Curious bivalves: systematic utility and unusual properties of anomalodesmatan mitochondrial genomes

S. T. Williams[a]*, P. G. Foster[a], C. Hughes[a], E. M. Harper[b], J. D. Taylor[a], D. T. J. Littlewood[a], P. Dyal[a], K. P. Hopkins[a,c] & A. G. Briscoe[a]

[a] Natural History Museum, Cromwell Rd, London SW7 5BD, United Kingdom

[b] Department of Earth Sciences, University of Cambridge, Downing Street, Cambridge, CB2 3EQ, United Kingdom

[c] Present address: Institute of Zoology, Zoological Society of London, Regent's Park, London, NW1 4RY, United Kingdom

*Corresponding author: S. T. Williams, email: s.williams@nhm.ac.uk

Highlights

- The addition of seven anomalodesmatans improves relationships in bivalve mitogenomic tree
- Non-monophyly of Bivalvia may in part be the result of compositional heterogeneity
- Transcriptomic data helped determine boundaries for protein-coding genes
- We identified a possible case of gene duplication of ND5 in *Myadora brevis*
- We identified possible pseudogenes in published anomalodesmatan transcriptomic data

Abstract

Mitogenomic trees for Bivalvia have proved problematic in the past, but several highly divergent lineages were missing from these analyses and increased representation of these groups may yet improve resolution. Here, we add seven new sequences from the Anomalodesmata and one unidentified semelid species (*Bryopa lata, Euciroa cf. queenslandica, Laternula elliptica, Laternula truncata, Lyonsia norwegica, Myadora brevis, Tropidomya abbreviata, "Abra"* sp.). We show that relationships in a mitogenomic tree for the Class are improved by the addition of seven anomalodesmatans from this highly divergent clade, but are still not completely consistent with relationships recovered in studies of nuclear genes. We suggest that some anomalous relationships (for instance the non-monophyly of Bivalvia) may be partially explained by compositional heterogeneity in the mitogenome and suggest that the addition of more taxa may help resolve both this effect and possible instances of long branch attraction. We also identify several curious features about anomalodesmatan mitogenomes. For example, many protein-coding gene boundaries are poorly defined in marine bivalves, but particularly so in anomalodesmatans, primarily due to non-conserved boundary sequences. The use of transcriptomic and genomic data together enabled better definition of gene boundaries, the identification of possible pseudogenes and suggests that most genes are translated monocistronically, which contrasts with many other studies. We also identified a possible case of gene duplication of ND5 in *Myadora brevis* (Myochamidae). Mitogenome size in the Anomalodesmata ranges from very small compact molecules, with the smallest for *Laternula elliptica* (Laternulidae) only 14,622 bp, to *Bryopa lata* (Clavagellidae) which is at least 31,969 bp long and may be > 40,000bp. Finally, sampled species show a high degree of sequence divergence and variable gene order, although intraspecific variation in *Laternula elliptica* is very low.

Keywords: mitochondrial genome, gene boundaries, pseudogenes, bivalve, enrichment, systematics

1. Introduction

The use of mitochondrial genomes (mitogenomes) in phylogenetics has had a varied history, depending on the organismal groups studied and the level of taxonomic resolution required. Offering complexity and phylogenetic signal across multiple levels (nucleotide, amino acid, gene order, RNA secondary structure), these ubiquitous, usually circular and high copy number molecules have been difficult to characterize with Sanger sequencing, but are rich in characters once secured. With limited intergenic spaces, usually with one large non-coding region and most often inherited asexually via the maternal line, a mitogenome can be considered as a single locus; typical mtDNAs encode 13 protein-coding, 2 ribosomal RNA and 22 tRNA genes. Unusual or conflicting phylogenetic resolution using mitogenomes, when compared to single or multiple nuclear gene phylogenies, combined with the cost of sequencing, has meant uneven taxon sampling across the Metazoa. However, denser and strategic taxonomic sampling of mitogenomes, recently facilitated by next generation sequencing methods for mitogenome characterization (Smith 2016), has led to improved substitution models and a better understanding of mtDNA evolution (Ballard & Whitlock 2004). Together this has informed more robust phylogenetic analyses, which in turn has led to considering compelling evolutionary scenarios and testable hypotheses concerning interrelationships. Mitogenomes also offer some advantages over nuclear genes. For example, their high copy number makes them an ideal target for studies using ancient or degraded DNA samples (e.g. Dabney *et al.* 2013; Krause *et al.* 2010). Complete mitogenomes (or sequences from all transcribed genes) can also be obtained at no extra cost as the by product of NGS studies that may initially have another focus (e.g. Uribe *et al.* 2016a). As such, it is worth investigating the utility of the mt genome further as a phylogenetic marker.

Studies have shown that mitogenomes are capable of resolving robust phylogenies that are informative at both deep and shallow nodes in Mollusca. They have proved particularly useful for resolving cephalopod and gastropod relationships (e.g. Allcock *et al.* 2011; Uribe *et al.* 2016a; Williams *et al.* 2014). However, mitogenomes have proved less reliable in unravelling relationships within the Bivalvia. Recent phylogenetic analyses of bivalves using partial mitogenomes (Plazzi *et al.* 2011; Plazzi *et al.* 2010) resulted in well-supported phylogenies, but many nodes conflicted with previous phylogenies based both on nuclear genetic markers (e.g. Sharma *et al.* 2012; Taylor *et al.* 2007) and traditional morphologically-based concepts (Giribet & Wheeler 2002). This incongruence suggested partial mitogenomes alone may not be useful for resolving bivalve relationships (Sharma *et al.* 2012). A later study using 12 protein-coding mitochondrial genes (effectively complete mitogenomes) from 18 bivalve taxa (Yuan *et al.* 2012) also shows some differences from nuclear gene trees. Expanded strategic taxon sampling has seen improvements of phylogenies within some

molluscan subgroups (e.g. cephalopods, Allcock *et al.* 2011) and may yet help to resolve relationships within Bivalvia.

Taxon sampling across the bivalve clade has not been uniform in previous mitogenome studies, and anomalodesmatan bivalves have been poorly represented (Plazzi *et al.* 2011; Plazzi & Passamonti 2010; Yuan *et al.* 2012). This is acknowledged by Plazzi and colleagues, who suggest that poor sampling of anomalodesmatans may be responsible for borderline measures of Average Taxonomic Distinctness and high variance in Taxonomic Distinctness of samples (Plazzi *et al.* 2011). At family level the anomalodesmatans account for approximately one sixth of extant bivalve diversity. This clade of early divergent 'heterodonts' has a long evolutionary history and today is represented by a wide variety of highly specialised and disparate species (Harper *et al.* 2006; Morton 1985). It includes some of the most specialised bivalves, including the tube-dwelling clavagelloids and an array of carnivorous taxa, known as the septibranchs, which have radiated into the deep sea where they account for half the diversity of Mollusca (Knudsen 1970). It is, at least in part, because of this degree of specialisation into unique and difficult to sample habitats that they have been under-represented in phylogenetic analyses to date. Here we address the taxonomic imbalance through improved sampling.

Previously, one of the greatest obstacles to the use of mitochondrial genes in bivalve phylogenies has been that 'universal' mitochondrial primers do not work routinely across all taxa because the group is so divergent. We overcome this technical difficulty by using shotgun sequencing to obtain full mitogenomes for seven anomalodesmatans and one semelid species. One of these sequences was obtained using a new method for enrichment of mitochondrial DNA combined with shotgun sequencing which resulted in increased coverage. We use these data to determine whether the addition of anomalodesmatan bivalves helps resolve a bivalve phylogeny. We also discuss features of the mitochondrial genome focussing on those that are unusual or found predominantly in marine bivalves.

2. Materials & methods

2.1 Sample collection and DNA extraction

We sequenced the mitogenomes of seven anomalodesmatan species plus one semelid species, which had been originally identified erroneously as an anomalodesmatan. The taxa chosen provided good coverage of the diversity of Anomalodesmata (6 families) and included two carnivorous septibranch taxa (*Tropidomya abbreviata* and *Euciroa cf. queenslandica;* Table 1). Specimens were collected from a range of localities and depths and stored in 100% ethanol (see Table 1). A small (5mm^3) piece of tissue was removed from each sample for molecular analyses; the remainder of each specimen was accessioned in NHM (London) or MNHN (Paris) collections (Table 1). Total genomic DNA (gDNA) was extracted using the DNeasy Blood & Tissue Kit (Qiagen) following manufacturer's recommendations. DNA was quantified using a Qubit fluorometer (Invitrogen). A mitogenome for one of our chosen species, *Laternula elliptica*, was published by another lab after this project had begun (NC_022846, Park & Ahn 2015) (NC_022846, Park & Ahn 2015)and was used for comparison.

2.2 Mitochondrial DNA enrichment

In order to increase the proportion of mitochondrial DNA within each genomic extract, which in turn increases the number of samples that can be multiplexed during sequencing, thus reducing costs, a hybridisation enrichment method was trialled. The entire mitochondrial sequence of *Laternula elliptica* was obtained by shotgun sequencing following the methods of Williams et al. (2014). Enrichment of two anomalodesmatan mitogenomes (*L. rostrata* and *Bryopa lata*) was attempted using the entire *Laternula elliptica* mitogenome in two overlapping fragments as bait.

Two primer pairs were designed using Primer3 with BLAST and global alignment algorithm to screen primers against the Mollusca database in order to avoid non-specific amplifications (Ye *et al.* 2012) (SW_PCR2_int_F4, CTTCTTCTTACTATTAACGCTC, SW_LongPCR_6R_COI, AATCATACTTAGGGAAGTGCCCATT; SW_LongPCR_3R_NAD2, ACCCCAAATCAATGGGTACTTGATA, SW_LongPCR_3F_COI, TGAATACCCTTTCACAGAAATTGGC). The first primer pair amplified a region 5,836 bp long and the second primer pair amplified a region 9,316 bp long. The two fragments were produced by long range PCR using a TaKaRa LA PCR kit (Takara Bio Inc.) with 2.5 u of Taq, 0.4 uM of each primer, and cycling conditions: 94°C for 2 mins, followed by 30 cycles of 94°C for 20 s, 60°C for 30 s, and 68°C for 9.5 mins, followed by 10 mins at 68°C.

153 The two PCR fragments were used as molecular probes for hybridisation capture following
154 Maricic (2010) with minor but essential modifications. In order to account for an expected greater
155 pairwise difference between the target and the probe sequence than has previously been shown to
156 be successful, the annealing temperature during the hybridization step was reduced every 12 hours
157 by 5°C (from 60-45°C over 48 hours). To elute the enriched library the supernatant was removed and
158 discarded, without conducting any washes and the enriched library was eluted and cleaned following
159 the standard protocol.

160

161 **2.3 Mitochondrial genome sequencing and assembly**

162 Indexed libraries were prepared for sequencing using a TruSeq Nano preparation kit
163 (Illumina) incorporating the standard 8-cycle enrichment step. Due to degradation of gDNA, three
164 specimens (*Tropidomya abbreviata*, *Lyonsia norwegica*, and *Euciroa cf queenslandica*) were
165 subsequently amplified for 15 cycles in order to meet the concentration requirements of the library.
166 All eight samples were sequenced on 1/5 of a flowcell on an Illumina MiSeq platform (v.2 chemistry;
167 2x250 paired-end). In addition, for specimens used in enrichment studies, unenriched indexed gDNA
168 libraries were also sequenced in order to provide baseline coverage of the mitogenome from which
169 success of the enrichment process could be judged (v.2 chemistry; 2x300 paired-end).

170 Following on-instrument de-multiplexing, raw sequencing reads were analysed and
171 assembled using Geneious v.6.1.7. (https://www.geneious.com). The data were trimmed allowing
172 no ambiguous base calls and removing bases from the terminal ends of reads with an error
173 probability of 0.05 or higher (i.e. those with a greater than 5% chance of being incorrect). The
174 trimmed reads were first assembled *de novo* and the resulting contigs were interrogated against the
175 NCBI database, using BLAST (Altschul *et al.* 1990), in order to identify sequences of mitochondrial
176 origin. Unassembled reads were then iteratively mapped and reassembled to the mitochondrial
177 sequences until the resulting contig could be circularised. A published 28S sequence from *Abra alba*
178 (KF741656) was also used as a reference sequence to identify 28S sequence from NGS raw reads for
179 the semelid species (GenBank accession KY420905).

180

181 **2.4 Mitochondrial genome annotation**

182 To verify the gene boundaries of our *Laternula elliptica* mitogenome, transcriptomic data
183 was obtained from the NCBI Sequence Read Archive (SRA011054; (Clark *et al.* 2010) and assembled
184 to the individual ribosomal and protein-coding genes (PCGs). Gene boundaries, for the remaining

bivalves sequenced in this study, were putatively identified using MITOS (Bernt *et al.* 2013) and verified by visualization of open reading frames and comparison to alignments of molluscan mitochondrial genes including the corrected *Laternula elliptica* sequences.

Transfer RNA sequences (tRNAs) were identified using MITOS, DOGMA (Wyman *et al.* 2004) and ARWEN v. 1.2. (Laslett & Canbäck 2008) and boundaries were determined by comparison of secondary structures. In cases where multiple locations were suggested for a tRNA, the options with the highest likelihood score were used except where these were considered invalid because of a large overlap with other existing tRNAs exhibiting a higher score. Secondary structures of tRNAs were drawn using mt-tRNA-Draw (Youngblood and Masta, unpublished).

2.5 Phylogenetic analyses

The eight new mitogenomes were included in phylogenetic analyses along with 98 bivalve mitogenomes downloaded from RefSeq representing all species sequenced (only one individual per species; June 2015) along with additional mitogenomes for *Argopecten pupuratus* **(**KF601246), *Macoma balthica* (KM373200) and *Nucula nucleus* (EF211991). A total of 87 gastropod sequences were downloaded from RefSeq along with *Lunella aff cinerea* KF700096). After preliminary analyses two bivalve genomes (NC_013659 *Venustaconcha ellipsiformis* and NC_015477 *Utterbackia peninsularis*, both freshwater palaeoheterodont taxa) were excluded as both were highly divergent from other genomes and both species are known to demonstrate doubly uniparental inheritance (Breton *et al.* 2011b). The sequence for the gastropod *Lottia digitalis* NC_007782 was also excluded because it was highly divergent, and its placement was not relevant to the aims of this study. A final total of published sequences for 87 gastropods and 99 bivalve taxa (+ eight new) was used in phylogenetic analyses.

Concatenated amino-acid sequences for all protein-coding genes were used in Bayesian analyses implemented in PhyloBayes (Lartillot *et al.* 2009). Preliminary analyses showed that protein sequences were highly diverged within Bivalvia, and use of DNA sequences in phylogenetic analyses was judged inappropriate given likely saturation. Complete genes were translated using NCBI translation table 5 for invertebrate mitochondria. Amino-acid sequences were aligned with Clustal Omega v 1.1.0 (Sievers *et al.* 2011). Gblocks v 0.91b (Casastrena 2000) was used to identify conserved sites in the alignment, using default settings except that parameter b5 was set to h (alignment positions with half gaps allowed) (Table 2). Of the 99 bivalve mitogenomes from RefSeq, 51 had no ATP8 annotated, and one (NC_009081, pectinid *Mizuhopecten yessoensis*) had no COX2. Since so few species had ATP8 sequence, it was excluded from phylogenetic analyses. One bivalve

218 species (NC_014590, *Musculista senhousia*) had two copies of COX2; only the first one, the longest,
219 was used in analyses. ND4L was also excluded from analyses because most sites were removed by
220 Gblocks (Table 2).

221 The results of phylogenetic analyses of two datasets are reported here. The first dataset
222 included 87 gastropods and 107 bivalves (gastropods+bivalves). The final alignment was 1931 amino
223 acids in length. In a second dataset we excluded all gastropods and two bivalves (*Solemya velum*
224 NC017612 and *Nucula nucleus* EF211991) and re-ran phylogenetic analyses including the remaining
225 bivalves (Autobranchia). The final alignment for the Autobranchia tree was 1918 amino acids in
226 length and included 105 taxa. Phylogenetic analyses were undertaken using PhyloBayes (v. 4.1c
227 Lartillot *et al.* 2009) with the CAT-Poisson model for the Autobranchia dataset and CAT+GTR for the
228 gastropods+bivalves dataset, with the analysis run four times for the first dataset and five times for
229 the second. A consensus tree was obtained from trees obtained from all runs. The
230 gastropods+bivalves tree was drawn using the gastropods as outgroup to bivalves and the
231 Pteriomorphia (36 species) were used as the outgroup to the remaining bivalves in the Autobranchia
232 tree. The choice of outgroups was based on two recent phylogenomic studies (Kocot *et al.* 2011;
233 Smith *et al.* 2011).

234 Compositional heterogeneity occurs when character state frequencies change across the
235 tree (Foster 2004). Effects of heterogeneity were shown using a composition tree. For this, distance
236 matrices were made based only on Euclidean distances between compositions of each sequence
237 (equation 4 in Lockhart *et al.* 1994), then using BIONJ (Gascuel 1997) to tree those distances, and
238 with support for nodes based on 200 bootstraps.

239

240 **2.6 Identification of semelid sp.**

241 The highest BLAST hits to the new 28S and COI sequences for the small, unidentified semelid
242 were to *Macoma, Abra*, *Soletellina*, *Gari* and *Scrobicularia* species (E value = 0; identity >
243 95%). Although the semelid specimen was damaged and could not be identified to species
244 morphologically, the presence of an internal ligament is consistent with placement in the genus *Abra*
245 or *Scrobicularia* (or any semelid), but as *Scrobicularia* does not occur in the Indo-West Pacific we
246 tentatively refer to it here as "*Abra*" sp. Another possible generic placement is *Leptomya*, but there
247 are no published sequences for this genus.

248

3. Results

3.1 General features of the mitochondrial genome

Complete mitogenomes were sequenced for seven of the eight bivalves studied (all new GenBank accession numbers in Table 3). Only the mitogenome for *Bryopa lata* was incomplete because it could not be circularised unambiguously, although every gene was found within the 31,969 bp sequence. Of the genomes studied the smallest genome was 14,622bp (in *Laternula elliptica*), with very short regions of non-coding sequence (longest non-coding region 78 bp; Table 3). Despite being incomplete, the *B. lata* genome was the largest of the seven species, at 31,969 bp with three large regions of non-coding sequence (2,779, 3,817, 7,025bp) (Table 3). High coverage of a repetitive region at the 3' end of the linear fragment that was recovered (approximately eight times that seen across the rest of the genome), would suggest the complete *B. lata* mitogenome may be >40,000 bp in length. It is this highly repetitive region that prevented circularising and completing the mitogenome with confidence.

3.2. Gene order and content

Sequences corresponding to all 13 metazoan protein-coding genes, including ATP8, and two ribosomal (rRNA) genes were identified for all taxa. A possible instance of gene duplication was identified for ND5 in *Myadora brevis*.

All protein-coding, rRNA and tRNA genes were transcribed on the forward strand in the same orientation. Gene order, both of protein-coding genes and considering all genes, differed significantly among taxa in this study as indicated by whole genome alignments (Appendix A Supp. Data 2). However, gene order was identical for *L. elliptica* with a genome for the same species published on GenBank (NC_022846 Park & Ahn 2015). *Bryopa lata*, *L. elliptica* and *L. rostrata* share the same gene order for protein-coding genes. Their gene order differs from *Ly. norwegica* only in the position of ND1 and ND2, which are moved to positions 12 and 13 respectively in rank order (Table 3).

3.3. Gene boundaries

All protein-coding gene boundaries for *L. elliptica* were confirmed by comparison with transcriptomic data, with the exception of the 3' end of ATP8 and the 5' end of COX3. The annotation of gene boundaries in our sequence and that in GenBank (NC_022846) differ by a several base pairs for both rRNA genes (≤3 bp), several tRNAs (≤7 bp) and three protein-coding genes (5' ND4 – 21 bp, 5' CYTB – 6 bp and 5' COX3 – 9 bp). The extra 21 bp at the 5' end of ND4 in the published genomic

sequence included a stop codon, but was translated in both our sequence and the transcriptomic sequence.

Alignments of protein coding genes showed that for many genes (e.g. COX1, ND3) there were no clearly conserved start or stop positions among the Anomalodesmata, or indeed among most marine bivalves. In such cases the beginning and end of each gene were determined by finding the closest possible initiation and termination codons to boundaries determined by reference to an alignment of all bivalves, excluding freshwater bivalves and pteriomorphs, as these taxa often started or stopped on a different residue. A truncated T stop codon was used if no other stop codon was present or to avoid an overlap of several amino acid residues with another protein-coding gene.

Initial examination of transcriptomic data for *L. elliptica* suggested that there were two polycistronic transcripts, however, more careful examination of sequence data determined that there were small regions in reads occurring immediately before and after each gene that did not match when aligned to the complete genome. These sequences correspond to Illumina adaptors at the 5' end and a poly-A tail at the 3' end, suggesting that translation is monocistronic in this species with the exception of ATP8 and COX3, which appear to be translated together suggesting that there may be post-translational modification of these proteins. We also found an overlap of several base pairs between putative ATP8 and COX3 genes in *Ly. norwegica*, *L. elliptica* and *L. rostrata*; and between ATP8 and ND5 in *Ly. norwegica*. The fact that assembly of transcriptome reads produced apparently polycistronic products raises the possibility that reports of polycistronic translation may be exaggerated.

ND4 and ND4L overlapped by a few base pairs in *Euciroa queenslandica* and *Ly. norwegica*. More unusually in our initial annotation there was a single base overlap for ND3/ND4 in *B. lata* and for ND1/ND6 in *L. rostrata*. In both cases the overlap is the last base of a TAG stop codon. It is possible that these genes stop on T or TA, avoiding an overlap. Transcriptomic data suggest that all genes are transcribed separately (except ATP8 and COX3), which supports the idea of the shortened stop codon, however as no TA stop codons were identified for other genes or taxa in our study, we have conservatively retained the full stop codon in our annotation.

Protein-coding genes varied in length by 6-48 amino acid residues, with the greatest length variation seen in COX1 and COX2 (Table 3). More individual genes were longest in the semelid "*Abra"* sp.. Excluding this outgroup from the analysis reduced the length variation seen among protein-coding genes.

3.4. Initiation and termination of translation:

The most common initiation codon was ATG (44%; Tables 4 & 5), the least common initiation codon was ATC (<1%). An unusual start codon GTT was observed at relatively high frequency (9%), but did not occur in the transcriptomic data. This start codon has also been recorded in nematodes (Okimoto & Wolstenholme 1990), but not previously in bivalves to our knowledge. In every case where we have used the GTT start codon, the identification of the starting position was difficult to determine, so our choice must be considered tentative at this time.

Only three stop codons were identified: TAA (48%), TAG (38%) and a truncated stop codon where a single T is completed by the addition of two 3' A residues to the mRNA to produce a TAA stop codon (14%) (Tables 4 and 5). All three codons were found in the transcriptomic data for *L. elliptica*.

3.5. Pseudogenes

In addition to confirming the gene boundaries within the mitogenome of *Laternula elliptica*, a number of reads were found in the published transcriptome that closely matched the 5' end of the COI sequence but ended prematurely due to the presence of stop codons. This would suggest that pseudogenes were being translated and subsequently sequenced.

3.6. tRNAs

Sequences corresponding to 22 tRNAs were found for all species. Secondary structures could be determined for most putative tRNA genes, with the exception of some serine tRNAs (Appendix A Supp. Data 1). As in other bivalves some non-canonical conformations were observed including mismatches in base pairing in the aminoacyl acceptor arm and the anticodon stem.

3.7. Intraspecific divergence

With our new mitogenome sequence, three datasets were available for *L. elliptica*. Comparison of published transcriptomic sequences with our shotgun sequence for *L. elliptica* identified only one base pair difference. On the other hand, our sequence differs from the GenBank genomic DNA sequence by 9 bp (one each in COX1, rRNA-L, tRNA-Asp, ND4, ND3, ND2, ND5, overlap between ATP8 and COX3, and COX3).

3.8. Enrichment

Approximately four times enrichment was observed in the only congeneric tested, *Laternula rostrata,* with 0.38% of mitochondrial reads compared to 0.079% in an un-enriched library. There was no enrichment in test species from other genera.

3.9. Phylogenetic analyses

The runs did not converge in the Bayesian analysis of the gastropods+bivalves dataset. The topology of consensus trees of each of the five runs was not the same, but since all the differences were within the gastropod clade a consensus tree was calculated based on all trees (Fig. 1). The runs for the Autobranchia dataset converged on the same tree with a maxdiff among independent runs of 0.13 (Fig. 2).

In the gastropods+bivalves tree two protobranch bivalves (*Solemya velum* and *Nucula nucleus*) fall out in in the gastropod clade. One explanation for this finding may be that compositional effects are driving the apparent relationship between Protobranchia and the gastropods. A Pearson's chi-squared test on the gastropods+bivalves dataset using P4 (Foster 2004) was significant for compositional heterogeneity (p=0.008). To determine whether this created any artefactual phylogenetic relationships we built a tree based only on Euclidean distances between compositions (Appendix A Supp. Data 3). In this sort of tree, closely related taxa are expected to have similar compositions, and so form clades. However, deeper nodes should have low support and relationships should not reflect systematic relationships if there is no compositional heterogeneity. Anomalous relationships that have good support may suggest phylogenetic effects due to their compositions. In our tree (Appendix A Supp. Data 3) clades were observed among major groups that were expected (e.g. Ostreidae, Pectinoida and Unionida). However we also observed that the two protobranch bivalves clustered with the gastropods, suggesting a weak compositional effect.

4. Discussion

The explosive numbers of new mitogenomes appearing on public databases has been received with mixed reactions – are new mitogenomes more of the same or a useful addition? The relative ease with which mitogenomes can be fully characterised accurately, rapidly and cost-effectively using NGS methodologies has allowed for more sampling of taxa for mitogenomic-based phylogenetics. However, it is sampling strategy in the context of available mitogenomes that determines utility and nodal resolution in phylogenetics. As an example, consider the evolution of mitogenomic studies in the context of wider phylogenomics in resolving annelid phylogenies; mitogenomic studies have contributed signal iteratively towards greater resolution until a threshold of taxonomic diversity partnered with strategic sampling reveals the endeavour to be worthwhile (Bleidorn *et al.* 2009; Weigert & Bleidorn 2016; Weigert *et al.* 2016). Similar successes can be tracked for decapod crustaceans (Shen *et al.* 2013). In the current study, considering bivalve molluscs, a focus on using 11 mt genes and additional anomalodesmatan taxa has provided an iterative step towards greater phylogenetic resolution (e.g. restoring Anomalodesmata to the Euheterodonta, unlike some earlier studies relying only on mt genes e.g. Plazzi *et al.* 2011; Plazzi & Passamonti 2010) and suggests continued investment and focus still have more to offer.

4.1.1 Systematic relationships within Bivalvia

When gastropods are included in the phylogenetic analyses monophyly of the Bivalvia is not supported in the mitogenome tree. While the Autobranchia are recovered as a clade, the protobranch bivalve genera *Solemya* and *Nucula* form a clade sister to non-Euthyneuran gastropods rather than with the rest of the bivalves (Fig. 1). Although some molecular analyses also failed to establish bivalve monophyly (Combosch *et al.* 2016; Giribet & Distel 2003; Plazzi *et al.* 2013), recent analyses using multiple genes (Bieler *et al.* 2014), nuclear coding genes (Sharma *et al.* 2013) or transcriptomic RNA-seq (Gonzalez *et al.* 2015) with taxa from all molluscan classes as outgroups have recovered monophyly of the bivalves with the protobranch clade sister to all other bivalves. High support in the mitogenome tree for the unexpectedly close relationship between protobranch bivalves and gastropods may be explained in part by compositional heterogeneity biasing the analyses (e.g. Rota-Stabelli *et al.* 2013). Alternatively, this relationship may reflect some genuine shared relationship between protobranchs and gastropods. Both protobranch genera sampled have an extended fossil record originating in the early Ordovician and have some morphological features not present in other bivalves but similar to basal gastropod features, including the large

hypobranchial gland of *Solemya* species (Taylor *et al.* 2008) and the leaflet ctenidia not used for filter feeding; in *Solemya* leaflet ctenidia have been modified to house chemosymbiotic bacteria.

In the Autobranchia tree we recover Pteriomorphia and Heteroconchia with moderate support (PP=0.89; Fig. 2), although Heteroconchia is not recovered in the bivalves+gastropods tree, as Pteriomorphia is sister to Euheterodonta (Fig. 1). Paleoheterodonta and Euheterodonta are both recovered in all analyses, although with high support only in the gastropods+bivalves tree (PP≥0.98; Fig. 1). Superfamilies are well supported within the Autobranchia bivalve tree (Fig. 2) but the relationships among superfamilies are less clear. Trees from prior molecular analyses also show varying topologies for the relative positions of the different major clades (e.g. Bieler *et al.* 2014; Combosch & Giribet 2016; Giribet & Distel 2003; Giribet & Wheeler 2002; Sharma *et al.* 2012; Steiner & Hammer 2000).

A recent transcriptomics study of the Pteriomorpha establishes monophyly of the group and recovers Mytilida as sister to Ostreida, with Arcida as sister to all other pteriomorph taxa (Lemer *et al.* 2016). Our tree agrees with this study in that it finds a well-supported pteriomorph clade and clades for super families were also recovered. However, relationships among superfamilies differ markedly. Mytilida is sister to all pteriomorphs with Arcida then sister to Pectinoidea+Pterioidea. Pinnoidea, represented in this study by *Atrina pectinata*, clusters in a well-supported clade with Pterioidea and Ostreoidea. This topology is supported by some previous analyses (e.g. Giribet & Distel, 2003; 5 gene analysis of Bieler et al. 2014) but not others (e.g. Sharma et al 2012; nine gene analysis of Bieler et al. 2014) where *Pinna carnea* groups with Mytilidae. New genomic analysis (Lemer *et al.* 2016) places Pinnoidea as sister to Pterioidea+Ostreoidea. Shell microstructure would suggest a closer relationship of Pinnoidea with Pterioidea than with Mytiloidea (compare Checa *et al.* 2014; Checa *et al.* 2005).

The Heterodonta are highly divergent with the Palaeoheterodonta distantly separated from the Euheterodonta. Although there are a number of mitogenomes published for Palaeoheterodonta species (Unionoidea + Trigonoidea) it is disappointing that there are none published for any species from the Archiheterodonta (Crassatellidae, Carditidae, Astartidae, Condylocardiidae). This is unfortunate because in recent molecular analyses the palaeoheterodonts and archiheterodonts form early divergent branches of the heteroconch bivalves but their relative positions appear unresolved; in the nine gene analysis of Bieler et al. (2014) Archiheterodonta form the basal branch but transcriptome data (Gonzalez *et al.* 2015) indicates that the Palaeoheterodonta form the basal split. Surprisingly, the mitochondrial gene trees of Plazzi et al. (2011) show *Cardita* and *Astarte* grouping within the Pteriomorpha, a finding not supported by other studies.

Within the Euheterodonta, the Anomalodesmata form a sister clade to all the other euheterodonts confirming a topology recovered in molecular analyses based on nuclear genes (e.g. Bieler *et al.* 2014; Gonzalez *et al.* 2015; Sharma *et al.* 2012; Taylor *et al.* 2007), but differing from earlier studies based only on four mitochondrial genes (Plazzi *et al.* 2011; Plazzi & Passamonti 2010).

Within the Imparidentia (non-anomalodesmatan euheterodonts) major superfamilies are well differentiated and well supported but data for many families are lacking. The grouping of the two superfamilies Solenoidea and Hiatelloidea (Adapedonta) has been recognised in previous molecular analyses (Bieler *et al.* 2014; Giribet & Wheeler 2002; Taylor *et al.* 2007). This contradicted previous assertions based on morphology that the Hiatelloidea were members of the Myoida. To date there are no strong morphological apomorphies that unite the two clades. In the present mitogenome tree, *Hiatella* is separate on a long branch from the two *Panopea* species although in other published gene trees it groups with *Panopea* (Bieler *et al.* 2014; Taylor *et al.* 2007) as a monophyletic Hiatelloidea.

The Tellinoidea and Cardioidea grouping in the Autobranchia tree (Fig. 2), although not well supported, has been recovered in other molecular phylogenies (Bieler *et al.* 2014; Taylor *et al.* 2007). As yet there are no obvious shared morphological apomorphic characters.

The remainder of the Imparidentia heterodonts – superfamilies Myoidea, Mactroidea, Arcticoidea and Veneroidea form a well supported clade and this grouping with other included families, named Neoheterodontei by Taylor et al. (2007), has been recovered in other molecular analyses (Bieler *et al.* 2014; Combosch *et al.* 2016; Giribet & Distel 2003; Giribet & Wheeler 2002).

4.1.2 Systematics of Anomalodesmata

The coral boring clavagelloidean *Bryopa lata* is included in analyses for the first time and groups in a well-supported subclade with *Lyonsia norwegica*. Previous 18S RNA analyses also showed a relationship between the clavagelloideans (*Clavagella*, *Brechites* and *Penicillus*) and pandoroidean, *Lyonsia norwegica* (Harper et al., 2006). Although it is tempting to consider the carnivorous septibranchs as a single clade, the two morphologically and taxonomically distinct septibranch taxa (*Tropidomya* and *Euciroa*) do not cluster together in our analysis, as has proved the case in previous studies with different taxa (Bieler *et al.* 2014; Harper *et al.* 2006).

4.1.3 Systematic relationships within Gastropoda

Although not the focus of this study, a large number of gastropods were included in phylogenetic analyses (Fig. 1). As discussed above, Gastropoda is made non-monophyletic by the inclusion of protobranch bivalves as sister to Vetigastropoda and Caenogastropoda. The major clades Vetigastropoda, Caenogastropoda and Euthyneura are all recovered with strong support (PP=1). Vetigastropoda is sister to Caenogastropoda with high support (PP=1). Sampling within this clade is limited, but the sister relationship between Fissurellidae and Haliotidae is not supported by recent mitogenomic studies focussing on this group (Lee *et al.* 2016; Uribe *et al.* 2016a; Uribe *et al.* 2016b; Wort *et al.* 2016). Of the three caenogastropod orders, only Architaenioglossa is monophyletic; the monophyly of this group is consistent with morphological studies (Strong 2003). As in other molecular studies, Littorinimorpha and Neogastropoda are non-monophyletic (e.g. Cunha *et al.* 2009; Osca *et al.* 2015; Williams *et al.* 2014). The Euthyneura clade unexpectedly is more like previous molecular phylogenies including nuclear genes than previous mitogenomic studies, in that it has opisthobranchs as a grade diverging early in the tree and a clade of Panpulmonata (see Wägele *et al.* 2013, for a review).

4.2. Properties of anomalodesmatan mitochondrial genomes

A number of interesting or curious features were discovered in the mitogenomes for the specimens sequenced in this study. In particular we identified problems with determining gene boundaries, a possible incidence of gene duplication, a large range of genome sizes, evidence of pseudogenes in transcriptomic data, and evidence of low levels of intra-specific divergences but very high interspecific divergences. Some of these characteristics are common to all marine bivalves, but others are quite unusual.

4.2.1. Protein-coding genes

Determining gene boundaries for protein-coding genes (PCG) in the Anomalodesmata was not straightforward, exemplifying problems noted for other marine bivalves (e.g. Wang *et al.* 2010). Using GenBank sequences to infer protein-coding boundaries is problematic as this relies on a circular argument of support. Many gene boundaries have been determined using the same software programmes and few, if any, molluscan proteins have been confirmed experimentally. In this study we used transcriptomic data to help determine gene boundaries for protein-coding genes in *Laternula elliptica*, which helped inform our choice of boundaries in other species. This goes some

way towards improving the confidence with which we can state that marine bivalves do not seem to have the same conservation of initiation and termination boundaries seen in other groups. Surprisingly, we noticed that freshwater bivalves, gastropods and cephalopods show greater conservation of 5' gene boundaries than marine bivalves, especially in COX1.

Mitochondrial PCG are notorious for the use of non-standard initiation and truncated stop codons (Boore 2006; Carapelli *et al.* 2008). We found seven start codons, including one non-typical start codon (GTT), which occurred in six taxa and six genes (three times for ND3). This codon has also been found to initiate translation in nematodes (Okimoto & Wolstenholme 1990), but may also reflect difficulties in defining where some genes were initiated.

The ATP8 gene is often thought to be absent from marine bivalve mitogenomes (e.g. Boore *et al.* 2004; Mizi *et al.* 2005) and is also thought to be missing or truncated in Nematoda, Placozoa, Platyhelminthes, Annelida, and Sipunculida (Kim *et al.* 2013). Curiously, it is not missing from freshwater bivalves or gastropods. It has been suggested that the bivalve gene may have been transferred to the nucleus or may be difficult to identify and therefore overlooked during gene annotation (Wang *et al.* 2010; Wu *et al.* 2012a; Wu *et al.* 2012b). In this study we identified regions corresponding to a putative ATP8 gene for all eight taxa examined. There is a conserved start for freshwater bivalves (MPQLSPMSW) but for these anomalodesmatans only the first few residues (MPQ or MPH) are conserved although some later residues are also conserved. The MPH start is shared with lucinid genera *Loripes* and *Lucinella* and the pectinid *Mimachlamys*. MAQ is another conserved start found in venerids (*Fulvia*, *Paphia* and *Meretrix*) and *Panopea* (Hiatelloidea). Although all eight of our species, seven anomalodesmatans and one semelid, had an ATP8 gene, in some cases the gene is short or overlaps another protein-coding gene by several residues. ATP8 genes of similar lengths to the shorter anomalodesmatan genes are found in some other marine bivalves (e.g. lucinids and several *Paphia*, *Meretrix* and *Panopea* species). It is not clear whether such genes are functional; it may be possible to map gene loss over the bivalve tree and show that gene loss is phylogenetically constrained.

A possible case of gene duplication was observed in the mitogenome of the myochamid anomalodesmatan *Myadora brevis*. An open reading frame (ORF) of 1,623 bp with an ATG start and a TAA stop codon was identified between ND5 and ATP8. A BLASTX search against Mollusca showed that approximately 700 bp brings up matches to molluscan ND5 sequences and a putative conserved region matching the oxidored_q1 superfamily. The ORF is AT rich (AT: 67.6%), showing a similar composition to the (putatively) functioning ND5 gene (AT: 69%), but they share no significant similarity at the nucleotide level in a BLASTn comparison. We do not have any EST or transcriptomic data to determine if this region is being expressed, but novel genes have been identified in other

bivalve mitogenomes, leaving open the possibility that this ORF may serve a biological function. Curiously, an ORF of 963 bp was identified in the pteriomorph *Pinctada maxima* mitogenome (Wu *et al.* 2012b), with an ATG start codon, TAA stop codon and the same putative conserved domain as detected in our sequence. These authors found that the gene was highly expressed in EST data from GenBank and suggest that it may play an important role in biological functions (Wu *et al.* 2012b). Other genes found in the mitogenome in other studies include a protein-coding gene that is likely involved in bivalve species with doubly uniparental inheritance (DUI) (Breton *et al.* 2011a; Breton *et al.* 2011b) and a C-terminus extended, male-transmitted COX2 protein in a freshwater bivalve that is thought to play a role in reproduction (Chakrabati et al. 2007). The marine bivalve *Venerupis,* which also exhibits DUI, also has a second tandem copy of COX2 (Serb & Lydeard 2003). Rapid divergence following a duplication of ND2 may have given rise to two novel genes in two species of *Crassostrea* oysters (Wu *et al.* 2012a). Examples from other taxa are given in Breton et al. (2014)

4.2.2. Genome size

The mitogenomes in this study range from very small compact molecules, with the smallest for *L. elliptica* only 14,622 bp to very large genomes. The largest genome is found in *B. lata*, which is at least 31,969 bp long. This genome is incomplete, or at least cannot be readily circularized, because of an extensive repeat region in a non-coding area; a convincing overlap could not be found in the assembled sequence. It is likely that this genome is in fact significantly larger than 32,000 bp as coverage for the genome was not even, with the non-coding portion receiving eight times higher coverage than coding portions of the genome, suggesting a total size >40,000bp.

The range of genomes sizes found in anomalodesmata fit within the range observed for other molluscs. The smallest molluscan mitogenome found to date is in the snail *Biomphalaria glabrata* with a length of only 13.6 kb (DeJong *et al.* 2004) and the largest known was found in *Scapharca subcrenata* at 48.2 kb (Hou *et al.* 2016). The mitogenome for the sea scallop *Placopecten magellanicus* is 30.7 kb when one copy of a 1.4 kb repeat is present (Smith & Snyder 2007), but individuals may have anything from two to eight copies of this repeat region (Fuller & Zouros 1993; La Roche *et al.* 1990). Smaller repeat regions may also have multiple copy numbers (Fuller & Zouros 1993). This suggests that the true size of the genome for this species ranges from about 32 kb in some individuals to almost 41 kb in others (Smith & Snyder 2007).

The difficulties in identifying secondary structures for the tRNA coding for serine, suggests that some of these genes could be non-functional and their loss in future evolutionary trajectories

562 could further decrease genome size. It is interesting to note that we did not have trouble identifying
563 secondary structures for the tRNA coding for serine in the larger genomes.

564

565 **4.2.3. Pseudogenes**

566 Mitochondrial pseudogenes can be insidious and ubiquitous in some marine invertebrates
567 (Williams & Knowlton 2001), but are rarely a problem with molluscs. However, close assessment of
568 the assembly of transcriptomic data for *L. elliptica* showed that a number of reads did not match the
569 consensus sequence. Translation of these sequences into amino acids showed no stop codons using
570 the standard (nuclear) genetic code, but did show stop codons with the appropriate invertebrate
571 mitochondrial code. Given the large number of these reads, we interpret these sequences as
572 pseudogenes rather than sequencing errors. The lack of stop codons in the nuclear code (and the
573 presence of stop codons using the mitochondrial code) suggests that these are nuclear copies that
574 are being transcribed in low numbers. Transcribed copies of pseudogenes have been found in other
575 molluscs (e.g. Korneev *et al.* 2013) and other organisms, including humans, where they may play a
576 role in regulating genes transcriptionally and post-transcriptionally (Milligan *et al.* 2016).

577

578 **4.2.4. Enrichment**

579 Enrichment of mitochondrial DNA allows shotgun sequencing of non-model organisms and
580 multiplexing of samples to decrease running costs, however mitochondrial enrichment of divergent
581 lineages is often unsuccessful. Here we show by minor modification to an existing protocol that we
582 can use mitochondrial sequence from one species (*Laternula elliptica*) as bait to enrich for
583 mitochondria for highly divergent congeneric taxa (*L. rostrata*).

584 Although *Laternula elliptica* and *L. rostrata* are congeneric, they are considerably divergent.
585 Genetic identity of nucleotides shared between the two sequences for protein-coding genes is 67%
586 (based on pairwise Muscle alignments for each gene). The two sequences differ by more than 25%
587 over the barcoding region of COX1 and yet our slight modifications to the mitochondrial enrichment
588 process resulted in a significant increase in coverage for *L. rostrata* in this study. Attempts to enrich
589 the *Bryopa lata* sequence using this method failed suggesting that it is too divergent from the bait
590 species. Although this method was only successful within a single genus, the differences we observe
591 in this bivalve genus are equivalent to deeper taxonomic divergences in many other taxa suggesting
592 our modifications to the enrichment process might be useful in other studies.

593

4.2.5. Divergences

Our sequence for the *L. elliptica* mitogenome differs from a published genomic sequence by only one base pair across the whole genome and from transcriptomic data for protein-coding genes by only 9 bp. The 'universal' barcoding region of COX1 differs at most by only 1 bp. The similarity of these sequences obtained by different methods acts as confirmation of sequence scoring accuracy but also suggests that the mitogenome may not be useful for population genetic studies in *L. elliptica* although further sampling is warranted to test this hypothesis. It may also suggest that a single individual's mitogenome is representative of the entire species. Low mitogenomic variability within species is also consistent with nuclear markers; no genetic differences were detected using 68 polymorphic AFLP markers to examine 96 specimens of *L. elliptica* from two sites in Antarctica, despite population level differences in shell morphology and repair rates (Harper *et al.* 2012). The morphological differences were interpreted as being ecophenotypic in nature, suggesting a plastic response to environmental differences (Harper *et al.* 2012).

Species level divergences among species within Anomalodesmata on the other hand are very high. As discussed above, even species from the same genus show divergences in excess of 25%.

5. Conclusion

The mitogenome is an attractive marker for phylogenetic studies, but our bivalve tree is still not recovering all clades that are expected based on nuclear genes or morphology alone. The inclusion of seven anomalodesmatan and one semelid species into a mitogenome phylogeny for the Autobranchia has resulted in a tree with many robust clades although further improvement is desirable. Relationships among superfamilies or major clades (Pteriomorphia, Palaeoheterodonta, Anomalodesmata, Imparidentia) were not well supported but some major taxon groups are still missing. Compositional bias in the mitogenome may account in part for some anomalous phylogenetic relationships. Improved, more even sampling across the bivalve clade with the inclusion of under represented groups like Archiheterodonta, Neotrigonidae and basal euheterodonts is likely to improve the resolution of mitogenome trees. Increased taxon sampling will become easier as sequencing costs decrease and whole genome sequencing makes sequencing mitogenomes easier, making the need to determine the phylogenetic utility of this marker imperative.

Anomalodesmatan mitogenomes are also shown to have a variety of curious features not seen, or only rarely, in other taxa. For example, protein-coding gene boundaries are not as well defined as in gastropods or freshwater bivalves, gene order is extremely variable among species and species are highly divergent, although intraspecific variation in one species is very low. The use of

transcriptomic and genomic data together enabled better definition of gene boundaries, the identification of possible pseudogenes and suggests that most genes are translated monocistronically.

Acknowledgements

We thank Lisa Smith for help in the lab, Melody Clark for providing transcriptomic data for *Laternula elliptica*, Mark Youngblood & Susan Masta for providing the tRNA drawing script, Gonzalo Giribet for providing the photo of *Myadora brevis* in the graphical abstract and Prashant Sharma and one anonymous reviewer for helpful comments. We gratefully acknowledge the following workshop organisers and Philippe Bouchet (MNHN) for specimens used in this study. *Myadora* was collected during the International Marine Biological Workshop 2005 organised by Peter Davie. Samples from Singapore were collected as part of the Southern Singapore Marine Biodiversity Survey 2013 organized by Tan Koh Siang and colleagues and bivalves from Tjärnö were collected as part of a Swedish Biodiversity Centre workshop organized by Per Sundberg and Malin Strand in 2010. MNHN material was collected during AURORA and PANGLAO 2004 expeditions of the 'Tropical Deep-Sea Benthos' programme under PIs Philippe Bouchet, Danilo Largo (University of San Carlos) and Marivene Manuel (National Museum of the Philippines). PANGLAO 2004 was funded by grants from the Total Foundation, the French Ministry of Foreign Affairs, and the Asean Regional Center for Biodiversity Conservation, and operated under a permit from the Philippine Bureau of Fisheries and Aquatic Resources. AURORA 2007 was funded by the Lounsbery Foundation, and was carried under a *Census of Marine Life / Census of Margins* umbrella. The authors gratefully acknowledge funding from the Department of Life Sciences at the Natural History Museum.

References

Allcock AL, Cooke IR, Strugnell JM (2011) What can the mitochondrial genome reveal about higher-level phylogeny of the molluscan class Cephalopoda? *Zoological Journal of the Linnean Society* **161**, 573-586.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403-410.

Ballard JWO, Whitlock MC (2004) The incomplete natural history of mitochondria. *Mol Ecol* **13**, 729-744.

Bernt M, Donath A, Juhling F*, et al.* (2013) MITOS: improved de novo metazoan mitochondrial genome annotation. *Mol Phylogenet Evol* **69**, 313-319.

Bieler R, Mikkelsen PM, Collins TM*, et al.* (2014) Investigating the Bivalve Tree of Life – an exemplar-based approach combining molecular and novel morphological characters. *Invertebrate Systematics* **28**, 32.

Bleidorn C, Podsiadlowski L, Zhong M*, et al.* (2009) On the phylogenetic position of Myzostomida: can 77 genes get it wrong? *BMC Evol Biol* **9**, 150.

Boore JL (2006) The complete sequence of the mitochondrial genome of *Nautilus macromphalus* (Mollusca: Cephalopoda). *BMC Genomics* **7**, 182.

Boore JL, Medina M, Rosenberg LA (2004) Complete sequences of the highly rearranged molluscan mitochondrial genomes of the scaphopod *Graptacme eborea* and the bivalve *Mytilus edulis*. *Mol Biol Evol* **21**, 1492-1503.

Breton S, Ghiselli F, Passamonti M*, et al.* (2011a) Evidence for a fourteenth mtDNA-encoded protein in the female-transmitted mtDNA of marine mussels (Bivalvia: Mytilidae). *PLoS One* **6**, e19365.

Breton S, Milani L, Ghiselli F*, et al.* (2014) A resourceful genome: updating the functional repertoire and evolutionary role of animal mitochondrial DNAs. *Trends Genet* **30**, 555-564.

Breton S, Stewart DT, Shepardson S*, et al.* (2011b) Novel protein genes in animal mtDNA: a new sex determination system in freshwater mussels (Bivalvia: Unionoida)? *Mol Biol Evol* **28**, 1645-1659.

Carapelli A, Comandi S, Convey P, Nardi F, Frati F (2008) The complete mitochondrial genome of the Antarctic springtail *Cryptopygus antarcticus* (Hexapoda: Collembola). *BMC Genomics* **9**, 315.

Casastrena J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* **17**, 540-552.

Checa AG, Pina CM, Osuna-Mascaró AJ, Rodríguez-Navarro AB, Harper EM (2014) Crystalline organization of the fibrous prismatic calcitic layer of the Mediterranean mussel *Mytilus galloprovincialis*. *European Journal of Mineralogy* **26**, 495-505.

Checa AG, Rodriguez-Navarro AB, Esteban-Delgado FJ (2005) The nature and formation of calcitic columnar prismatic shell layers in pteriomorphian bivalves. *Biomaterials* **26**, 6404-6414.

686 Clark MS, Thorne MA, Vieira FA*, et al.* (2010) Insights into shell deposition in the Antarctic bivalve
687 *Laternula elliptica*: gene discovery in the mantle transcriptome using 454 pyrosequencing.
688 *BMC Genomics* **11**, 362.

689 Combosch DJ, Collins TM, Glover EA*, et al.* (2016) A family-level tree of life for bivalves based on a
690 Sanger- sequencing approach. *Mol Phylogenet Evol* **107**, 191-208.

691 Combosch DJ, Giribet G (2016) Clarifying phylogenetic relationships and the evolutionary history of
692 the bivalve order Arcida (Mollusca: Bivalvia: Pteriomorphia). *Mol Phylogenet Evol* **94**, 298-
693 312.

694 Cunha RL, Grande C, Zardoya R (2009) Neogastropod phylogenetic relationships based on entire
695 mitochondrial genomes. *BMC Evol Biol* **9**, 210.

696 Dabney J, Knapp M, Glocke I*, et al.* (2013) Complete mitochondrial genome sequence of a Middle
697 Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc Natl Acad Sci U S A*
698 **110**, 15758-15763.

699 Darling ACE, Mau B, Blattner FR, Perna NT (2004) Mauve: multiple alignment of conserved genomic
700 sequence with rearrangements. *Genome Research* **14**, 1394-1403.

701 DeJong RJ, Emery AM, Adema CM (2004) The mitochondrial genome of *Biomphalaria glabrata*
702 (Gastropoda: Basommatophora), intermediate host of *Schistosoma mansoni*. *Journal of*
703 *Parasitology* **90**, 991-997.

704 Foster PG (2004) Modeling compositional heterogeneity. *Systematic Biology* **53**, 485-495.

705 Fuller KM, Zouros E (1993) Dipsersed discrete length polymorphism of mitochondrial DNA in the
706 scallop *Placopecten magellanicus*. *Current Genetics* **23**, 365-369.

707 Gascuel O (1997) BIONJ: an improved version of the NJ algorithm based on a simple model of
708 sequence data. *Mol Biol Evol* **14**, 685-695.

709 Giribet G, Distel DL (2003) Bivalve phylogeny and molecular data. In: *Molecular Systematics and*
710 *Phylogeography of Mollusks* (eds. Lydeard C, Lindberg DR), pp. 45-90. Smithsonian Books,
711 Washington, D.C.

712 Giribet G, Wheeler W (2002) On Bivalve Phylogeny- A High-Level Analysis of the Bivalvia (Mollusca)
713 Based on Combined Morphology and DNA Sequence Data. *Invertebrate Biology* **121**, 271-
714 324.

715 Gonzalez VL, Andrade SC, Bieler R*, et al.* (2015) A phylogenetic backbone for Bivalvia: an RNA-seq
716 approach. *Proceedings of the Royal Society B* **282**, 20142332.

717 Harper EM, Clark MS, Hoffman JI*, et al.* (2012) Iceberg scour and shell damage in the Antarctic
718 bivalve *Laternula elliptica*. *PLoS One* **7**, e46341.

719 Harper EM, Dreyer H, Steiner G (2006) Reconstructing the Anomalodesmata (Mollusca: Bivalvia):
720 morphology and molecules. *Zoological Journal of the Linnean Society.*

721 Hou Y, Wu B, Liu Z-H*, et al.* (2016) Complete mitochondrial genome of ark shell *Scapharca*
722 *subcrenata*. *Mitochondrial DNA Part A* **27**, 939-940.

Kim S, Lim BJ, Min GS, Choi HG (2013) The complete mitochondrial genome of Arctic *Calanus hyperboreus* (Copepoda, Calanoida) reveals characteristic patterns in calanoid mitochondrial genome. *Gene* **520**, 64-72.

Knudsen J (1970) The systematics and biology of abyssal and hadal Bivalvia. *Galathea* **11**.

Kocot KM, Cannon JT, Todt C*, et al.* (2011) Phylogenomics reveals deep molluscan relationships. *Nature* **477**, 452-456.

Korneev SA, Kemenes I, Bettini NL*, et al.* (2013) Axonal trafficking of an antisense RNA transcribed from a pseudogene is regulated by classical conditioning. *Sci Rep* **3**, 1027.

Krause J, Fu Q, Good JM*, et al.* (2010) The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature* **464**, 894-897.

La Roche J, Snyder M, Cook DI, Fuller K, Zouros E (1990) Molecular characterization of a repeat element causing large-scale size variation in the mitochondrial DNA of the sea scallop *Placopecten magellanicus*. *Mol Biol Evol* **7**, 45-64.

Lartillot N, Lepage T, Blanquart S (2009) PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25**, 2286-2288.

Laslett D, Canbäck B (2008) ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. *Bioinformatics* **24**, 172-175.

Lee H, Samadi S, Puillandre N*, et al.* (2016) Eight new mitogenomes for exploring the phylogeny and classification of Vetigastropoda. *Journal of Molluscan Studies* **82**, 534-541.

Lemer S, Gonzalez VL, Bieler R, Giribet G (2016) Cementing mussels to oysters in the pteriomorphian tree: a phylogenomic approach. *Proceedings of the Royal Society B* **283**.

Lockhart PJ, Steel MA, Hendy MD, Penny D (1994) Recovering evolutionary trees under a more realistic model of sequence evolution. *Mol Biol Evol* **11**, 605-612.

Maricic T, Whitten M, Paabo S (2010) Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS One* **5**, e14004.

Milligan MJ, Harvey E, Yu A*, et al.* (2016) Global Intersection of Long Non-Coding RNAs with Processed and Unprocessed Pseudogenes in the Human Genome. *Front Genet* **7**, 26.

Mizi A, Zouros E, Moschonas N, Rodakis GC (2005) The complete maternal and paternal mitochondrial genomes of the Mediterranean mussel *Mytilus galloprovincialis*: implications for the doubly uniparental inheritance mode of mtDNA. *Mol Biol Evol* **22**, 952-967.

Morton B (1985) Adaptive radiation in the Anomalodesmata. In: *The Mollusca, Vol. 10. Evolution.* (eds. Trueman ER, Clarke MR), pp. 405-459. Academic Press, Orlando.

Okimoto R, Wolstenholme DR (1990) A set of tRNAs that lack either the TYC arm or the dihydrouridine arm: towards a minimal tRNA adaptor. *EMBO Journal* **9**, 3405-3411.

Osca D, Templado J, Zardoya R (2015) Caenogastropod mitogenomics. *Mol Phylogenet Evol* **93**, 118-128.

759 Park H, Ahn dH (2015) Complete mitochondrial genome of the Antarctic soft-shelled clam, *Laternula*
760 *elliptica* (Bivalvia; Laternulidae). *Mitochondrial DNA* **26**, 642-643.

761 Plazzi F, Ceregato A, Taviani M, Passamonti M (2011) A molecular phylogeny of bivalve mollusks:
762 ancient radiations and divergences as revealed by mitochondrial genes. *PLoS One* **6**, e27147.

763 Plazzi F, Ferrucci RR, Passamonti M (2010) Phylogenetic representativeness: a new method for
764 evaluating taxon sampling in evolutionary studies. *BMC Bioinformatics* **11**, 209-223.

765 Plazzi F, Passamonti M (2010) Towards a molecular phylogeny of Mollusks: bivalves' early evolution
766 as revealed by mitochondrial genes. *Mol Phylogenet Evol* **57**, 641-657.

767 Plazzi F, Ribani A, Passamonti M (2013) The complete mitochondrial genome of *Solemya velum*
768 (Mollusca: Bivalvia) and its relationships with Conchifera. *BMC Genomics* **14**, 409.

769 Rota-Stabelli O, Lartillot N, Philippe H, Pisani D (2013) Serine codon-usage bias in deep
770 phylogenomics: pancrustacean relationships as a case study. *Syst Biol* **62**, 121-133.

771 Serb JM, Lydeard C (2003) Complete mtDNA sequence of the North American freshwater mussel,
772 *Lampsilis ornata* (Unionidae): an examination of the evolution and phylogenetic utility of
773 mitochondrial genome organization in Bivalvia (Mollusca). *Mol Biol Evol* **20**, 1854-1866.

774 Sharma PP, Gonzalez VL, Kawauchi GY*, et al.* (2012) Phylogenetic analysis of four nuclear protein-
775 encoding genes largely corroborates the traditional classification of Bivalvia (Mollusca). *Mol*
776 *Phylogenet Evol* **65**, 64-74.

777 Sharma PP, Zardus JD, Boyle EE*, et al.* (2013) Into the deep: a phylogenetic approach to the bivalve
778 subclass Protobranchia. *Mol Phylogenet Evol* **69**, 188-204.

779 Shen H, Braband A, Scholtz G (2013) Mitogenomic analysis of decapod crustacean phylogeny
780 corroborates traditional views on their relationships. *Mol Phylogenet Evol* **66**, 776-789.

781 Sievers F, Wilm A, Dineen D*, et al.* (2011) Fast, scalable generation of high-quality protein multiple
782 sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539.

783 Smith DR (2016) The past, present and future of mitochondrial genomics: have we sequenced
784 enough mtDNAs? *Brief Funct Genomics* **15**, 47-54.

785 Smith DR, Snyder M (2007) Complete mitochondrial DNA sequence of the scallop *Placopecten*
786 *magellanicus*: evidence of transposition leading to an uncharacteristically large
787 mitochondrial genome. *J Mol Evol* **65**, 380-391.

788 Smith SA, Wilson NG, Goetz FE*, et al.* (2011) Resolving the evolutionary relationships of molluscs with
789 phylogenomic tools. *Nature* **480**, 364-367.

790 Steiner G, Hammer S (2000) Molecular phylogeny of the Bivalvia inferred from 18S rDNA sequences
791 with particular reference to the Pteriomorphia. In: *The evolutionary biology of the Bivalvia*
792 (eds. Harper EM, Taylor JD, Crame JA), pp. 11-29. The Geological Society of London, London,
793 UK.

794 Strong EE (2003) Refining molluscan characters: morphology, character coding and a phylogeny of
795 the Caenogastropoda. *Zoologcal Journal of the Linnean Society* **137**, 447-554.

796 Taylor JD, Glover EA, Williams ST (2008) Ancient chemosynthetic bivalves: systematics of Solemyidae
797 from eastern and southern Australia (Mollusca: Bivalvia). In: *Proceedings of the Thirteenth*
798 *International Marine Biological Workshop, The Marine Fauna and Flora of Moreton Bay,*
799 *Queensland* (eds. Davie PJF, Phillips JA), pp. 75-104. Memoirs of the Queensland Museum —
800 Nature Brisbane.

801 Taylor JD, Williams ST, Glover EA, Dyal P (2007) A molecular phylogeny of heterodont bivalves
802 (Mollusca: Bivalvia: Heterodonta): new analyses of 18S and 28S rRNA genes. *Zoologica*
803 *Scripta* **36**, 587-606.

804 Uribe JE, Kano Y, Templado J, Zardoya R (2016a) Mitogenomics of Vetigastropoda: insights into the
805 evolution of pallial symmetry. *Zoologica Scripta* **45**, 145-159.

806 Uribe JE, Williams ST, Templado J, Abalde S, Zardoya R (2016b) Denser mitogenomic sampling
807 improves resolution of the phylogeny of the superfamily Trochoidea (Gastropoda:
808 Vetigastropoda). *Journal of Molluscan Studies*.

809 Wägele H, Klussmann-Kolb A, Verbeek E, Schrödl M (2013) Flashback and foreshadowing—a review
810 of the taxon Opisthobranchia. *Organisms Diversity & Evolution* **14**, 133-149.

811 Wang H, Zhang S, Li Y, Liu B (2010) Complete mtDNA of Meretrix lusoria (Bivalvia: Veneridae) reveals
812 the presence of an atp8 gene, length variation and heteroplasmy in the control region.
813 *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics* **5**, 256-264.

814 Weigert A, Bleidorn C (2016) Current status of annelid phylogeny. *Organisms Diversity & Evolution*
815 **16**, 345-362.

816 Weigert A, Golombek A, Gerth M*, et al.* (2016) Evolution of mitochondrial gene order in Annelida.
817 *Mol Phylogenet Evol* **94**, 196-206.

818 Williams ST, Foster PG, Littlewood DT (2014) The complete mitochondrial genome of a turbinid
819 vetigastropod from MiSeq Illumina sequencing of genomic DNA and steps towards a resolved
820 gastropod phylogeny. *Gene* **533**, 38-47.

821 Williams ST, Knowlton N (2001) Mitochondrial pseudogenes are pervasive and often insidious in the
822 snapping shrimp genus *Alpheus*. *Mol Biol Evol* **18**, 1484-1493.

823 Wort EJG, Fenberg PB, Williams ST (2016) Testing the contribution of individual genes in
824 mitochondrial genomes for assessing phylogenetic relationships in Vetigastropoda. *Journal of*
825 *Molluscan Studies*.

826 Wu X, Li X, Li L*, et al.* (2012a) New features of Asian *Crassostrea* oyster mitochondrial genomes: a
827 novel alloacceptor tRNA gene recruitment and two novel ORFs. *Gene* **507**, 112-118.

828 Wu X, Li X, Li L, Yu Z (2012b) A unique tRNA gene family and a novel, highly expressed ORF in the
829 mitochondrial genome of the silver-lip pearl oyster, *Pinctada maxima* (Bivalvia: Pteriidae).
830 *Gene* **510**, 22-31.

831 Wyman SK, Jansen RK, Boore JL (2004) Automatic annotation of organellar genomes with DOGMA.
832 *Bioinformatics* **20**, 3252-3255.

833 Ye J, Coulouris G, Zaretskaya I*, et al.* (2012) Primer-BLAST: A tool to design target-specific primers for
834 polymerase chain reaction. *BMC Bioinformatics* **13**, 134.

835 Yuan Y, Li Q, Yu H, Kong L (2012) The complete mitochondrial genomes of six heterodont bivalves
836 (Tellinoidea and Solenoidea): variable gene arrangements and phylogenetic implications.
837 *PLoS One* **7**, e32353.
838

Figure legends

Graphical abstract. Phylogenetic relationships among 105 autobranch bivalves using amino acid sequences from 11 mitochondrial genes as recovered after Bayesian analysis implemented in PhyloBayes. Photographed specimen is *Myadora brevis* from Moreton Bay, Queensland Australia. Photo copyright Gonzalo Giribet.

Fig. 1. Phylogenetic relationships among 107 bivalves and 86 gastropods using amino acid sequences from 11 mitochondrial genes as recovered after Bayesian analysis implemented in PhyloBayes. The tree is a consensus of all trees obtained from five runs, with splits not common to all five runs highlighted in red; all differences are in the gastropod clade. Nodal support is provided by posterior probabilities. For the sake of clarity, not all support values are shown. New mitogenomes sequenced in this study are highlighted in red font.

Fig. 2. Phylogenetic relationships among 105 autobranch bivalves (two protobranch species are excluded) using amino acid sequences from 11 mitochondrial genes as recovered after Bayesian analysis implemented in PhyloBayes. The tree is a consensus of all trees obtained from four runs. Nodal support is provided by posterior probabilities. New mitogenomes sequenced in this study are highlighted in red font.

Appendix A. Supplementary material

Supplementary data 1. Putative secondary structures for mitochondrial tRNAs determined for seven anomalodesmatans and one semelid bivalve. Watson-Crick pairing is shown by lines and G-T pairing by dots. Arrows indicate non-canonical pairings. Secondary structures are not shown for all serine tRNAs.

Supplementary data 2. Whole genome alignments showing gene rearrangements produced by Mauve (Darling *et al.* 2004) implemented in Geneious.

Supplementary data 3. Composition tree based on Euclidean distances. Support values are bootstraps. Nodes with less than 10% support have been collapsed. New mitogenomes sequenced in this study are highlighted in pink font. Two protobranch taxa are highlighted in orange font.

Table 1. Sample details. Registration number for voucher, family and species identifications for specimens, sampling depth (m) and locality (with longitude and latitude and expedition and station number, where available).

Reg. No.	Family	Species	Depth (m)	Sampling Locality
No voucher	Laternulidae	*Laternula elliptica* (King, 1832)	–	Rothera Point, Adelaide Island Ryder Bay, Antarctica 67˚34’S, 68˚8’W
NHMUK 20160581	Laternulidae	*Laternula rostrata* (G. B. Sowerby II, 1839)	0-5	Sungei Buroh, Strait of Johor, Singapore; 1°26’39.3’’ N, 103°43'39.39’’E
NHMUK 20160582	Clavagellidae	*Bryopa lata* (Broderip 1834)	4-8	Singapore; 1°13.9N, 103°52'E
NHMUK 20160583	Cuspidariidae	*Tropidomya abbreviata* (Forbes, 1843)	70	Tjarno; 58°56.608N, 11°05.314'E
NHMUK 20160584	Lyonsiidae	*Lyonsia norwegica* (Gmelin, 1791)	11	Tjarno; 58°52.424N, 11°06.178'N
NHMUK 20070216	Myochamidae	*Myodora brevis* (G. B. Sowerby I, 1827)	31	N Moreton I., Moreton Bay, Qld, Australia; 26°56.60'S, 153°24.25'E
MNHN IM-2009-10750	Euciroidae	*Euciroa cf queenslandica* Lamprell & Healy, 1997	422-431	Philippines; 15°58'N, 121°49'E; MNHN AURORA 2007, CP2658
MNHN IM-2009-10741	Semelidae	*“Abra”* sp	10	Bohol I., Manga, Philippines; 9°41.8'N, 123°51'E; MNHN Panglao 2004, S20

Table 2. Number of amino acid residues (aa) in alignment before and after removal of nonconserved positions with Gblocks. Note that ATP8 was not included in this analysis as it was excluded from phylogenetic analyses given many bivalve taxa were missing this gene. ND4L (in grey font) was also excluded from phylogenetic analyses given that so few sites remained after Gblocks.

Dataset – (n) /Gene	**Original length (aa)**	**Length after Gblocks (aa)**	**% aa remaining**
Bivalves+gastropods – 193 taxa			
ND1	385	180	46.8
ND2	594	72	12.1
ND3	180	44	24.4
ND4	589	218	37.0
ND4L	242	0	0
ND5	775	217	28.0
ND6	282	14	5.0
COX1	660	484	73.3
COX2	895	147	16.4
COX3	487	177	36.3
ATP6	499	79	15.8
CYTB	540	299	55.4
all	5886	1931	32.8
Autobranchia only – 105 taxa			
ND1	366	182	49.7
ND2	493	64	13.0
ND3	183	49	26.8
ND4	536	260	48.6
ND4L	200	17	8.5
ND5	752	232	30.9
ND6	285	32	11.2
COX1	653	474	72.6
COX2	895	81	90.6
COX3	436	166	38.1
ATP6	500	63	12.6
CYTB	544	315	58.0
all	5843	1935	33.1

Table 3. Details of mitochondrial genome for eight bivalves: order of protein-coding genes (order), position of initial base pair (min), position of final base pair (max), length of gene in nucleotides (Len), maximum size difference for each gene (Size Range), longest non-coding region (NCR), size of genome (total genome). The largest gene/genome is highlighted with black and the smallest with grey. Gene order of protein-coding genes is identical in species with purple shading and bold font. New GenBank accession numbers are given under species names.

Species	***Bryopa lata***				***Euciroa cf. queenslandica***				***Laternula elliptica***				***Laternula rostrata***				***Lyonsia norwegica***				***Myadora brevis***				***Tropidomya abbreviata***				***"Abra" sp.***				**Size**
GenBank Acc #	**KX815957**				**KX815958**				**KX815959**				**KX815963**				**KX815960**				**KX815961**				**KX815962**				**KX815956**				**Range**
Gene	**Order**	**Min**	**Max**	**Len**	**Order**	**Min**	**Max**	**Len**	**Order**	**Min**	**Max**	**Len**	**Order**	**Min**	**Max**	**Len**	**Order**	**Min**	**Max**	**Len**	**Order**	**Min**	**Max**	**Len**	**Order**	**Min**	**Max**	**Len**	**Order**	**Min**	**Max**	**Len**	**(bp/aa)**
ATP6	**7**	9,830	10,615	786	8	8,802	9,533	732	**7**	8,265	8,960	696	**7**	9,734	10,429	696	6	4,823	5,540	718	9	11,532	12,311	780	10	11,378	12,052	675	11	13,230	13,943	714	111/37
ATP8	**10**	13,547	13,708	162	9	9,755	9,913	159	**10**	11,813	11,917	105	**10**	13,312	13,416	105	8	7,394	7,504	111	5	6,730	6,933	204	3	2,064	2,291	228	9	11,157	11,294	138	123/41
COX1	**1**	1	1,597	1,597	1	1	1,545	1,545	**1**	1	1,581	1,581	**1**	1	1,569	1,569	1	1	1,569	1,569	1	1	1,596	1,596	1	1	1,605	1,605	1	1	1,689	1,689	144/48
COX2	**12**	14,655	15,383	729	5	5,273	5,980	708	**12**	12,713	13,408	696	**12**	14,346	15,030	685	10	8,296	8,988	693	2	1,665	2,372	708	9	10,405	11,307	903	8	8,380	9,237	858	144/48
COX3	**11**	13,783	14,625	843	13	14,240	15,037	798	**11**	11,908	12,712	805	**11**	13,395	14,201	807	9	7,488	8,295	808	7	7,444	8,295	852	13	16,006	16,797	792	12	14,030	14,842	813	60/20
CYTB	**13**	15,385	16,509	1,125	7	7,658	8,792	1,135	**13**	13,420	14,551	1,132	**13**	15,040	16,179	1,140	11	9,020	10,177	1,158	11	13,446	14,591	1,146	6	4,207	5,382	1,176	7	7,135	8,376	1,242	117/39
ND1	**3**	2,781	3,704	924	10	10,056	10,988	933	**3**	4,211	5,111	901	**3**	2,149	3,072	924	12	10,245	11,157	913	12	14,737	15,669	933	11	13,018	13,948	945	5	4,335	5,258	924	44/14
ND2	**8**	10,632	11,669	1,038	3	2,285	3,278	994	**8**	9,029	10,051	1,023	**8**	10,560	11,600	1,041	13	11,280	12,255	976	10	12,392	13,433	1,042	7	5,559	6,581	1,023	13	14,857	15,918	1,062	47/15
ND3	**6**	9,143	9,481	339	6	7,293	7,652	360	**6**	7,858	8,196	339	**6**	9,056	9,385	330	5	4,234	4,564	331	8	8,371	8,727	357	4	3,170	3,526	357	3	3,224	3,586	363	33/11
ND4	**5**	7,800	9,143	1,344	12	12,791	14,116	1,326	**5**	6,106	7,458	1,353	**5**	7,704	9,050	1,347	4	2,894	4,228	1,335	13	17,705	19,072	1,368	8	6,610	7,977	1,368	2	1,690	3,030	1,341	42/14
ND4L	**4**	7,522	7,806	285	11	12,513	12,791	279	**4**	5,504	5,776	273	**4**	4,316	4,588	273	3	2,619	2,900	282	3	2,664	2,948	285	2	1,704	1,985	282	4	3,721	4,011	291	18/6
ND5	**9**	11,828	13,528	1,701	4	3,351	5,045	1,695	**9**	10,127	11,800	1,674	**9**	11,607	13,274	1,668	7	5,734	7,417	1,684	4	3,104	4,861	1,758	12	14,267	15,982	1,716	6	5,338	7,077	1,740	90/30
ND6	**2**	2,173	2,673	501	2	1,736	2,210	475	**2**	1,650	2,126	477	**2**	1,688	2,149	462	2	2,077	2,553	477	6	6,940	7,428	489	5	3,654	4,148	495	10	11,351	11,893	543	81/27
12S rRNA	–	19,518	20,428	911	–	6,438	7,292	855	–	3,365	4,210	846	–	6,709	7,567	859	–	12,493	13,371	879	–	10,392	11,284	893	–	12,089	12,964	876	–	9,442	10,320	879	65
16S rRNA	–	21,429	22,799	1,371	–	11,344	12,481	1,138	–	2,251	3,364	1,114	–	5,564	6,700	1,137	–	13,407	14,586	1,180	–	8,873	10,050	1,178	–	8,769	9,929	1,161	–	11,978	13,151	1,174	257
Longest NCR				7,025				242				78				826				75				1,754				601				759	
Total genome				**>31,969**				**15,042**				**14,622**				**16,363**				**14,673**				**19,292**				**16,829**				**16,270**	

Table 4. Initiation and termination codons for 13 protein-coding genes in eight bivalve mitochondrial genomes.

Gene/ Species	*Bryopa lata*	*Euciroa cf. queenslandica*	*Laternula elliptica*	*Laternula truncata*	*Lyonsia norwegica*	*Myadora brevis*	*Tropidomya abbreviata*	*"Abra"* sp.
ATP6	GTG/TAG	TTG/TAA	GTG/TAA	ATG/TAG	ATG/T	TTG/TAA	ATA/TAG	ATG/TAA
ATP8	ATG/TAA	ATG/TAA	ATC/TAA	ATG/TAA	ATG/TAA	ATG/TAG	GTG/TAA	ATT/TAG
COX1	ATT/T	ATA/TAG	GTG/TAG	ATA/TAA	ATT/TAG	ATA/TAA	ATT/TAG	GTT/TAA
COX2	ATG/TAA	ATG/TAA	ATG/TAG	ATG/T	ATG/TAA	ATG/TAG	GTG/TAA	ATG/TAA
COX3	ATT/TAA	ATG/TAA	ATT/T	ATG/TAG	ATA/T	ATG/TAG	ATG/TAA	GTG/TAA
CYTB	GTT/TAA	ATG/T	ATA/T	GTG/TAG	GTG/TAA	GTG/TAG	ATA/TAA	ATA/TAG
ND1	GTG/TAA	ATG/TAA	ATG/T	GTG/TAG	GTG/T	ATG/TAA	GTT/TAA	ATT/TAG
ND2	ATG/TAG	ATG/T	ATG/TAA	ATG/TAG	ATG/T	GTT/T	ATA/TAA	ATG/TAG
ND3	GTG/TAA	ATG/TAA	GTG/TAA	GTT/TAA	GTT/T	GTT/TAA	ATG/TAA	ATG/TAA
ND4	ATT/TAG	ATA/TAA	GTG/TAG	GTG/TAA	GTT/TAG	ATT/TAA	GTT/TAA	ATA/TAG
ND4L	ATG/TAG	GTG/TAA	ATG/TAG	ATG/TAG	GTG/TAA	ATG/TAG	ATA/TAA	ATG/TAA
ND5	ATG/TAG	ATG/TAG	ATG/TAA	ATG/TAA	ATG/T	ATG/TAG	ATA/TAA	ATA/TAG
ND6	GTG/TAG	GTG/T	ATG/TAA	ATG/TAG	GTG/TAG	GTG/TAG	ATA/TAG	ATG/TAG

Table 5. Frequency with which start/stop codons are used. (Note: truncated T stop codon becomes TAA by the addition of 3' A residues to the mRNA)

Start	**Stop**
ATG – 46	TAA – 50
GTG – 22	TAG – 39
ATA – 15	T – 15
GTT – 9	
ATT – 9	
TTG – 2	
ATC – 1	

Figure 1

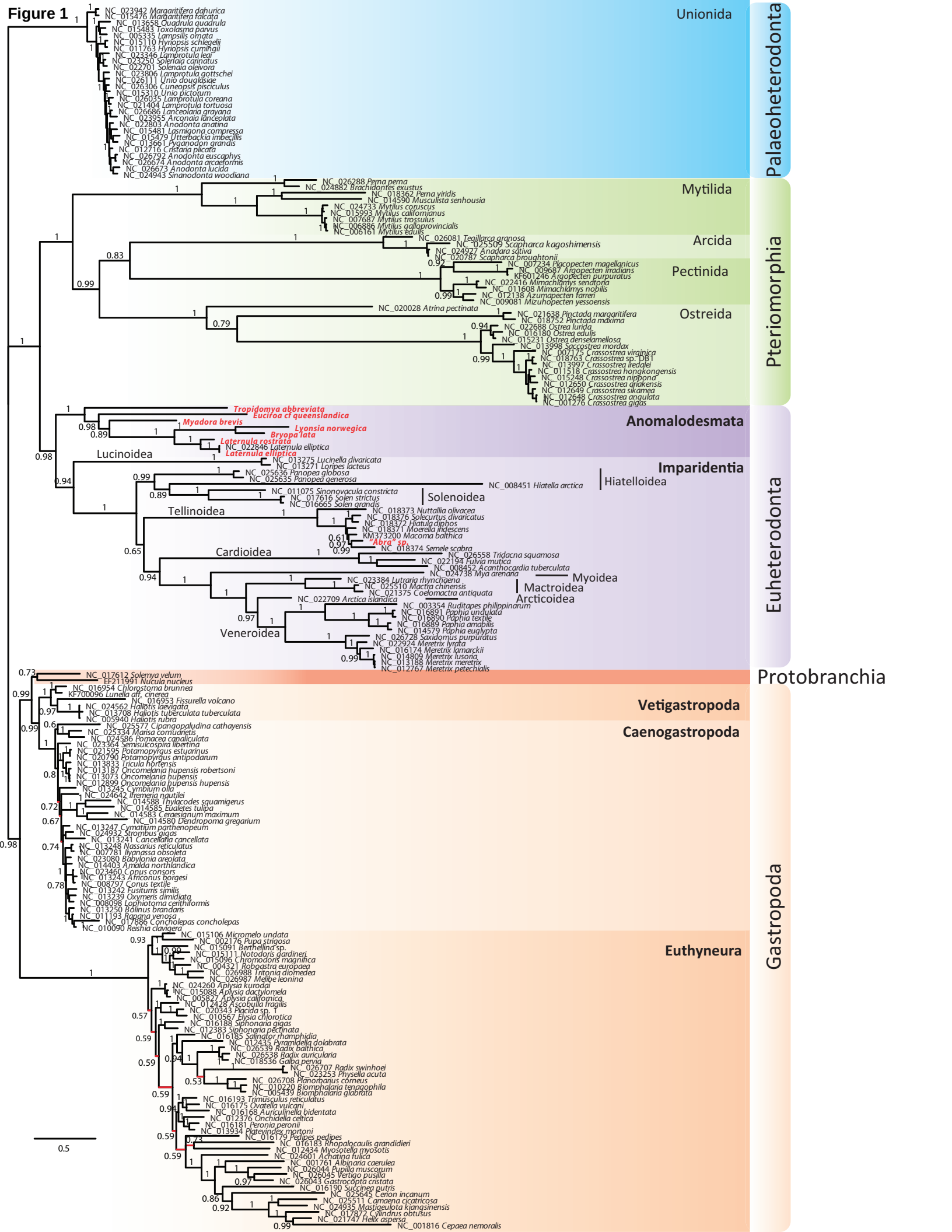
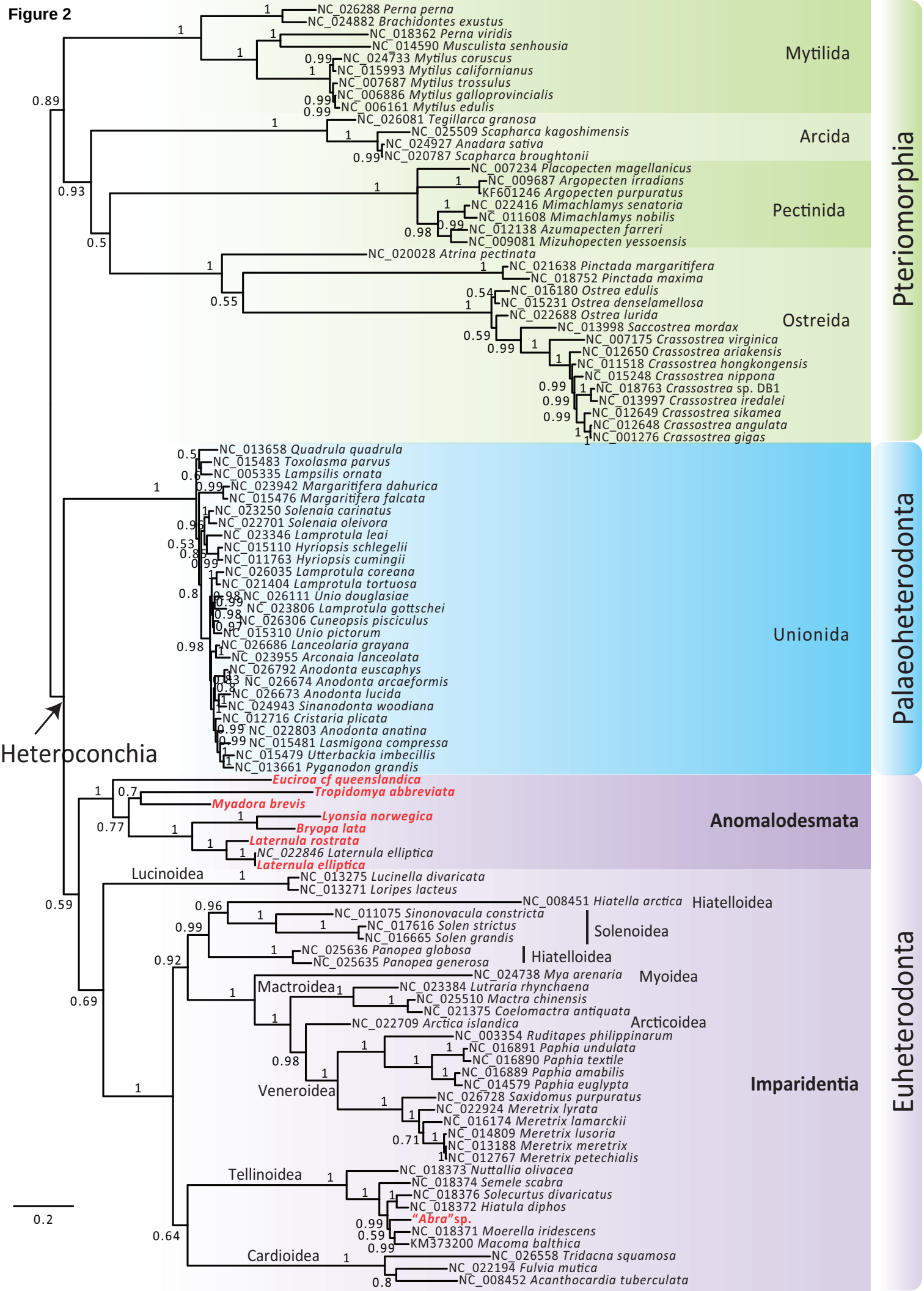

Unionida
Palaeoheterodonta
Mytilida
Arcida
Pectinida
Ostreida
Pteriomorphia
Anomalodesmata
Imparidentia
Hiatelloidea
Solenoidea
Myoidea
Mactroidea
Arcticoidea
Lucinoidea
Tellinoidea
Cardioidea
Veneroidea
Euheterodonta
Protobranchia
Vetigastropoda
Caenogastropoda
Euthyneura
Gastropoda
0.5

Figure 2


NC_026288 Perna perna
NC_024882 Brachidontes exustus
NC_018362 Perna viridis
NC_014590 Musculista senhousia
NC_024733 Mytilus coruscus
NC_015993 Mytilus californianus
NC_007687 Mytilus trossulus
NC_006886 Mytilus galloprovincialis
NC_006161 Mytilus edulis
Mytilida
NC_026081 Tegillarca granosa
NC_025509 Scapharca kagoshimensis
NC_024927 Anadara sativa
NC_020787 Scapharca broughtonii
Arcida
NC_007234 Placopecten magellanicus
NC_009687 Argopecten irradians
KF601246 Argopecten purpuratus
NC_022416 Mimachlamys senatoria
NC_011608 Mimachlamys nobilis
NC_012138 Azumapecten farreri
NC_009081 Mizuhopecten yessoensis
Pectinida
NC_020028 Atrina pectinata
NC_021638 Pinctada margaritifera
NC_018752 Pinctada maxima
NC_016180 Ostrea edulis
NC_015231 Ostrea denselamellosa
NC_022688 Ostrea lurida
NC_013998 Saccostrea mordax
NC_007175 Crassostrea virginica
NC_012650 Crassostrea ariakensis
NC_011518 Crassostrea hongkongensis
NC_015248 Crassostrea nippona
NC_018763 Crassostrea sp. DB1
NC_013997 Crassostrea iredalei
NC_012649 Crassostrea sikamea
NC_012648 Crassostrea angulata
NC_001276 Crassostrea gigas
Ostreida
Pteriomorphia
NC_013658 Quadrula quadrula
NC_015483 Toxolasma parvus
NC_005335 Lampsilis ornata
NC_023942 Margaritifera dahurica
NC_015476 Margaritifera falcata
NC_023250 Solenaia carinatus
NC_022701 Solenaia oleivora
NC_023346 Lamprotula leai
NC_015110 Hyriopsis schlegelii
NC_011763 Hyriopsis cumingii
NC_026035 Lamprotula coreana
NC_021404 Lamprotula tortuosa
NC_026111 Unio douglasiae
NC_023806 Lamprotula gottschei
NC_026306 Cuneopsis pisciculus
NC_015310 Unio pictorum
NC_026686 Lanceolaria grayana
NC_023955 Arconaia lanceolata
NC_026792 Anodonta euscaphys
NC_026674 Anodonta arcaeformis
NC_026673 Anodonta lucida
NC_024943 Sinanodonta woodiana
NC_012716 Cristaria plicata
NC_022803 Anodonta anatina
NC_015481 Lasmigona compressa
NC_015479 Utterbackia imbecillis
NC_013661 Pyganodon grandis
Unionida
Palaeoheterodonta
Heteroconchia
Euciroa cf queenslandica
Tropidomya abbreviata
Myadora brevis
Lyonsia norwegica
Bryopa lata
Laternula rostrata
NC_022846 Laternula elliptica
Laternula elliptica
Anomalodesmata
Lucinoidea
NC_013275 Lucinella divaricata
NC_013271 Loripes lacteus
NC_008451 Hiatella arctica
Hiatelloidea
NC_011075 Sinonovacula constricta
NC_017616 Solen strictus
NC_016665 Solen grandis
Solenoidea
NC_025636 Panopea globosa
NC_025635 Panopea generosa
Hiatelloidea
NC_024738 Mya arenaria
Myoidea
Mactroidea
NC_023384 Lutraria rhynchaena
NC_025510 Mactra chinensis
NC_021375 Coelomactra antiquata
NC_022709 Arctica islandica
Arcticoidea
NC_003354 Ruditapes philippinarum
NC_016891 Paphia undulata
NC_016890 Paphia textile
NC_016889 Paphia amabilis
NC_014579 Paphia euglypta
Veneroidea
NC_026728 Saxidomus purpuratus
NC_022924 Meretrix lyrata
NC_016174 Meretrix lamarckii
NC_014809 Meretrix lusoria
NC_013188 Meretrix meretrix
NC_012767 Meretrix petechialis
Imparidentia
Tellinoidea
NC_018373 Nuttallia olivacea
NC_018374 Semele scabra
NC_018376 Solecurtus divaricatus
NC_018372 Hiatula diphos
"Abra" sp.
NC_018371 Moerella iridescens
KM373200 Macoma balthica
Cardioidea
NC_026558 Tridacna squamosa
NC_022194 Fulvia mutica
NC_008452 Acanthocardia tuberculata
Euheterodonta
0.2