



Cite this article: Lemer S, González VL, Bieler R, Giribet G. 2016 Cementing mussels to oysters in the pteriomorphian tree: a phylogenomic approach. *Proc. R. Soc. B* **283**: 20160857.
<http://dx.doi.org/10.1098/rspb.2016.0857>

Received: 15 April 2016

Accepted: 3 June 2016

Subject Areas:

evolution, genomics, taxonomy and systematics

Keywords:

phylogenomics, genome, phylogeny, Mollusca, Bivalvia, evolutionary rate

Author for correspondence:

Sarah Lemer

e-mail: sarah.lemer@gmail.com

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rspb.2016.0857> via <http://rspb.royalsocietypublishing.org>.

Cementing mussels to oysters in the pteriomorphian tree: a phylogenomic approach

Sarah Lemer¹, Vanessa L. González², Rüdiger Bieler³ and Gonzalo Giribet¹

¹Museum of Comparative Zoology, Department of Organismic and Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, MA 02138, USA

²Department of Invertebrate Zoology, National Museum of Natural History, Smithsonian Institution, Washington, DC 20013, USA

³Integrative Research Center, Field Museum of Natural History, 1400 South Lake Shore Drive, Chicago, IL 60605, USA

SL, 0000-0003-0048-7296

Mussels (Mytilida) are a group of bivalves with ancient origins and some of the most important commercial shellfish worldwide. Mytilida consists of approximately 400 species found in various littoral and deep-sea environments, and are part of the higher clade Pteriomorphia, but their exact position within the group has been unstable. The multiple adaptive radiations that occurred within Pteriomorphia have rendered phylogenetic classifications difficult and uncertainty remains regarding the relationships among most families. To address this phylogenetic uncertainty, novel transcriptomic data were generated to include all five orders of Pteriomorphia. Our results, derived from complex analyses of large datasets from 41 transcriptomes and evaluating possible pitfalls affecting phylogenetic reconstruction (matrix occupancy, heterogeneity, evolutionary rates, evolutionary models), consistently recover a well-supported phylogeny of Pteriomorphia, with the only exception of the most complete but smallest data matrix (*Matrix 3*: 51 genes, 90% gene occupancy). Maximum-likelihood and Bayesian mixture model analyses retrieve strong support for: (i) the monophyly of Pteriomorphia, (ii) Mytilida as a sister group to Ostreida, and (iii) Arcida as sister group to all other pteriomorphians. The basal position of Arcida is congruent with its shell microstructure (solely composed of aragonitic crystals), whereas Mytilida and Ostreida display a combination of a calcitic outer layer with an aragonitic inner layer composed of nacre tablets, the latter being secondarily lost in Ostreoida.

1. Introduction

Mussels (the members of the superfamily Mytiloidea) are a ubiquitous and common group of bivalves and a main source of protein for humans and non-human animals alike. Despite being the dominant organisms in many littoral, shallow sub-littoral, deep-sea hydrothermal vent and cold seep ecosystems [1]—including rocky and sediment shores on open coasts and in estuaries and marshes [2]—the position of mussels in relation to other pteriomorphian bivalves remains unresolved (e.g. [3]). Because of their economic and ecological importance, mussels have been the subject of considerable research effort. Taxonomically, Mytiloidea is a diverse group of pteriomorphian bivalves recognized by characteristics of shell form, hinge and mussel attachment scars [4].

Mytilidae, by most authors considered the only family in the superfamily Mytiloidea and the order Mytilida (but see Carter [5]), is an ancient group with roots extending to the Devonian [6]. It comprises eight recognized extant subfamilies [7,8], with approximately 400 species worldwide [9], probably constituting the largest family of bivalves in number of species. Like mussels, in particular, bivalves of the subclass Pteriomorphia in general form a commercially important clade with

species harvested and cultured worldwide both for food (e.g. oysters, scallops, ark shells), and ornament industry (pearl oysters). They are characterized by a large variability in ligament structure and in shell size, shape and composition. Some shells are constituted of aragonite tablets only (e.g. Arcida), others of calcite crystals (e.g. Ostreoida), whereas some species present a shell that has both crystal forms of calcium carbonate, calcite and aragonite (e.g. Mytilida, Pectinida, Limida, Pinnoidea and Pterioidea) [10]. Shell composition is important because most pteriomorphians are sessile as adults, their calcified shells being the only mechanism against predation and desiccation [11]. In addition, understanding the evolution of shell formation can help assess the impacts of ocean acidification on bivalves.

Pteriomorphia appears in the fossil record in the Early Ordovician where mytiloid, arcoid and pteroid forms are recognized. These groups experienced periods of radiating diversification in the Late Devonian, as well as in the beginning and the end of the Mesozoic [12]. The diversity of this group reflects several adaptive radiations and the shell forms have attracted the interest of many palaeontologists and neontologists (e.g. [13–21]). The multiple adaptive radiations have rendered phylogenetic classifications difficult owing to convergence and/or parallel evolution at various levels [22]. Although the most recent phylogenies agree on the monophyly of Pteriomorphia, uncertainty remains regarding the internal relationships of the majority of pteriomorphian clades (families and superfamilies), despite considerable phylogenetic (e.g. [16,18,23–27]) and even recent phylogenomic [28] efforts.

Pteriomorphia comprises five widely recognized extant orders (Arcida, Limida, Mytilida, Ostreida and Pectinida) [27]. Among the most unstable results in bivalve phylogeny are the relative positions of Arcida and Mytilida ([3,29]). As one of the most diverse and studied bivalve groups, it is unsettling that to this day relationships among pteriomorphians in general and mytiloids in particular remain unresolved, impinging on future studies aiming at exploring extinction, diversification and biogeographic patterns of this group, including dating and inference of the evolution of lineages through time. With the aim of improving resolution within Pteriomorphia, we generated a novel Illumina-based dataset to evaluate the internal branching patterns of this clade from a fresh perspective and infer on the evolution of shell microstructure within pteriomorphians. For this, we sequenced, assembled and analysed 12 new pteriomorphian transcriptomes and combined them with 13 additional transcriptomes previously generated by us [28], and 22 transcriptomes and one genome from publicly available data, for a total of 41 pteriomorphian samples and eight outgroups. The taxonomic sampling used here reflects what is now acknowledged to be the span of pteriomorphian diversity.

2. Material and methods

(a) Taxon sampling, cDNA library construction and next-generation sequencing

We sequenced cDNA from 12 pteriomorphian specimens using an Illumina HiSeq 2500 platform and combined these with 35 published transcriptomes and one genome, including 13 libraries previously sequenced in our laboratory [28]. Information about the sampled specimens can be found in the electronic supplementary material, table S1 and in the MCZ online collections database (<http://mczbase.mcz.harvard.edu>). All tissues were collected fresh and immediately flash frozen in liquid nitrogen or fixed in

RNAlater[®] (Life Technologies, Carlsbad, CA, USA) and stored at -80°C . Total RNA was extracted using TRIzol (Life Sciences) and purification of mRNA was performed using the Dynabeads (Invitrogen) following the manufacturer's instructions and as described in [28]. For each sample, quality of mRNA was assessed with a picoRNA assay in an Agilent 2100 Bioanalyzer (Agilent Technologies) and quantity measured with an RNA assay in a Qubit fluorometer (Life Technologies).

All cDNA libraries were constructed using the PrepX mRNA kit for Apollo 324 (Wafergen). Libraries were sequenced on the Illumina HiSeq 2500 platform with paired-end reads of 150 bp at the FAS Center for Systems Biology at Harvard University, after their concentration and quality were assessed.

(b) Transcriptome assembly

All reads generated for this study are deposited in the National Center for Biotechnology Information Sequence Read Archive (NCBI-SRA; electronic supplementary material, table S1). Each sample, except for the genome of *Pinctada fucata*, was prepared as in [28], as detailed in the electronic supplementary material, S2. De novo assemblies were conducted for each sample with Trinity r2014-04-13 [30,31] using paired read files and default parameters except for '-path_reinforcement_distance 50', which seems to produce slightly better assemblies, with higher N50 values and longer contigs. Reduction of redundant reads, peptide prediction and peptide filtration were conducted as in [28] (detailed in the electronic supplementary material, S2).

(c) Orthology assignment and matrix construction

Orthology assignment for the dataset assemblies was performed using stand-alone OMA v. 0.99z.2 [32,33] (detailed parameters in the electronic supplementary material, S2). Three initial data matrices following occupancy thresholds [34] were generated for phylogenetic analyses: the first one, *Matrix 1*, targeting a minimum gene occupancy of 50%, was constructed by selecting the OMA orthogroups present in 24 or more taxa (1205 orthogroups). *Matrix 2* includes orthogroups present in 36 or more taxa (gene occupancy greater than 75%; 277 orthogroups). *Matrix 3* was constructed by selecting the orthogroups found in 44 or more taxa (greater than 90% gene occupancy; 51 orthogroups). The orthogroup selection based on minimum taxon occupancy was executed using a custom Python script (all scripts and data matrices are available online at https://dataverse.harvard.edu/dataverse/Pteriomorphia_phylogenomics). Alignments were generated for each matrix using MUSCLE v. 3.6 [35] (details in the electronic supplementary material, S2).

In order to assess the effects of rate of molecular evolution and heterotachy on tree topology, 10 additional matrices were constructed by selecting subsets of *Matrix 2* based on evolutionary rate, for which per cent pairwise identity was employed as a proxy (see figure 1 and the electronic supplementary material, S2, for details). This method was chosen to approximate evolutionary rate because it is agnostic to tree topology. Accumulated conservation values were generated for each locus using TRIMAL 1.2b (-sct flag). Matrices were produced as follows: (i) *Matrices A–D* were constructed with incremental addition of loci to produce four matrices (figure 1). This strategy has been used in previous phylogenomic analyses [36–38], but masks the possible contribution of the fastest evolving genes, as the slowest ones are always used. In order to investigate the contribution of the different blocks of evolutionary rates in the absence of the other genes, we designed another strategy: *Matrices E–J* were constructed by parsing 50 loci matrices and one 27 loci matrix, with no addition (figure 1). Compositional heterogeneity among taxa and within each orthogroup can affect phylogenetic results and lead to incorrect tree reconstructions. In order to discern if this was the case in our dataset we used the package BACoCa v. 1.1r [39]

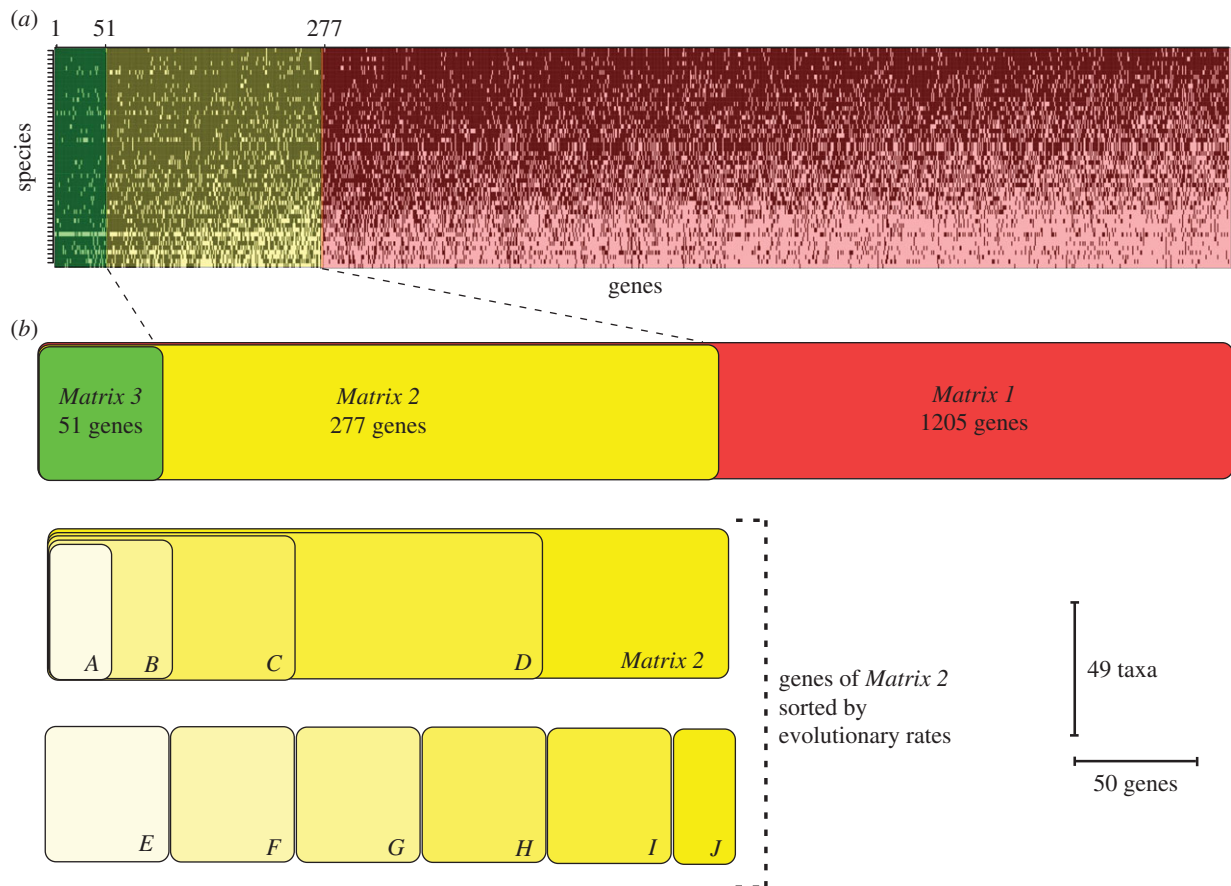


Figure 1. (a) Gene occupancy representation per species with maximum occupancy towards the top left. A light cell indicates a non-sampled gene. *Mytilus galloprovincialis* is the best-represented species, whereas *Saccostrea palmula* is the worst-represented one. *Matrix 1* is represented in red, *Matrix 2* in yellow and *Matrix 3* in green. (b) Schematic of the 13 matrices generated for the phylogenetic, per cent pairwise identity and compositional heterogeneity analyses. Colour codes are as in (a). (Online version in colour.)

to estimate relative composition frequency variability (RCFV) in *Matrix 2*.

(d) Phylogenetic analyses

All matrices, but *Matrix 1*, were analysed using maximum-likelihood inferences conducted by PHYML-PCMA [40]. PHYML-PCMA estimates a model through the use of a principal component (PC) analysis. We selected 10 PCs in the PHYML-PCMA analyses. Because of its large size and the intense computation required by PHYML-PCMA, *Matrix 1* was analysed with EXAML v. 3.0 [41] (see the electronic supplementary material, S2, for detailed parameters).

The initial three matrices (*Matrices 1, 2* and *3*) were also analysed using Bayesian inference with EXABAYES v. 1.21 with openmpi v. 1.64 ([42]; see the electronic supplementary material, S2, for detailed parameters). Additional Bayesian tree searches were also conducted for *Matrices 2* and *3* in PHYLOBAYES MPI v. 1.4e [43] using the site-heterogeneous CAT-GTR model of evolution [44]. Four independent Markov chain Monte Carlo (MCMC) runs were conducted for 2196–3187 cycles. The initial cycles in each MCMC run were discarded as burn-in and determined using the ‘tracecomp’ executable. Convergence was assessed using the ‘bpcomp’ executable and chains were considered to have converged when the maximum bipartition discrepancies (maxdiff) across a minimum of two independent chains reached 0.2.

Finally, because *Matrix 3* showed inconsistent tree topologies across analyses (see Results), we conducted additional tree searches after partitioning this matrix into genes, using RAxML 7.7.5 [45] with the GAMMA model of rate heterogeneity, the WAG protein substitution model and 1000 bootstraps.

To test for putative gene incongruence, we inferred individual gene trees for each orthogroup included in each of the three initial matrices (*Matrices 1, 2* and *3*) using RAxML 7.7.5 [45] and SUPERQ v. 1.1 [46] (see the electronic supplementary material, S2, for detailed parameters).

3. Results

(a) Phylogenetic relationships based on the three main matrices

The number of sequence reads, used reads, accession numbers, contigs, and other values to assess the quality of the assembled transcriptomes, can be found in the electronic supplementary material, table S1. Orthology assessment of this 49-taxon dataset with the OMA stand-alone algorithm recovered 149 182 orthogroups. The three super-matrices generated yielded 1205 (*Matrix 1*: occupancy of more than 50%; 316 219 aa), 277 (*Matrix 2*: occupancy of more than 75%; 64 318 aa) and 51 (*Matrix 3*: occupancy of more than 90%; 11 066 aa) orthologs, respectively.

All the phylogenetic analyses conducted on all three matrices revealed well-supported consistent topologies for most pteriomorphian superfamilies (figure 2). The maximum-likelihood (ML: PHYML, EXAML and RAxML) and Bayesian (EXABAYES and PHYLOBAYES) phylogenetic analyses conducted on the three main matrices (*Matrix 1, Matrix 2, Matrix 3*) recovered the monophyly of all five pteriomorphian orders, superfamilies and families, except for Arcoidea, which was

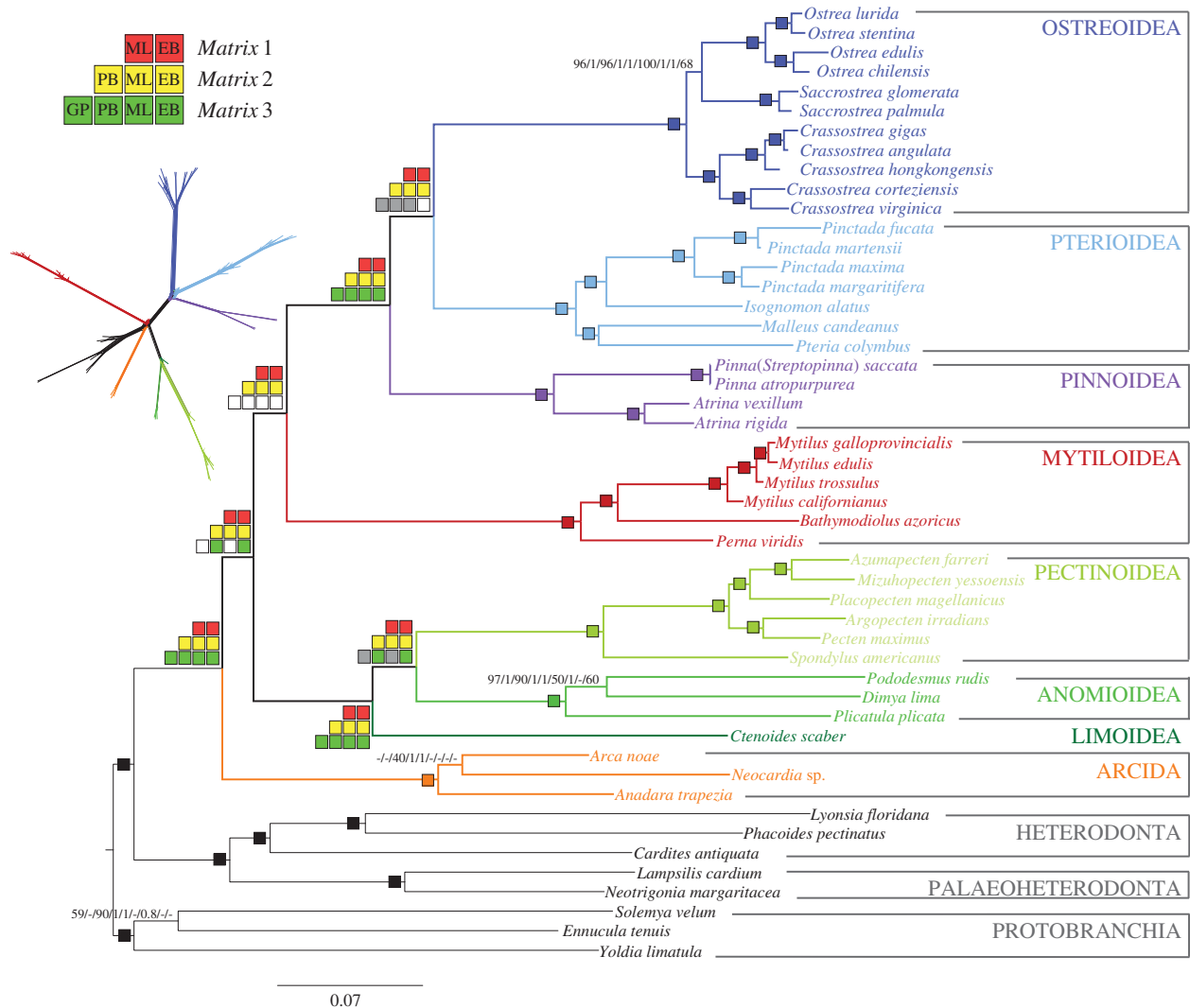


Figure 2. Phylogenetic hypothesis based on *Matrix 2* analysed in PHyML-PCMA ($\text{Ln } L = -1\,132\,164.323494$) with support values plotted as follows: checked boards in major deep nodes represent nodal support for the different analyses in *Matrix 1* (red), *Matrix 2* (yellow) and *Matrix 3* (green). PHyML-PCMA and EXA ML are abbreviated as ML, EXA BAYES: EB, PHYLO BAYES: PB, and RAxML with gene partitions: GP. Filled squares indicate nodal support values higher than 99% BS (ML) and a pp of 0.99 or higher (EB). Grey squares indicate lower nodal support and white squares indicate unrecovered nodes in the specified analysis. Single squares on internal nodes indicate maximum support in all six analyses. Lower nodal support values of internal nodes are reported as follows: *Matrix 1* ML/*Matrix 1* EB/*Matrix 2* ML/*Matrix 2* EB/*Matrix 2* PB/*Matrix 3* ML/*Matrix 3* EB/*Matrix 3* PB/*Matrix 3* GP. Pteriomorphian families are represented in different shades of colour. Top left: supernetwork representation of quartets derived from the individual ML gene trees for *Matrix 2*. The same colour scheme is applied to all figures. (Online version in colour.)

not recovered as monophyletic and included Limopsoidea in all analyses conducted with *Matrix 2* (figure 2). Ostreoida and Pterioidea appeared as sister clades with maximum support (100% bootstrap support (BS) and a posterior probability (pp) of 1.00; for all ML and Bayesian analyses, respectively) with Pinnoidea as their sister group in all analyses except in analyses conducted with *Matrix 3*, mostly refuting prior morphology-based hypotheses (e.g. [47]; figure 4). The relationships among superfamilies of Pectinida were consistent and well supported in all analyses: Anomioidea and Pectinoidea were sister clades (figure 2). Similarly, Limida was unvaryingly recovered as the sister group to Pectinida with 100% BS or a pp of 1.00 in all analyses and with all the analysed matrices.

The position of Mytilida was consistent and well supported in the five analyses with *Matrices 1* and *2* (figure 2), constituting the sister group to Ostreida with a 99% BS and a pp of 1.00. However, the phylogenetic analyses with *Matrix 3* were inconsistent across methods and placed Mytilida either more basally, as sister group to Arcida (ML and PHYLO BAYES; 27% BS and pp = 1.00, respectively), as sister group to the clade constituted by Pectinida and Limida (EXA BAYES; pp = 0.89), or as a sister

group to all pteriomorphians except Arcida (RAxML with gene partitions; 47% BS). The lack of nodal support for this smallest matrix thus indicates that not enough information was available in *Matrix 3* to resolve the position of Mytilida and that more than 51 genes may be needed to resolve this ancient divergence. Nonetheless, Arcida was always recovered at the base of the pteriomorphian tree, either as sister group to all other pteriomorphians, or as sister group to Mytilida, even with *Matrix 3*.

The supernetworks obtained for each matrix using SUPERQ v. 1.1, displayed a tree-like structure with relatively long edges and similar topologies as to the concatenated species trees (figure 2; electronic supplementary material, figure SA). All three networks show a long edge leading to the clade formed by Ostreida (*sensu* [27]); however, they also indicated gene conflict for the position of Mytilida with respect to Arcida, and the clade including Limida and Pectinida, especially in the supernetwork for *Matrix 3*. Reticulations were also visible in the supernetwork obtained with *Matrix 3* relative to the position of Pterioidea and gene conflict was detected in more derived nodes within Ostreida,

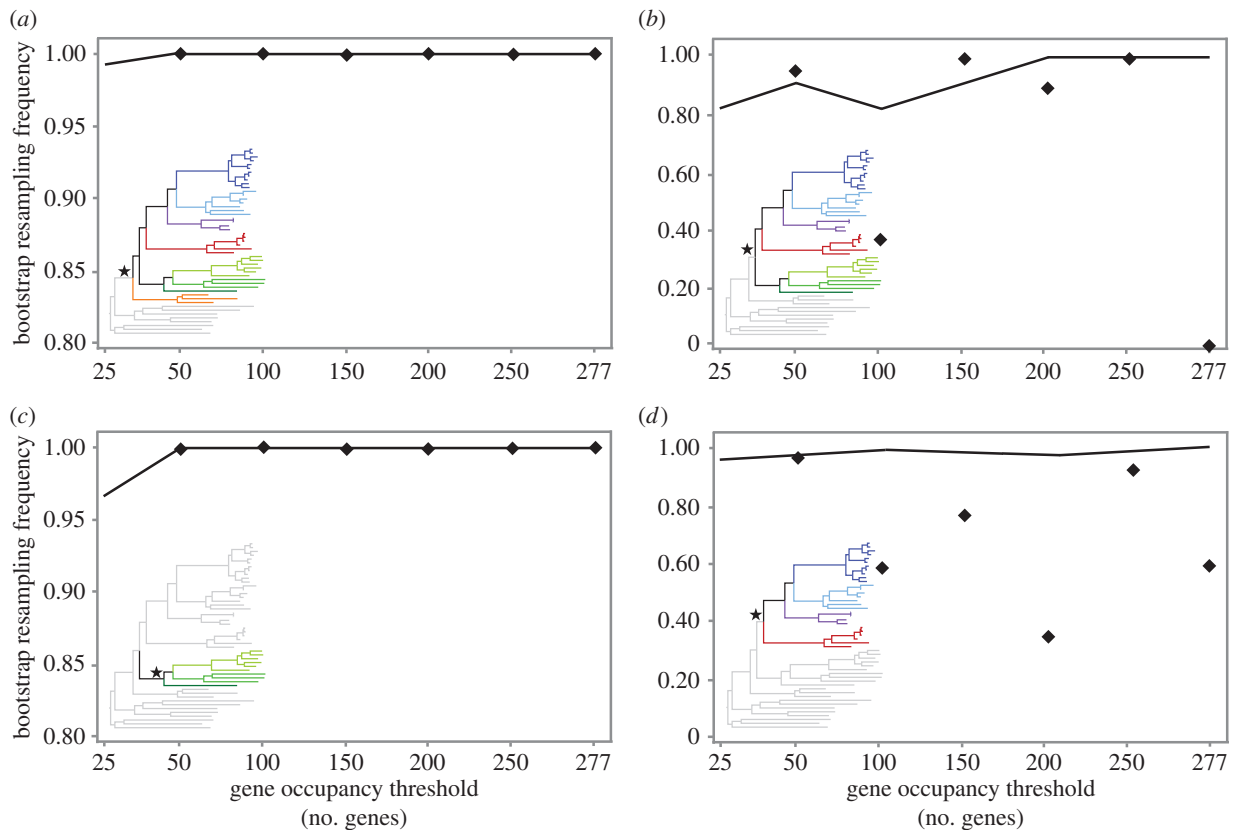


Figure 3. Bootstrap resampling frequency for phylogenetic hypotheses (a) monophyly of Pteriomorphia, (b) basal position of Arcida (=monophyly of all other pteriomorphians), (c) monophyly of Limida + Pectinida, (d) Mytilida sister group to Ostreida; inferred from *Matrices A–D* and *Matrix 2*, represented in full line (incremental addition of orthogroups in order of per cent pairwise identity, from most to least conserved); *Matrices E–J*, represented by diamonds (increments of non-additive 50 orthogroups in order of per cent pairwise identity, from most to least conserved). (Online version in colour.)

probably reflecting on the poor quality of the *P. fucata* genome (electronic supplementary material, figure SA).

(b) Concatenation by per cent pairwise identity and compositional heterogeneity

As the position of Mytilida has been in flux, we focused our analyses on the putative node uniting Mytilida with Ostreida. When evaluating the matrices obtained by sequentially adding genes based on their evolutionary rates (*Matrices A–D* and *Matrix 2*; see figure 1 for details), we observed a monotonic trajectory of nodal support increasing from 97% to 100% BS (full line in figure 3d). A similar result was obtained for the monophyly of Pteriomorphia and the sister group relationship of Limida and Pectinida (full lines in figure 3a,c). The nodal support for the basal placement of Arcida, i.e. monophyly of all other pteriomorphians (full line in figure 3b), showed a more gradual and fluctuating increase when adding sets of genes with higher evolutionary rates, going from 81% to 100% BS. By contrast, when looking at *Matrices E–J* (no addition; figure 1), we observed a fluctuating trajectory of BS for the sister group relationship of Mytilida and Ostreida, with maximum support obtained with the 50 most conserved genes (96% BS) but thereafter nodal support oscillated between 35 and 93%, with no clear trend (diamonds in figure 3d). The support for the basal placement of Arcida also showed some fluctuation: between 90 and 100% BS was found for all but two sets of 50 genes. The second set of fastest evolving genes only displayed 37% BS and the last set of the fastest-evolving 27 genes failed to support the topology altogether and instead placed

Arcida as sister group to Pectinida (diamonds in figure 3b). The latter placement of Arcida was not retrieved in any of the conducted phylogenetic analyses and probably reflects on the poor ability of a small subset of fast evolving genes to recover deep phylogenetic nodes in Pteriomorphia. The monophyly of Pteriomorphia and the sister group relationship of Limida and Pectinida (diamonds in figure 3a,c) were supported in all analyses, which found 100% BS for all sets of 50 genes, irrespective of their evolutionary rate.

The RCFV per taxon and per amino acid ranged from 0.0001 to 0.007 (electronic supplementary material, figure SB), indicating compositional homogeneity throughout all of the amino acids and taxa included in *Matrix 2* and thus eliminating possible biases owing to compositional heterogeneity.

4. Discussion

(a) Cementing mussels to oysters

The phylogenetic dataset generated in this study is, to our knowledge, the largest ever gathered to clarify the position of mussels and to test relationships with their closest relatives. As in many other studies, our results support the monophyly of Pteriomorphia with maximum support and this node is found under all analyses. Nevertheless, the tree topologies recovered by the ML and Bayesian analyses of *Matrix 3* (51 genes) recapitulate a recurring phenomenon in molecular phylogenetic studies of Pteriomorphia: the unstable position of Mytilida. A major concern in phylogenetic reconstruction has been the amount of missing data,

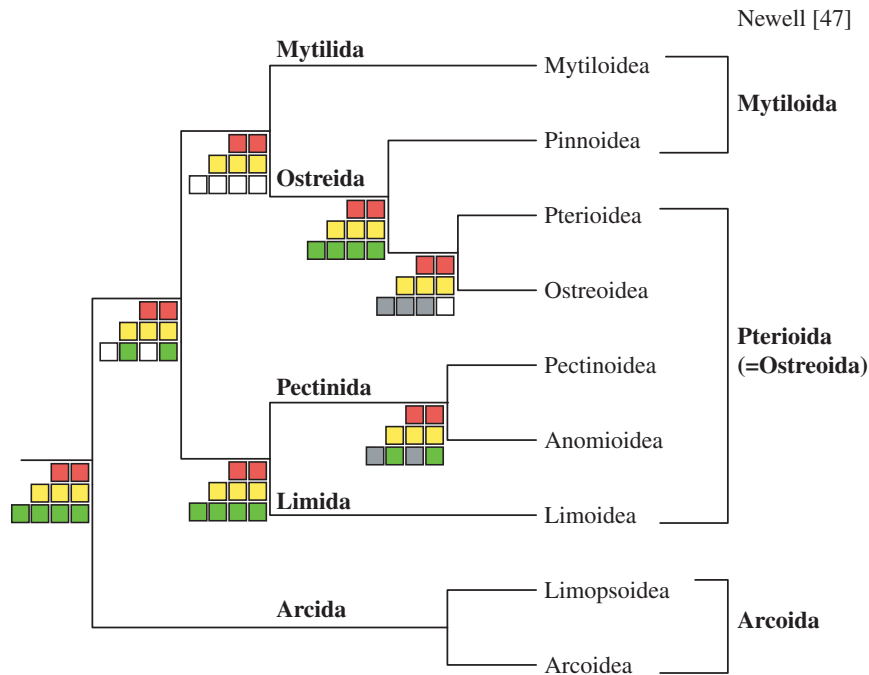


Figure 4. Summary tree representing relationships among the superfamilies and orders of Pteriomorpha. Checked boards are coded as in figure 2 and represent nodal support provided by *Matrices 1–3*. Newell's [47] classification is given for comparison. (Online version in colour.)

as they can produce misleading estimates of topology and branch lengths (e.g. [48,49]). However, with the generalized use of transcriptome-based data matrices for phylogenetic reconstruction, large matrices, even if typically incomplete, have shown the ability to reconstruct relationships with strong support [34]. Perhaps, less intuitive has been the finding that in some cases the most complete matrices are not always providing the most accurate reconstructions, even if nearing 100% completeness, a poorly studied phenomenon, but that may be owing to limited amount of available genetic information in the genes selected for completeness [50]. In our case, *Matrix 3* is the only one with conflicting topologies when using different phylogenetic methods. The PHYLML and PHYLOBAYES analyses conducted with *Matrix 3* place Mytilida as the sister group of Arcida, these two constituting the sister group of all remaining Pteriomorpha; whereas the EXABAYES and gene-partitioned RAxML analyses place Mytilida as sister group to the clade formed by Pectinida and Limida or as sister group to all pteriomorphians excluding Arcida, respectively. None of these topologies, however, received significant support. We thus conclude that the position of Mytilida resolved by the two larger matrices, as sister group to Ostreida, is the best corroborated hypothesis; this result is also recovered by several other submatrices of genes sorted by evolutionary rate (*Matrices A–D, E, I*; figures 1 and 3*d*), including slow evolving genes (*Matrices A, B, E*) and fast evolving genes (*Matrices C, D, I*). The fact that several small matrices, composed of different sets of genes, yield different tree topologies probably explains the disparity of phylogenetic results from previous molecular analyses (e.g. [18,24,26,27]).

The position of Arcida has also been unstable throughout a series of molecular data analyses [18,24,26,27] and hardly any morphological analysis has proposed a sister group relationship to all other pteriomorphians, as many authors have placed Mytilida in a more basal position (e.g. [12,16]). Our results, however uniformly place Arcida in a basal position, as sister group to all other pteriomorphians (as in the morphological analyses of Giribet & Wheeler [24,27]),

with the exception of two analyses (PHYML and PHYLOBAYES, *Matrix 3*) that place Mytilida as their sister group—a partially unsupported relationship (27% BS, but pp = 1.00) that had been suggested previously [51]. These findings do not support the hypothesis of Carter *et al.* [5] that arranged Pteriomorpha into two major clades, Mytilomorphi (with Mytilida) and Ostreomorphi (with the other groups treated herein).

(b) Implications for the evolution of shell microstructure

Early bivalves built aragonite shells, a character retained in Nuculida, Trigoniida, Unionida and most Anomalodesmata [12,52,53]. Pteriomorphians all develop an outer calcitic layer in addition to the inner aragonitic shell layer, with the exception of Arcida, whose shell microstructure still consists of both outer and inner crossed-lamellar layers of aragonite crystals only [10,27]. Since Arcida displays a basal position with respect to other pteriomorphians, this character requires no homoplasy, as arcoids present the plesiomorphic state found in closely related bivalve outgroups. The development of an outer calcitic layer is thought to be a recent adaptation against shell dissolution in cold and undersaturated waters [54,55]. There are two main types of calcite prisms from the viewpoint of the microstructure: simple prismatic or foliated. The simple prismatic microstructure is found in the outer layers of Pterioidea (except for Malleidae), Pinnoidea and Mytilida (the latter has the particularity of also being fibrous) [10,14]. The outer simple calcitic prismatic layer is a flexible structure that allows tight sealing of the two valves, thus better isolating the inner chamber from changing ambient water conditions and minimizes damage from unsuccessful predation. This shell ultrastructure is usually associated with epi- and endo-byssate life-habits (e.g. Mytilida, Pterioidea, Pinnoidea). By contrast, the clades that developed a foliated calcite layer (Pectinida, Ostreoidea and the family Malleidae) gained the opportunity to produce more convex valves with a wide repertoire of ornamentations and new

shell morphologies, making this character a key innovation in the adaptive radiation of Pteriomorphia. The foliate microstructure is a derived feature that is more rigid but less resistant to breakage than the simple prismatic structure [56]; clades that adopted this arrangement are more widespread, have higher generic diversity and have repeatedly evolved more derived life-habits, e.g. cementing (Ostreoidea), swimming (Pectinoidea) or free reclining (Anomioidea) [57]. The calcitic outer layer of Limida consists of a combination of calcitic fibrous prismatic and crossed-foliated structures, and because this combination is not observed in the other pteriomorphians it has been argued to justify the ordinal level of Limida [19]. The inner aragonite shell layer of all pteriomorphians is composed of cross-lamellar crystals, with the exception of Ostreida and some Mytilida, which display nacreous tablets (Pinnoidea and Pterioidea) or in the case of Ostreoidea, a complete loss of aragonite [58,59]. In addition, the clade comprising Limida and Pectinida is the only one that combines an outer foliated calcitic structure with an inner crossed-lamellar aragonitic layer, which makes it a synapomorphy for this clade. Altogether, the distribution of this character is concordant with the classification recovered by our analyses. Interestingly, both mussels and oysters have compound latero-frontal cilia in their gills, unlike the rest of pteriomorphians [60].

(c) Revised higher-level classification

Our results, derived from complex analyses of large datasets, and evaluating possible pitfalls affecting phylogenetic reconstruction (matrix occupancy, heterogeneity, evolutionary rates, evolutionary models) consistently recover a well-supported phylogeny of Pteriomorphia, with the only exception of the most complete (but smallest) *Matrix* 3. After exploring this potential phylogenetic inconsistency, we conclude that this result is not necessarily owing to the

small matrix size, but that it may be related to the specific genes of this partition, which cause considerable conflict (electronic supplementary material, figure SA).

To conclude, we recognize five orders as shown in the figure 4: Mytilida (including superfamily Mytiloidea), Ostreida (including superfamilies Pinnoidea, Pterioidea, and Ostreoidea), Pectinida (including superfamilies Pectinoidea and Anomioidea), Limida (including superfamily Limoidea) and Arcida (including superfamilies Limopsoidea and Arcoidea).

Data accessibility. All reads generated for this study are deposited online in the NCBI-SRA repository and all transcriptome accession numbers and museum ID are listed in the electronic supplementary material, table S1. Detailed information about the sampled specimens can be found on the MCZ online collections database: <http://mczbase.mcz.harvard.edu>. All in-house scripts used to construct the raw data matrices, and the raw orthogroup data matrices, are available online at https://dataverse.harvard.edu/dataverse/Pteriomorphia_phylogenomics.

Authors' contributions. S.L., G.G. and R.B. conceived the study. S.L. carried out the molecular laboratory work, analysed the data and wrote the paper. V.L.G. participated in data analysis. G.G., R.B. and V.L.G. helped draft the manuscript. All authors gave final approval for publication.

Competing interests. We have no competing interests.

Funding. This project was supported by the National Science Foundation AToL Program: Phylogeny on the half-shell—Assembling the Bivalve Tree of Life (no. DEB-0732854, 0732903, 0732860) to R.B., G.G. and Paula M. Mikkelsen.

Acknowledgements. The other members of the BivAToL team are acknowledged for discussions and support. The Harvard FAS Center for Systems Biology provided technical support. The Harvard FAS Research Computing team, especially Bob Freeman, is acknowledged for its continued support and patience with our demanding analyses. Associate editor Davide Pisani and four anonymous reviewers are acknowledged for their comments and constructive criticism. Fieldwork for this study was mostly supported by the BivAToL award and by Museum of Comparative Zoology funds.

References

- Lutz RA, Kennish MJ. 1993 Ecology of deep-sea hydrothermal vent communities: a review. *Rev. Geophys.* **31**, 211–242. (doi:10.1029/93rg01280)
- Bayne BL (ed.). 1976 *Marine mussels: their ecology and physiology*. Cambridge, UK: Cambridge University Press.
- Giribet G. 2008 Bivalvia. In *Phylogeny and evolution of the mollusca* (eds WF Ponder, DR Lindberg), pp. 105–141. Berkeley, CA: University of California Press.
- Soot-Ryen J. 1955 A report on the family Mytilidae (Pelecypoda). *Allan Hancock Pac. Expedition* **20**, 1–174.
- Carter JG *et al.* 2011 A synoptical classification of the Bivalvia (Mollusca). *Paleontol. Contrib.* **4**, 1–47.
- Distel DL. 2000 Phylogenetic relationships among Mytilidae (Bivalvia): 18S rRNA data suggest convergence in mytilid body plans. *Mol. Phylogenet. Evol.* **15**, 25–33. (doi:10.1006/mpev.1999.0733)
- Bieler R, Carter JG, Coan EV. 2010 Classification of bivalve families. *Malacologia* **52**, 113–133.
- Bouchet P, Rocroi J-P. 2010 Nomenclator of bivalve families. *Malacologia* **52**, 1–184. (doi:10.4002/040.052.0201)
- Huber M. 2010 *Compendium of bivalves*. Hackenheim, Germany: ConchBooks.
- Esteban-Delgado FJ, Harper EM, Checa AG, Rodríguez-Navarro AB. 2008 Origin and expansion of foliated microstructure in pteriomorph bivalves. *Biol. Bull.* **214**, 153–165. (doi:10.2307/25066672)
- Zhang G *et al.* 2012 The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* **490**, 49–54. (doi:10.1038/nature11413)
- Carter JG. 1990 Evolutionary significance of shell microstructure in the Palaeotaxodonta, Pteriomorphia and Isofilibranchia (Bivalvia: Mollusca). In *Skeletal biomineralization: patterns, processes and evolutionary trends*, (ed. JG Carter), vol. 1 pp. 135–296. New York, NY: Van Nostrand Reinhold.
- Carter JG, Tevesz MJS. 1978 The shell structure of *Ptychodesma* (Cyrtodontidae; Bivalvia) and its bearing on the evolution of the Pteriomorphia. *Phil. Trans. R. Soc. Lond. B* **284**, 367–374. (doi:10.1098/rstb.1978.0073)
- Waller TR. 1978 Morphology, morphoclines and a new classification of the Pteriomorphia (Mollusca: Bivalvia). *Phil. Trans. R. Soc. Lond. B* **284**, 345–365. (doi:10.1098/rstb.1978.0072)
- Cope JCW. 1996 The early evolution of the Bivalvia. In *Origin and evolutionary radiation of the Mollusca* (ed. JD Taylor), pp. 361–370. Oxford, UK: Oxford University Press.
- Waller TR. 1998 Origin of the molluscan class Bivalvia and a phylogeny of major groups. In *Bivalves: an eon of evolution—palaebiological studies honoring Norman D. Newell* (eds PA Johnston, JW Haggart), pp. 1–45. Calgary, Canada: University of Calgary Press.
- Healy JM, Keys JL, Daddow LYM. 2000 Comparative sperm ultrastructure in pteriomorph bivalves with special reference to phylogenetic and taxonomic implications. In *The evolutionary biology of the Bivalvia* (eds EM Harper, JD Taylor, JA Crame),

- pp. 169–190. London, UK: The Geological Society of London.
18. Steiner G, Hammer S. 2000 Molecular phylogeny of the Bivalvia inferred from 18S rDNA sequences with particular reference to the Pteriomorphia. In *The evolutionary biology of the Bivalvia* (eds EM Harper, JD Taylor, JA Crame), pp. 11–29. London, UK: The Geological Society of London.
 19. Matsumoto M. 2003 Phylogenetic analysis of the subclass Pteriomorphia (Bivalvia) from mtDNA COI sequences. *Mol. Phylogenet. Evol.* **27**, 429–440. (doi:10.1016/S1055-7903(03)00013-7)
 20. Malchus N. 2004 Constraints in the ligament ontogeny and evolution of pteriomorphian Bivalvia. *Palaeontology* **47**, 1539–1574. (doi:10.1111/j.0031-0239.2004.00419.x)
 21. Xue D-X, Wang H-Y, Zhang T, Zhang S-P, Xu F-S. 2012 Phylogenetic analysis of the subclass Pteriomorphia (Bivalvia) based on partial 28S rRNA sequence (in Chinese). *Oceanol. Limnol. Sin.* **43**, 348–356.
 22. Stanley SM. 1977 Trends, rates, and patterns of evolution in the Bivalvia. In *Patterns of evolution* (ed. A Hallam), pp. 209–250. Amsterdam, The Netherlands: Elsevier.
 23. Campbell DC, Hoekstra KJ, Carter JG. 1998 18S ribosomal DNA and evolutionary relationships within the Bivalvia. In *Bivalves: an eon of evolution—palaeobiological studies honoring Norman D. Newell* (eds PA Johnston, JW Haggart), pp. 75–85. Calgary, Canada: University of Calgary Press.
 24. Giribet G, Wheeler WC. 2002 On bivalve phylogeny: a high-level analysis of the Bivalvia (Mollusca) based on combined morphology and DNA sequence data. *Invertebr. Biol.* **121**, 271–324. (doi:10.1111/j.1744-7410.2002.tb00132.x)
 25. Giribet G, Distel DL. 2003 Bivalve phylogeny and molecular data. In *Molecular systematics and phylogeography of Mollusks* (eds C Lydeard, DR Lindberg), pp. 45–90. Washington, DC: Smithsonian Books.
 26. Sharma PP *et al.* 2012 Phylogenetic analysis of four nuclear protein-encoding genes largely corroborates the traditional classification of Bivalvia (Mollusca). *Mol. Phylogenet. Evol.* **65**, 64–74. (doi:10.1016/j.ympev.2012.05.025)
 27. Bieler R *et al.* 2014 Investigating the Bivalve Tree of Life: an exemplar-based approach combining molecular and novel morphological characters. *Invertebr. Syst.* **28**, 32–115. (doi:10.1071/IS13010)
 28. Lemer S, Kawauchi GY, Andrade SCS, González VL, Boyle MJ, Giribet G. 2015 Re-evaluating the phylogeny of Sipuncula through transcriptomics. *Mol. Phylogenet. Evol.* **83**, 174–183. (doi:10.1016/j.ympev.2014.10.019)
 29. Bieler R, Mikkelsen PM. 2006 Bivalvia: a look at the branches. *Zool. J. Linn. Soc.* **148**, 223–235. (doi:10.1111/j.1096-3642.2006.00255.x)
 30. Grabherr MG *et al.* 2011 Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652. (doi:10.1038/Nbt.1883)
 31. Haas BJ *et al.* 2013 De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protocols* **8**, 1494–1512. (doi:10.1038/nprot.2013.084)
 32. Altenhoff AM, Gil M, Gonnet GH, Dessimoz C. 2013 Inferring hierarchical orthologous groups from orthologous gene pairs. *PLoS ONE* **8**, e53786. (doi:10.1371/journal.pone.0053786)
 33. Altenhoff AM, Schneider A, Gonnet GH, Dessimoz C. 2011 OMA 2011: orthology inference among 1000 complete genomes. *Nucleic Acids Res.* **39**, D289–D294. (doi:10.1093/Nar/Gkq1238)
 34. Hejnal A *et al.* 2009 Assessing the root of bilaterian animals with scalable phylogenomic methods. *Proc. R. Soc. B* **276**, 4261–4270. (doi:10.1098/rspb.2009.0896)
 35. Edgar RC. 2004 MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797. (doi:10.1093/nar/gkh340)
 36. Fernández R, Hormiga G, Giribet G. 2014 Phylogenomic analysis of spiders reveals nonmonophyly of orb weavers. *Curr. Biol.* **24**, 1772–1777. (doi:10.1016/j.cub.2014.06.035)
 37. Sharma PP, Kaluziak S, Pérez-Porro AR, González VL, Hormiga G, Wheeler WC, Giribet G. 2014 Phylogenomic interrogation of Arachnida reveals systemic conflicts in phylogenetic signal. *Mol. Biol. Evol.* **31**, 2963–2984. (doi:10.1093/molbev/msu235)
 38. Andrade SCS, Novo M, Kawauchi GY, Worsaae K, Pleijel F, Giribet G, Rouse GW. 2015 Articulating ‘archannelids’: phylogenomics and annelid relationships, with emphasis on meiofaunal taxa. *Mol. Biol. Evol.* **32**, 2860–2875. (doi:10.1093/molbev/msv157)
 39. Kück P, Struck TH. 2014 BaCoCa: a heuristic software tool for the parallel assessment of sequence biases in hundreds of gene and taxon partitions. *Mol. Phylogenet. Evol.* **70**, 94–98. (doi:10.1016/j.ympev.2013.09.011)
 40. Zoller S, Schneider A. 2013 Improving phylogenetic inference with a semiempirical amino acid substitution model. *Mol. Biol. Evol.* **30**, 469–479. (doi:10.1093/molbev/mss229)
 41. Kozlov AM, Aberer AJ, Stamatakis A. 2015 ExaML version 3: a tool for phylogenomic analyses on supercomputers. *Bioinformatics* **31**, 2577–2579. (doi:10.1093/bioinformatics/btv184)
 42. Aberer AJ, Kobert K, Stamatakis A. 2014 ExaBayes: massively parallel Bayesian tree inference for the whole-genome era. *Mol. Biol. Evol.* **31**, 2553–2556. (doi:10.1093/molbev/msu236)
 43. 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. (doi:10.1093/Sysbio/Syt022)
 44. 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. (doi:10.1093/molbev/msh112)
 45. Berger SA, Krompass D, Stamatakis A. 2011 Performance, accuracy, and Web server for evolutionary placement of short sequence reads under maximum likelihood. *Syst. Biol.* **60**, 291–302. (doi:10.1093/sysbio/syr010)
 46. Grünewald S, Spillner A, Bastkowski S, Bogershausen A, Moulton V. 2013 SuperQ: computing supernetworks from quartets. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **10**, 151–160. (doi:10.1109/TCBB.2013.8)
 47. Newell ND. 1965 Classification of the Bivalvia. *Am. Mus. Novit.* **2206**, 1–25.
 48. Lemmon AR, Brown JM, Stanger-Hall K, Lemmon EM. 2009 The effect of ambiguous data on phylogenetic estimates obtained by maximum likelihood and Bayesian inference. *Syst. Biol.* **58**, 130–145. (doi:10.1093/Sysbio/Syp017)
 49. Dell’Ampio E *et al.* 2014 Decisive data sets in phylogenomics: lessons from studies on the phylogenetic relationships of primarily wingless insects. *Mol. Biol. Evol.* **31**, 239–249. (doi:10.1093/molbev/mst196)
 50. Fernández R, Edgecombe GD, Giribet G. In press. Exploring phylogenetic relationships within Myriapoda and the effects of matrix composition and occupancy on phylogenomic reconstruction. *Syst. Biol.* (doi:10.1093/sysbio/syw041)
 51. Morton B. 1996 The evolutionary history of the Bivalvia. In *Origin and evolutionary radiation of the Mollusca* (ed. JD Taylor), pp. 337–359. Oxford, UK: Oxford University Press.
 52. Taylor JD. 1973 The structural evolution of the bivalve shell. *Palaeontology* **16**, 519–534.
 53. Cope JCW. 1997 The early phylogeny of the class Bivalvia. *Palaeontology* **40**, 713–746.
 54. Taylor JD, Reid DG. 1990 Shell microstructure and mineralogy of the Littorinidae; ecological and evolutionary significance. *Hydrobiologia* **193**, 199–215. (doi:10.1007/BF00028077)
 55. Harper EM. 2000 Are calcitic layers an effective adaptation against shell dissolution in the Bivalvia? *J. Zool.* **251**, 179–186. (doi:10.1111/j.1469-7998.2000.tb00602.x)
 56. Taylor JD, Layman M. 1972 The mechanical properties of bivalve (Mollusca) shell structures. *Palaeontology* **15**, 73–87.
 57. Stanley SM. 1970 Relation of shell form to life habits in the Bivalvia (Mollusca). *Geol. Soc. Am. Mem.* **125**, 1–182. (doi:10.1130/MEM125-p1)
 58. Taylor JD, Kennedy WJ, Hall A. 1969 The shell structure and mineralogy of the Bivalvia. Introduction: Nuculacea-Trigonacea. *Bull. Br. Mus. Nat. Hist. Zool.* 1–125.
 59. Furuhashi T, Schwarzingler C, Miksik I, Smrz M, Beran A. 2009 Molluscan shell evolution with review of shell calcification hypothesis. *Comp. Biochem. Physiol. B* **154**, 351–371. (doi:10.1016/j.cbpb.2009.07.011)
 60. Owen G. 1978 Classification and the bivalve gill. *Phil. Trans. R. Soc. Lond. B* **284**, 377–385. (doi:10.1098/rstb.1978.0075)