the group. Early on, comprehensive morphological analyses tried to establish relationships and identify synapomorphies (Haszprunar 1988; Ponder and Lindberg 1997; Salvini-Plawen and Haszprunar 1987; Sasaki 1998). On the molecular side, early studies used a handful of markers and provided incremental contributions to solving vetigastropod relationships (Geiger and Thacker 2005; Harasewych et al. 1997; Yoon and Kim 2005). Some key studies, though with few markers, presented densely sampled phylogenies, helping place not only the most diverse groups, but also minute and hard-to-collect taxa (Aktipis and Giribet 2011; Kano 2008; Williams et al. 2008). Finally, a recent burst of publications based on mitochondrial genomes have targeted deep vetigastropod relationships, albeit with limited taxon sampling (Lee et al. 2016; Uribe et al. 2016, 2017; Wort et al. 2017).

These efforts led to few stable results. Pleurotomariidae has been recovered as the sister group to all other vetigastropods since the first molecular phylogenies (Harasewych et al. 1997). This contradicted morphological hypotheses in which pleurotomariids were more closely related to other families with symmetric, paired pallial organs (e.g. haliotids, fissurellids, scissurellids) and to the asymmetric trochids (Ponder and Lindberg 1997; Salvini-Plawen 1980; Sasaki 1998). Trochoidea, the most diverse superfamily of vetigastropods, has also received considerable attention, with many dedicated studies (Uribe et al. 2017; Williams 2012; Williams et al. 2010, 2008; Williams and Ozawa 2006). Phasianellidae and Angariidae, once considered subfamilies of Turbinidae, have later been elevated to family and even superfamily status, being recovered as early divergences in the tree (Williams et al. 2008; Williams and Ozawa 2006). More recently, however, they have been consistently recovered within the superfamily Trochoidea with mitochondrial genomes and a small sample of vetigastropod transcriptomes (Lee et al. 2016; Uribe et al. 2016, 2017; Zapata et al. 2014).

Except for the position of Pleurotomariidae, and despite the numerous studies on Vetigastropoda and Trochoidea, relationships between superfamilies and families remain contentious. Most datasets have non-overlapping taxa representation and do not have enough sequence data to resolve such old

divergences. In addition, mitochondrial genomes suffer from having fast evolutionary rates, which regularly create artifacts of long branch attraction (LBA) (Uribe et al. 2016, 2019). This issue is accentuated by insufficient taxon sampling of vetigastropods and outgroups. Particularly problematic taxa that have been placed in very different positions depending on the study, and that seem to be more affected by LBA, are Fissurellidae, Haliotidae, Lepetelloidea and Lepetodriloidea (Aktipis and Giribet 2011; Lee et al. 2016; Uribe et al. 2016, 2017; Williams et al. 2008). Familial relationships within Trochoidea have been redefined recently with relatively good support, but pending the inclusion of several missing families (Uribe et al. 2017).

Here we present new transcriptomes for 41 genera of vetigastropods, and combine them with previously published sequences. Our sampling covers all superfamilies and about half of the described families to resolve deep relationships at those levels. This is the first genomic-scale dataset focused on vetigastropods, including a large representation of taxa. Our recovered topology is well supported and congruent among inference methods for the majority of nodes.

1.2   Methods

1.2.1   Sampling and sequencing

We sequenced the transcriptomes of 41 vetigastropod genera and added another eight genera with previously published sequences, for a total of 49 ingroup terminals. We further used 20 outgroup gastropods and one bivalve outgroup, for a total of 62 terminals. This sampling covers all eight superfamilies of vetigastropods, and 18 of the 38 families. All new data and selected published sequences are paired-end Illumina reads. New samples were fixed in RNA*later* (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 the Harvard University.

1.2.2   Transcriptome assembly and matrix construction

Raw reads were assembled *de novo* with Trinity v2.3.2 (Grabherr et al. 2011; Haas et al. 2013) with the same pipeline and scripts from Cunha and Giribet 2019. The completeness of the assemblies was evaluated with BUSCO v3.0.2 by comparison with the Metazoa database (Simão et al. 2015). Orthology assignment of the peptide assemblies was done with OMA v2.2.0 (Altenhoff et al. 2018). We built four matrices to account for extreme evolutionary rates, amino acid composition heterogeneity and two levels of matrix completeness. We first used a custom python script (*selectslice.py*) to select all orthogroups for which at least half of the terminals were represented (50% taxon occupancy),

resulting in a matrix with 1027 genes (matrix 1) (Figure 1.1). Each orthogroup was aligned with MAFFT v7.309 (Katoh and Standley 2013), and the ends of the alignment were trimmed to remove positions with more than 80% missing data with a custom bash script (*trimEnds.sh*). A subset of 259 genes with 70% taxon occupancy constitutes matrix 2. To avoid possible biases, saturation and long-branch attraction, matrix 3 was built by removing from matrix 1 the 20% slowest and the 20% fastest evolving genes, as calculated with TrimAl (Capella-Gutierrez et al. 2009), for a final size of 615 genes (Figure 1.1). Matrix 4 is the subset of 894 genes from matrix 1 that are homogeneous regarding amino acid composition. Homogeneity for each gene was determined with a simulation-based test from the python package p4 (Foster 2004, 2018), with a custom script modified from Laumer et al. (2018) (*p4_compo_test.py*) and a conservative p-value of 0.1. For inference methods that require concatenation, genes were concatenated using Phyutility (Smith and Dunn 2008). We further reduced compositional heterogeneity in matrices 1 and 2 by recoding amino acids into the six Dayhoff categories (Dayhoff et al. 1978) with a custom bash script (*recdayhoff.sh*). All scripts are available from Cunha and Giribet 2019.
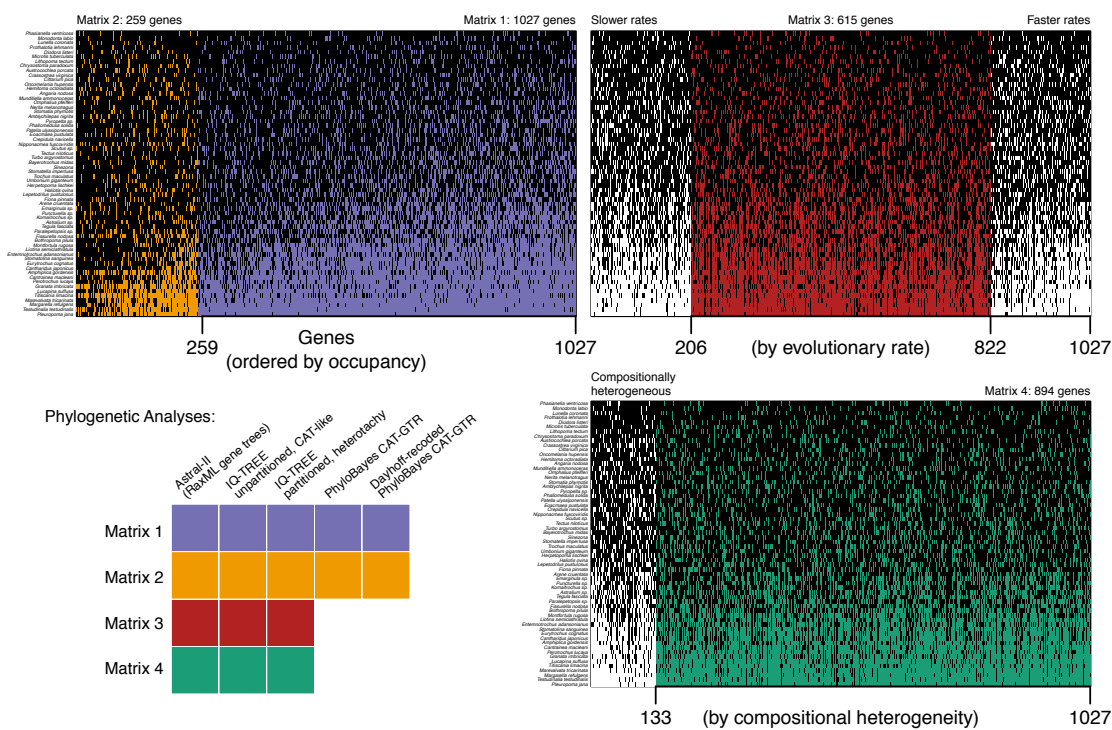

FIGURE 1.1: Matrices and methods used to infer vetigastropod relationships. Matrix 1 includes 1027 genes with 50% taxon occupancy. Matrix 2 is the subset of 259 genes with 70% taxon occupancy. Genes and species are sorted with the best sampling on the upper left. Matrix 3 is the subset of 615 genes after ordering all genes by evolutionary rate and removing the 20% slowest and 20% fastest evolving genes. Matrix 4 includes the 894 genes that are homogeneous in amino acid composition. Black cells indicate genes present for each species. Check Methods for details.

1.2.3 Phylogenetic analyses

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 (Chernomor et al. 2016; Kalyaanamoorthy et al. 2017; Nguyen et al. 2015), and with Bayesian inference (BI) in PhyloBayes MPI v1.7a (Lartillot et al. 2013). The two Dayhoff-recoded matrices were analyzed in PhyloBayes (Figure 1.1). 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 (-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) (ML variant of the Bayesian CAT model (Lartillot and Philippe 2004)) (-bb 1500 -m MFP -mset LG,WAG -rcluster 10 -mfreq F+C20 -mrate G,R). 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 the CAT-GTR model on matrices 1 and 2, discarding constant sites to speed up computation. Tree figures were edited with the R package *ggtree* (Yu et al. 2017).

1.3 Results

1.3.1 Phylogenetic analyses

Our 16 analyses included amino acid and Dayhoff-recoded matrices, four schemes of gene subsampling, and Bayesian inference, maximum likelihood and coalescent-based approaches to infer deep relationships between superfamilies and families of vetigastropods (Figure 1.1). The result is an overall well resolved and well supported phylogeny that covers all of the eight superfamilies and about half of the extant family diversity (Figure 1.2). Vetigastropoda is monophyletic, and Pleurotomarioidea is sister group to all other vetigastropods. Seguenzioidea, here represented by the family Chilodontaidae, then diverges from remaining groups, followed by the divergence of Haliotoidea (Figure 1.2). The remaining diversity is then split in two large clades, Trochoidea on one side, and Lepetelloidea, Lepetodriloidea, Scissurelloidea and Fissurelloidea as its sister group. Within the latter, Lepetelloidea (represented by Pseudococculinidae and Pyropeltidae) shares a more recent common ancestor with Lepetodriloidea (represented by Lepetodrillidae), and this clade is sister group to Scissurelloidea (represented by Scissurellidae) (Figure 1.2).

Within Trochoidea, Skeneidae is sister group to all other trochoid families. One clade then unites Calliostomatidae and Trochidae, and another encompasses the rest of the familial diversity of
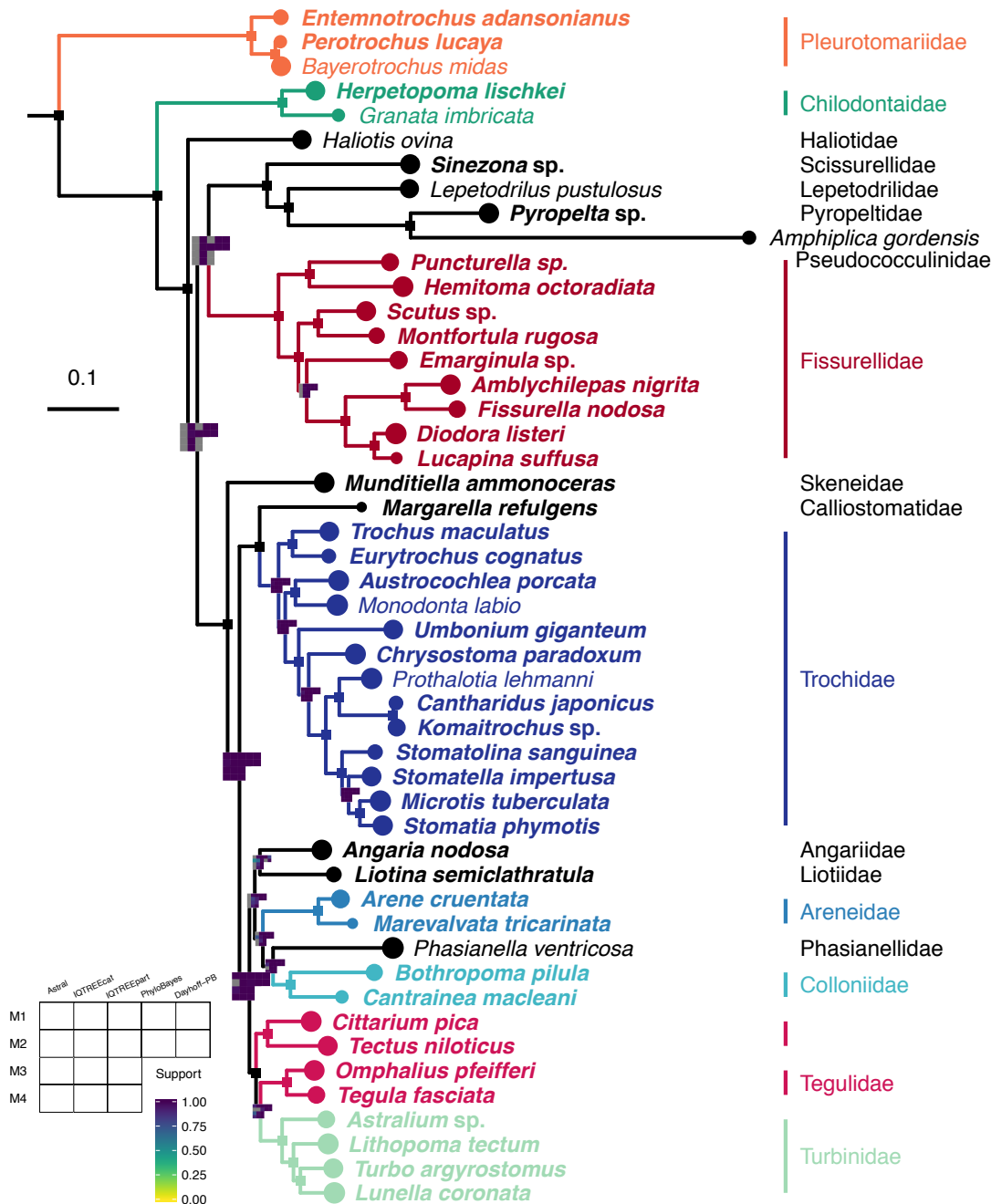
FIGURE 1.2: Vetigastropod phylogeny from matrix 1 with IQTREEcat: ML profile mixture model. Single squares mark branches with full support in all analyses; remaining nodes have plots colored according to support value (0-1) for each analysis. Grey squares represent splits that were absent in a given analysis. New transcriptomes in bold. M1-M4: matrices 1-4; IQTREEpart: ML partitioned; Dayhoff-PB: BI Dayhoff-recoded matrix. See Methods for details.

Trochoidea. Colloniidae and Phasianellidae are sister group to Areneidae, and this clade is recovered as the sister group of Angariidae and Liotiidae (Figure 1.2). Tegulidae is not monophyletic, with the clade containing *Tegula* being more closely related to Turbinidae than to other tegulids. With the exception of Tegulidae, all families with more than one sampled genus are monophyletic and well supported in all analyses. In Pleurotomariidae, *Perotrochus* and *Bayerotrochus* are more closely related than either is to *Entemnotrochus*. In Fissurellidae, Diodorinae and Fissurellinae are sister subfamilies, but Emarginulinae is not monophyletic, with *Emarginula* being more closely related to Diodorinae and Fissurellinae than to *Scutus* and *Montfortula*. This entire clade is then sister group to the subfamily Zeidorinae (*Puncturella* and *Hemitoma*) (Figure 1.2). In Trochidae, all subfamilies with more than one genus are monophyletic. Trochinae is the sister group to all other trochids, with Monodontinae then diverging from the remaining subfamilies. Stomatellinae (*Stomatella* and relatives) is the sister group to Cantharidinae (*Cantharidus* and relatives). These two subfamilies are sister to Chrysostomatinae (*Chrysostoma*), and the clade is sister group to Umboniinae (*Umbonium*) (Figure 1.2). In Turbinidae, *Astralium* is recovered as the sister group to other sampled genera, with *Lithopoma* then diverging from *Lunella* and *Turbo*.

Inference methods are largely congruent across the tree, however two nodes show conflicting results: the branch leading to the clade that is sister group to Haliotidae, and the branch leading to the clade of Scissurelloidea, Lepetodriloidea, Lepetelloidea and Fissurelloidea. In both cases, Bayesian inferences and maximum likelihood analyses with the more complex profile mixture model result in the tree shown in Figure 1.2 with full support (Figures A.1, A.2). On the other hand, coalescent-based analyses and most gene-partitioned maximum likelihood analyses result in a different topology, where Haliotoidea is the sister group to Trochoidea, this clade is sister group to Fissurellidae, and finally this large clade is sister group to the clade of Scissurelloidea, Lepetelloidea and Lepetodriloidea (Figures A.3, A.4).

1.4 Discussion

We present the first comprehensive phylogenomic framework for Vetigastropoda, including all super-families and about half of the family diversity (Figure 1.2). As expected based on previous molecular studies (Harasewych et al. 1997; Kano 2008; Williams et al. 2008), Pleurotomarioidea is the sister group to all other vetigastropods. Apart from that, previous work on this major gastropod lineage has resulted in many conflicting and poorly resolved topologies. Our resulting backbone of the vetigastropod tree is largely concordant among matrices and inference methods, but a methodological discordance still exists in two key basal nodes discussed below.

1.4.1 Methodological discordance

Short branches lead to the two basal nodes with conflicting results, which likely indicates this was a time of rapid divergence between vetigastropod clades, making it hard to infer the true history

of divergences. Results did not differ based on the strategy of gene subsampling, but based on the model of inference that was used. Bayesian analyses, which used the CAT model for site heterogeneity, and the maximum likelihood analyses that similarly used a CAT-like profile mixture model both resulted in the tree shown in Figure 1.2. A different placement of Haliotoidea and Fissurelloidea is however recovered in analyses ran with gene-partitioned maximum likelihood and with the shortcut coalescent-based approach (Figures A.3, A.4; see section 1.3).

The discordance is interesting because some recent phylogenomic studies have found agreement between coalescent-based approaches and concatenated datasets analyzed with the more complex profile mixture models of sequence heterogeneity (Cunha and Giribet 2019; Schwentner et al. 2017). Approaches based on the multispecies coalescent account for incomplete lineage sorting and therefore are theoretically more sound (Degnan and Rosenberg 2009; Edwards 2009). The agreement with some concatenated datasets indicates that more complex models were effectively accounting for the heterogeneity present in such large concatenated matrices. In practice, however, fully coalescent approaches that infer gene trees and the species tree simultaneously are computationally intensive and currently impossible to run on densely-sampled genomic datasets. Shortcut methods were then developed that use maximum likelihood gene trees to infer the species tree, accelerating computation time (Mirarab and Warnow 2015). One of the issues of shortcut methods is therefore gene tree estimation error, derived from systematic inference error or from insufficient information contained in short alignments (Meiklejohn et al. 2016; Molloy and Warnow 2018; Sayyari et al. 2017).

It has recently been shown that excessive partitioning leads to a compounded effect of small sample bias over many single protein partitions (Wang et al. 2019). This in turn results in a marked long branch attraction (LBA) issues for analyses using protein-wise partitioning (Wang et al. 2019). Because shortcut coalescent approaches such as Astral rely on single-gene analyses, they are also expected to be susceptible to errors based on gene partitioning. We argue that the congruence between our partitioned ML and coalescent results indicates a likely bias of gene-wise partitioning and LBA. Interestingly, it has been observed that using gene trees with less informative alignments caused a shift towards more pectinate (asymmetric) species trees (Meiklejohn et al. 2016). This was also observed in our results, with the alternative position of Fissurelloidea (Figures A.3, A.4), producing a more asymmetric tree than the one shown in Figure 1.2. Modeling site-heterogeneity is one of the most important strategies to avoid biases in large genomic datasets (Meiklejohn et al. 2016; Schwentner et al. 2018), and it has recently been recommended that gene-partitioning should be avoided due to strong LBA biases (Wang et al. 2019). We therefore consider the results of our site-heterogeneous models to provide a more accurate representation of the true evolutionary history of vetigastropods at the conflicting nodes (Figure 1.2).

1.4.2 Relationships among vetigastropod superfamilies

We find a well supported and stable clade composed of Scissurelloidea, Lepetodriloidea and Lepetelloidea (Figure 1.2). 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 in molecular phylogenies (but see Kano 2008). Lepetelloidea in particular is absent in most studies, but has been recovered as the sister group to Patellogastropoda, which was nested in Vetigastropoda, in a seven-gene phylogeny (Aktipis and Giribet 2011). The unexpected position of both clades and their exceptional long branches in that tree indicate a clear artifact of LBA. Here we find that Lepetelloidea is closely related to Lepetodriloidea and to other superfamilies in the order Lepetellida, in agreement with current classification (Bouchet et al. 2017; WoRMS 2019). Scissurelloidea and Lepetodriloidea have been recovered as sister groups in many molecular studies in which Lepetelloidea was absent (Aktipis and Giribet 2011; Geiger and Thacker 2005; Kano 2008; Williams et al. 2008; Williams and Ozawa 2006; Yoon and Kim 2005), which is concordant with our results. In mitochondrial genome analyses, Scissurelloidea has not been sampled yet, and Lepetodriloidea is in turn placed with Haliotoidea and Seguenzioidea (Lee et al. 2016; Uribe et al. 2016). These three superfamilies display very long branches in mitogenome studies (Lee et al. 2016; Uribe et al. 2016), again indicating a likely effect of LBA.

Seguenzioidea (here represented by Chilodontaidae) has been recovered with Haliotoidea in several mitogenome and nuclear studies (Aktipis and Giribet 2011; Lee et al. 2016; Uribe et al. 2016, 2017) and more broadly with Haliotoidea, Scissurelloidea and Lepetodriloidea in other phylogenies (Kano 2008; Williams et al. 2008). Here this superfamily is the sister group to all vetigastropods except pleurotomariids (Figure 1.2). Our results could be interpreted as an artifact of LBA, pushing Seguenzioidea to diverge earlier in the tree. However, its branch is not particularly long, and the position of this superfamily is completely stable and fully supported across all our analyses, many of which use site-heterogeneous models that are known to be more robust against LBA artifacts (Lartillot et al. 2007; Le et al. 2008). Interestingly, Seguenziidae has been placed in a similar position in early morphological studies (Ponder and Lindberg 1997; Sasaki 1998). This 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 difficult (Quinn Jr. 1983). The independent position of Seguenzioidea in our topology is in agreement with this hypothesis.

Haliotoidea is the only superfamily currently assigned to Lepetellida (Bouchet et al. 2017; WoRMS 2019) that actually is not recovered in this vetigastropod order in our results. Instead, our phylogeny puts Haliotidae as the sister group to a clade of Lepetellida *s.s.* and Trochida (Figure 1.2). Even

if one considers the alternative topology of the coalescent-based and partitioned ML analyses (Figures A.3, A.4), Haliotidae is still not included with the remaining Lepetellida, instead being sister to Trochida alone. Haliotoidea therefore appears to be an independent lineage of Vetigastropoda, not pertaining to any of the four currently recognized orders. The exact position of Haliotidae is however still not certain. Some previous studies based on morphological and a few molecular markers have found Haliotidae as closely related to Trochoidea (Aktipis and Giribet 2010; Geiger and Thacker 2005; Kano 2008; Ponder and Lindberg 1997), but our analyses indicate this could be a biased result (see subsection 1.4.1). Although Haliotidae does not present a particularly long branch in our tree, increasing taxon sampling for the family is a possible strategy to help determine its phylogenetic placement. Haliotidae has only one extant genus, which is why only one species was sampled here. Their fossil record is relatively recent, with most records being from the Cenozoic and only a few from the late Mesozoic (Groves and Alderson 2008). If the extant diversity of abalones is indeed a recent radiation, it is not clear how much further sampling could help in resolving the position of the family among vetigastropods.

Some previous molecular studies find Fissurelloidea as one of the lineages that first diverges from remaining vetigastropods (Aktipis and Giribet 2010; Kano 2008; Lee et al. 2016; Uribe et al. 2016). This is particularly worrisome in mitogenome analyses, where Fissurellidae always presents the longest branch in the tree, having a different gene arrangement and faster rates of mitochondrial evolution compared to other vetigastropods (Lee et al. 2016; Uribe et al. 2016, 2017). Such results indicate a likely effect of LBA pushing fissurellids closer to the outgroups. Our analyses, on the other hand, have an extensive sample of fissurellid taxa and support the placement of the family in Lepetellida *s.s.* as the sister group to a clade of Lepetodriloidea, Lepetelloidea and Scissurelloidea (Figure 1.2), a relationship that has been recovered before (Williams et al. 2008; Williams and Ozawa 2006). Within Fissurellidae, relationships are well resolved and congruent with the ones recently recovered in a study with denser taxon sampling, but fewer molecular markers (Cunha et al. 2019). Zeidorinae is sister group to the other sampled fissurellids, and Diodorinae and Fissurellinae form a clade. Emarginulinae is not monophyletic, with *Emarginula* being more closely related to Diodorinae and Fissurellinae than to other emarginuline.

Trochoidea is the most diverse superfamily of Vetigastropoda, with over 2,000 described species, and is here recovered as monophyletic with an extensive representation of ten out of 13 families. Early molecular results had elevated Phasianellidae and Angariidae to superfamilies Phasianelloidea and Angarioidea (Williams et al. 2008), finding that they represented earlier splits in the vetigastropod tree (Aktipis and Giribet 2011; Williams et al. 2008). Our results do not support this hypothesis, and instead agree with recent mitogenomic analyses that reinstated these families as members of Trochoidea (Uribe et al. 2017).

1.4.3 Relationships within Trochoidea

Although mitogenome studies have suffered from poor taxon sampling and long branches for many vetigastropod superfamilies, they have provided more stable and relatively better-sampled trees for Trochoidea (Lee et al. 2016; Uribe et al. 2016, 2017). Where sampling overlaps, our results are in agreement with mitogenomic analyses: Calliostomatidae and Trochidae are sister groups, and together they are sister group to a clade of Angariidae and Phasianellidae, Tegulidae (not monophyletic) and Turbinidae (Figure 1.2). We further sampled additional families, with over 75% of the family diversity being represented. With this sampling we find some similarities, but many discordances with more densely sampled studies of Trochoidea that used a handful of nuclear and mitochondrial genes (Williams 2012; Williams et al. 2010).

Skeneidae is the sister group to all other trochoids, as previously reported (Williams 2012). While past work has found Liotiidae closely related to Calliostomatide or Tegulidae (Aktipis and Giribet 2011; Kano 2008; Williams 2012; Williams et al. 2008), we recover Liotiidae as sister group to Angariidae (Figure 1.2). A similar clade had already been recovered before (Aktipis and Giribet 2011), but without support from other studies on trochoids.

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. We also find that *Stomatella* is more closely related to *Stomatia* than to *Stomatolina*, as recovered in Williams et al. 2010.

Finally, Tegulidae is divided in two clades, one of which is more closely related to Turbinidae (Figure 1.2). This result has been found in other studies (Uribe et al. 2017; Williams 2012), but the clade including *Tectus* remains treated as unassigned, and a new family designation is necessary. Of the nine genera still classified as Tegulidae, six have been sampled here and in other recent studies, with *Cittarium*, *Rochia* and *Tectus* consistently recovered in the unassigned clade (Uribe et al. 2017; Williams 2012).

1.5 Conclusion

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 allows for a robust inference of backbone relationships within the most diverse vetigastropod clades: superfamily Trochoidea, and families Trochidae and Fissurellidae. Divergences at all levels are well supported and largely concordant between inference methods.

We find that Trochoidea (sole member of the order Trochida) is the sister group to Lepetellida *s.s.* (including Lepetodriloidea, Lepetelloidea, Scissurelloidea and Fissurelloidea). Haliotoidea is sister group to this large clade, which in turn is the sister group to Seguenzioidea. The position of Haliotoidea remains the most uncertain node in the tree. Additional sampling of abalones could potentially provide

more confidence in the results recovered here, but because modern Haliotidae are a relatively recent radiation, it is unclear how much phylogenetic signal further sampling will add to these ancient divergences.

With the exception of the position of Pleurotomarioidea as the sister group to all other vetigastropods, our results differ considerably from topologies previously inferred for vetigastropods. In particular, the large taxon and gene sampling and the use of complex models of site-heterogeneity allowed us to resolve relationships that have been greatly affected by long branch attraction artifacts in previous studies. Such artifacts have been widespread in vetigastropod analyses, from datasets of a handful of molecular markers to complete mitochondrial genomes. Difficulty in resolving deep relationships in this group are likely due to the ancient and fast divergence between the main lineages, and the lack of power in previous datasets derived from insufficient sequence data. Our data illustrate that it is possible to reconstruct these ancient splits with a comprehensive sampling of taxa and sequence data.

1.6   Author's contributions

TJC and GG conceived the study, collected and identified specimens. TJC carried out lab work, analyzed the data and drafted the manuscript. Both authors improved the manuscript and gave final approval for publication.

1.7   Funding

1.8   Acknowledgments

Appendix A

FIGURE A.1: Vetigastropod phylogeny inferred with maximum likelihood and a profile mixture model on each of four matrices: (A) matrix 1, (B) matrix 2, (C) matrix 3, (D) matrix 4. Single squares mark branches with full support; remaining nodes show the bootstrap value. New transcriptomes in bold. See Methods for details.
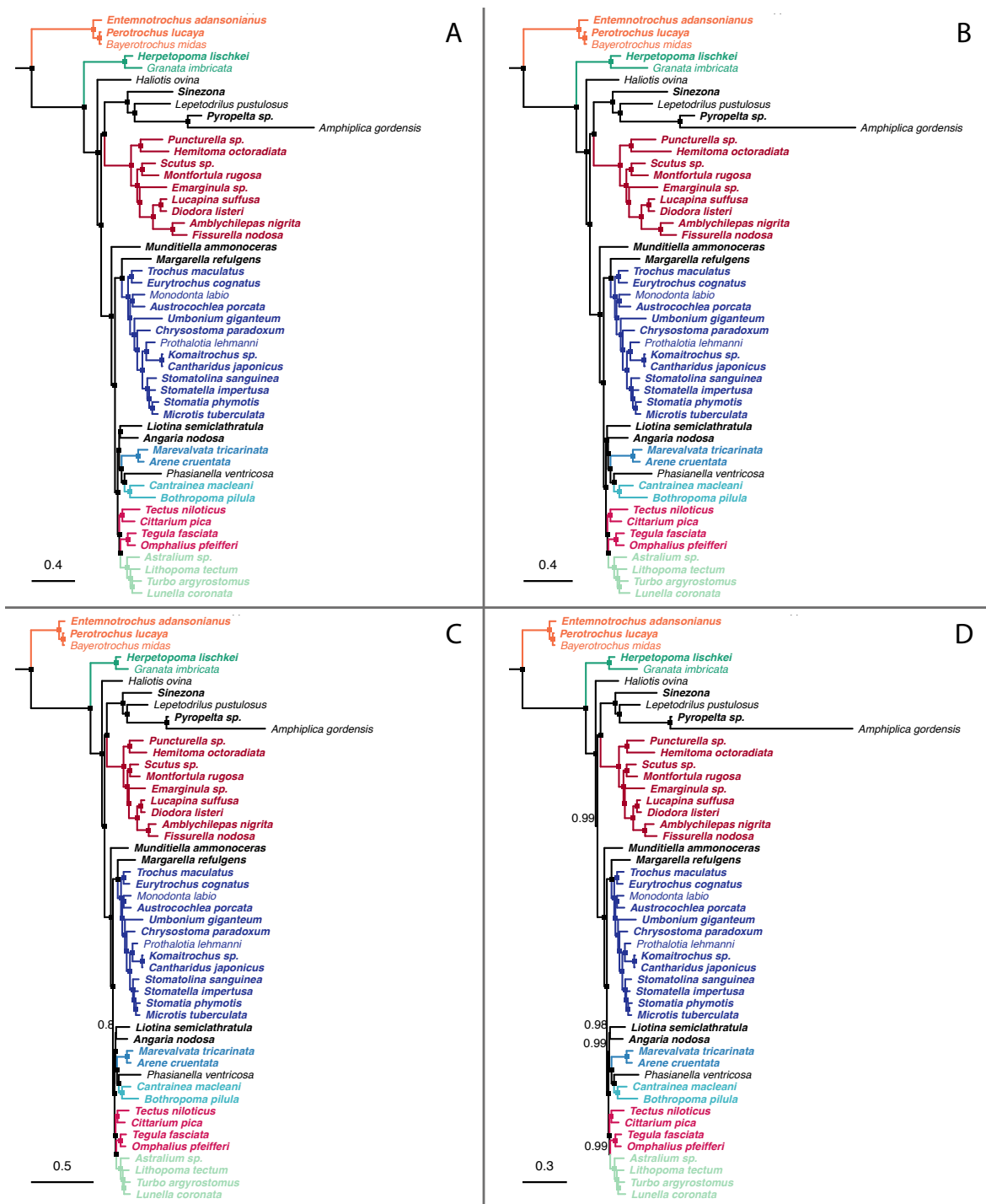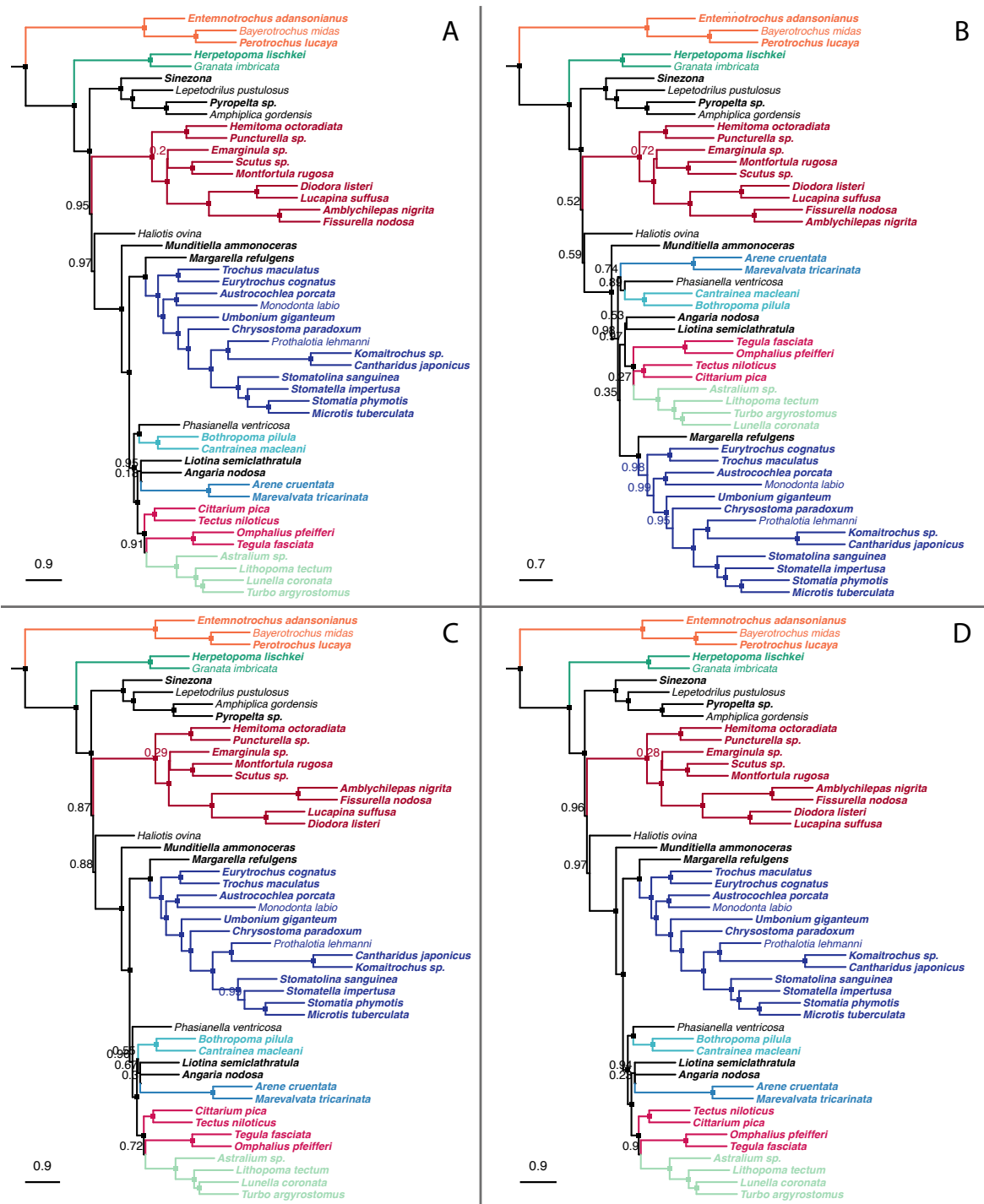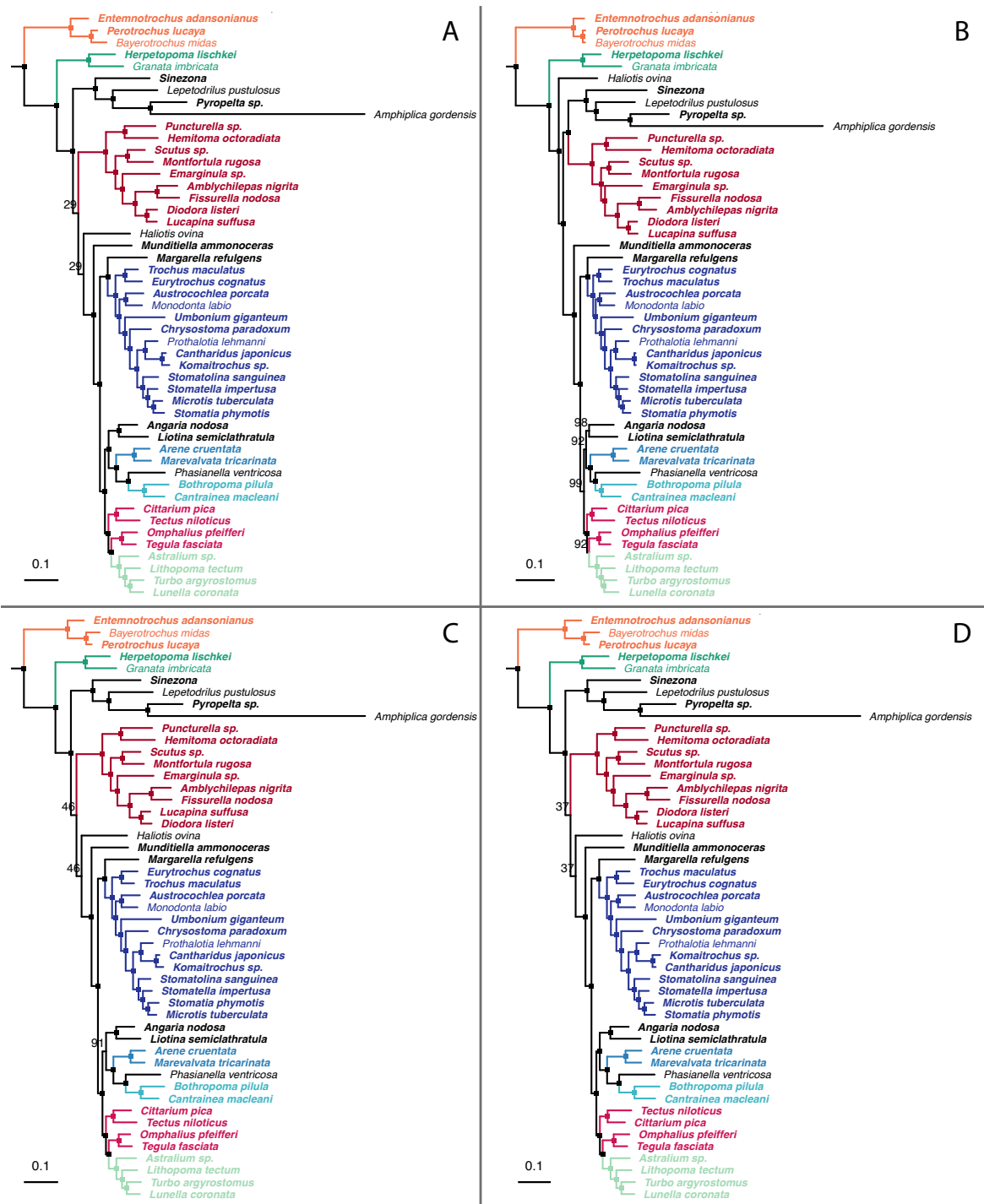
FIGURE A.2: Vetigastropod phylogeny inferred with Bayesian inference on each of four matrices: (A) matrix 1, (B) matrix 2, (C) matrix 3, (D) matrix 4. Single squares mark branches with full support; remaining nodes show the posterior probability. New transcriptomes in bold. See Methods for details.

FIGURE A.3: Vetigastropod phylogeny inferred with a shortcut coalescent-based method on each of four matrices: (A) matrix 1, (B) matrix 2, (C) matrix 3, (D) matrix 4. Single squares mark branches with full support; remaining nodes show the branch support value. New transcriptomes in bold. See Methods for details.

FIGURE A.4: Vetigastropod phylogeny inferred with maximum likelihood and genewise partitioning on each of four matrices: (A) matrix 1, (B) matrix 2, (C) matrix 3, (D) matrix 4. Single squares mark branches with full support; remaining nodes show the bootstrap value. New transcriptomes in bold. See Methods for details.

Bibliography

Aktipis, S. W. and G. Giribet (2010). "A phylogeny of Vetigastropoda and other "archaeogastropods": reorganizing old gastropod clades". In: *Invertebrate Biology* 129, pp. 220–240. DOI: 10.1111/j.1744-7410.2010.00198.x.

— (2011). "Testing relationships among the vetigastropod taxa: a molecular approach". In: *Journal of Molluscan Studies* 78, pp. 12–27. DOI: 10.1093/mollus/eyr023.

Altenhoff, A. M., N. M. Glover, C.-M. Train, K. Kaleb, A. Warwick Vesztrocy, et al. (2018). "The OMA orthology database in 2018: retrieving evolutionary relationships among all domains of life through richer web and programmatic interfaces". In: *Nucleic Acids Research* 46, pp. D477–D485. DOI: 10.1093/nar/gkx1019.

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

Capella-Gutierrez, S., J. M. Silla-Martinez, and T. Gabaldon (2009). "trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses". In: *Bioinformatics* 25, pp. 1972–1973. DOI: 10.1093/bioinformatics/btp348.

Chernomor, O., A. von Haeseler, and B. Q. Minh (2016). "Terrace aware data structure for phylogenomic inference from supermatrices". In: *Systematic Biology* 65, pp. 997–1008. DOI: 10.1093/sysbio/syw037.

Cunha, T. J. and G. Giribet (2019). "A congruent topology for deep gastropod relationships". In: *Proceedings of the Royal Society B: Biological Sciences* 286, p. 20182776. DOI: 10.1098/rspb.2018.2776.

Cunha, T. J., S. Lemer, P. Bouchet, Y. Kano, and G. Giribet (2019). "Putting keyhole limpets on the map: phylogeny and biogeography of the globally distributed marine family Fissurellidae (Vetigastropoda, Mollusca)". In: *Molecular Phylogenetics and Evolution* 135, pp. 249–269. DOI: 10.1016/j.ympev.2019.02.008.

Dayhoff, M. O., R. M. Schwartz, and B. C. Orcutt (1978). "A model of evolutionary change in proteins". In: *Atlas of Protein Sequence and Structure*. Ed. by M. O. Dayhoff. National Biomedical Research Foundation, pp. 345–352. DOI: 10.1.1.145.4315.

Degnan, J. H. and N. a. Rosenberg (2009). "Gene tree discordance, phylogenetic inference and the multispecies coalescent". In: *Trends in Ecology and Evolution* 24, pp. 332–340. DOI: 10.1016/j.tree.2009.01.009.

Edwards, S. V. (2009). "Is a new and general theory of molecular systematics emerging?" In: *Evolution* 63, pp. 1–19. DOI: 10.1111/j.1558-5646.2008.00549.x.

Foster, P. G. (2004). "Modeling compositional heterogeneity". In: *Systematic Biology* 53, pp. 485–495. DOI: 10.1080/10635150490445779.

— (2018). *p4 (Python package)*. URL: http://p4.nhm.ac.uk/tutorial/tut_compo.html.

Fryda, J., A. Nützel, and P. J. Wagner (2008). "Paleozoic Gastropoda". In: *Phylogeny and Evolution of the Mollusca*. Ed. by W. F. Ponder and D. R. Lindberg. University of California Press. Chap. 10, pp. 239–270.

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

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, et al. (2011). "Full-length transcriptome assembly from RNA-Seq data without a reference genome". In: *Nature Biotechnology* 29, pp. 644–652. DOI: 10.1038/nbt.1883.

Groves, L. T. and J. M. Alderson (2008). "Earliest record of the genus *Haliotis* (Mollusca: Gastropoda) from the Late Cretaceous (Campanian) of Los Angeles County, California". In: *Veliger* 50, pp. 24–26.

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

Harasewych, M. G., S. L. Adamkewicz, J. A. Blake, D. Saudeck, T. Spriggs, and C. J. Bult (1997). "Phylogeny and relationship of pleurotomariid gastropods (Mollusca: Gastropoda): an assessment based on partial 18S rDNA and cytochrome c oxydase I sequences". In: *Molecular Marine Biology and Biotechnology* 6, pp. 1–20.

Haszprunar, G. (1988). "On the origin and evolution of major gastropods group, with special reference to the streptoneura". In: *Journal of Molluscan Studies* 54, pp. 367–441.

Kalyaanamoorthy, S., B. Q. Minh, T. K. F. Wong, A. von Haeseler, and L. S. Jermiin (2017). "ModelFinder: fast model selection for accurate phylogenetic estimates". In: *Nature Methods* 14, pp. 587–589. DOI: 10.1038/nmeth.4285.

Kano, Y. (2008). "Vetigastropod phylogeny and a new concept of Seguenzioidea: independent evolution of copulatory organs in the deep-sea habitats". In: *Zoologica Scripta* 37, pp. 1–21. DOI: 10.1111/j.1463-6409.2007.00316.x.

Katoh, K. and D. M. Standley (2013). "MAFFT multiple sequence alignment software version 7: improvements in performance and usability". In: *Molecular Biology and Evolution* 30, pp. 772–780. DOI: 10.1093/molbev/mst010.

Lartillot, N., H. Brinkmann, and H. Philippe (2007). "Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model." In: *BMC evolutionary biology* 7 Suppl 1, S4. DOI: 10.1186/1471-2148-7-S1-S4.

Lartillot, N. and H. Philippe (2004). "A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process". In: *Molecular Biology and Evolution* 21, pp. 1095–1109. DOI: 10.1093/molbev/msh112.

Lartillot, N., N. Rodrigue, D. Stubbs, and J. Richer (2013). "PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment". In: *Systematic Biology* 62, pp. 611–615. DOI: 10.5061/dryad.c459h.

Laumer, C. E., H. Gruber-Vodicka, M. G. Hadfield, V. B. Pearse, A. Riesgo, J. C. Marioni, and G. Giribet (2018). "Support for a clade of Placozoa and Cnidaria in genes with minimal compositional bias". In: *eLife* 7, pp. 1–32. DOI: 10.7554/eLife.36278.

Le, S. Q., O. Gascuel, and N. Lartillot (2008). "Empirical profile mixture models for phylogenetic reconstruction". In: *Bioinformatics* 24, pp. 2317–2323. DOI: 10.1093/bioinformatics/btn445.

Lee, H., S. Samadi, N. Puillandre, M. H. Tsai, C. F. Dai, and W. J. Chen (2016). "Eight new mitogenomes for exploring the phylogeny and classification of Vetigastropoda". In: *Journal of Molluscan Studies* 82, pp. 534–541. DOI: 10.1093/mollus/eyw027.

Meiklejohn, K. A., B. C. Faircloth, T. C. Glenn, R. T. Kimball, and E. L. Braun (2016). "Analysis of a rapid evolutionary radiation using Ultraconserved Elements: evidence for a bias in some multispecies coalescent methods". In: *Systematic Biology* 65, pp. 612–627. DOI: 10.1093/sysbio/syw014.

Mirarab, S. and T. Warnow (2015). "ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes". In: *Bioinformatics* 31, pp. i44–i52. DOI: 10.1093/bioinformatics/btv234.

Molloy, E. K. and T. Warnow (2018). "To include or not to include: the impact of gene filtering on species tree estimation methods". In: *Systematic Biology* 67, pp. 285–303. DOI: 10.1093/sysbio/syx077.

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

Ponder, W. F. and D. R. Lindberg (1997). "Towards a phylogeny of gastropod molluscs: an analysis using morphological characters". In: *Zoological Journal of the Linnean Society* 119, pp. 83–265. DOI: 10.1006/zjls.1996.0066.

Quinn Jr., J. F. (1983). "A revision of the Seguenziacea Verrill, 1884 (Gastropoda: Prosobranchia). I. Summary and evaluation of the superfamily". In: *Proceedings of the Biological Society of Washington* 96, pp. 725–757.

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

Salvini-Plawen, L. V. and G. Haszprunar (1987). "The Vetigastropoda and the systematics of streptoneurous Gastropoda (Mollusca)". In: *Journal of Zoology* 211, pp. 747–770. DOI: 10.1111/j.1469-7998.1987.tb04485.x.

Sasaki, T. (1998). "Comparative anatomy and phylogeny of the Recent Archaeogastropoda (Mollusca: Gastropoda)". In: *The University of Tokyo Bulletin* 38, pp. 1–223.

Sayyari, E., J. B. Whitfield, and S. Mirarab (2017). "Fragmentary gene sequences negatively impact gene tree and species tree reconstruction". In: *Molecular Biology and Evolution* 34, pp. 3279–3291. DOI: 10.1093/molbev/msx261.

Schwentner, M., D. J. Combosch, J. Pakes Nelson, and G. Giribet (2017). "A Phylogenomic solution to the origin of insects by resolving crustacean-hexapod relationships". In: *Current Biology* 27, pp. 1818–1824. DOI: 10.1016/j.cub.2017.05.040.

Schwentner, M., S. Richter, D. C. Rogers, and G. Giribet (2018). "Tetraconatan phylogeny with special focus on Malacostraca and Branchiopoda: highlighting the strength of taxon-specific matrices in phylogenomics". In: *Proceedings of the Royal Society B: Biological Sciences* 285, p. 20181524. DOI: 10.1098/rspb.2018.1524.

Simão, F. A., R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, and E. M. Zdobnov (2015). "BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs". In: *Bioinformatics* 31, pp. 3210–3212. DOI: 10.1093/bioinformatics/btv351.

Smith, S. A. and C. W. Dunn (2008). "Phyutility: a phyloinformatics tool for trees, alignments and molecular data". In: *Bioinformatics* 24, pp. 715–716. DOI: 10.1093/bioinformatics/btm619.

Stamatakis, A. (2014). "RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies". In: *Bioinformatics* 30, pp. 1312–1313. DOI: 10.1093/bioinformatics/btu033.

Uribe, J. E., D. J. Colgan, L. R. Castro, Y. Kano, and R. Zardoya (2016). "Phylogenetic relationships among superfamilies of Neritimorpha (Mollusca: Gastropoda)". In: *Molecular Phylogenetics and Evolution* 104, pp. 21–31. DOI: 10.1016/j.ympev.2016.07.021.

Uribe, J. E., I. Irisarri, J. Templado, and R. Zardoya (2019). "New patellogastropod mitogenomes help counteracting long-branch attraction in the deep phylogeny of gastropod mollusks". In: *Molecular Phylogenetics and Evolution* 133, pp. 12–23. DOI: 10.1016/j.ympev.2018.12.019.

Uribe, J. E., S. T. Williams, J. Templado, S. Abalde, and R. Zardoya (2017). "Denser mitogenomic sampling improves resolution of the phylogeny of the superfamily Trochoidea (Gastropoda: Vetigastropoda)". In: *Journal of Molluscan Studies* 83, pp. 111–118. DOI: 10.1093/mollus/eyw049.

Wang, H.-C., E. Susko, and A. J. Roger (2019). "The relative importance of modeling site pattern heterogeneity versus partition-wise heterotachy in phylogenomic inference". In: *Systematic Biology*, pp. 1–56. DOI: 10.1093/sysbio/syz021.

Williams, S. T. (2012). "Advances in molecular systematics of the vetigastropod superfamily Trochoidea". In: *Zoologica Scripta* 41, pp. 571–595. DOI: 10.1111/j.1463-6409.2012.00552.x.

Williams, S. T., K. M. Donald, H. G. Spencer, and T. Nakano (2010). "Molecular systematics of the marine gastropod families Trochidae and Calliostomatidae (Mollusca: Superfamily Trochoidea)". In: *Molecular Phylogenetics and Evolution* 54, pp. 783–809. DOI: 10.1016/j.ympev.2009.11.008.

Williams, S. T., S. Karube, and T. Ozawa (2008). "Molecular systematics of Vetigastropoda: Trochidae, Turbinidae and Trochoidea redefined". In: *Zoologica Scripta* 37, pp. 483–506. DOI: 10.1111/j.1463-6409.2008.00341.x.

Williams, S. T. and T. Ozawa (2006). "Molecular phylogeny suggests polyphyly of both the turban shells (family Turbinidae) and the superfamily Trochoidea (Mollusca: Vetigastropoda)". In: *Molecular Phylogenetics and Evolution* 39, pp. 33–51. DOI: 10.1007/3-540-30023-6_3.

WoRMS (2019). *World Register of Marine Species*. URL: http://www.marinespecies.org.

Wort, E. J., P. B. Fenberg, and S. T. Williams (2017). "Testing the contribution of individual genes in mitochondrial genomes for assessing phylogenetic relationships in Vetigastropoda". In: *Journal of Molluscan Studies* 83, pp. 123–128. DOI: 10.1093/mollus/eyw044.

Yoon, S. H. and W. Kim (2005). "Phylogenetic relationships among six vetigastropod subgroups (Mollusca, Gastropoda) based on 18S rDNA sequences." In: *Molecules and cells* 19, pp. 283–288.

Yu, G., D. K. Smith, H. Zhu, Y. Guan, and T. T.-Y. Lam (2017). "ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data". In: *Methods in Ecology and Evolution* 8. Ed. by G. McInerny, pp. 28–36. DOI: 10.1111/2041-210X.12628.

Zapata, F., N. G. Wilson, M. Howison, S. C. S. Andrade, K. M. Jorger, M. Schrodl, F. E. Goetz, G. Giribet, and C. W. Dunn (2014). "Phylogenomic analyses of deep gastropod relationships reject Orthogastropoda". In: *Proceedings of the Royal Society B: Biological Sciences* 281, p. 20141739. DOI: 10.1098/rspb.2014.1739.