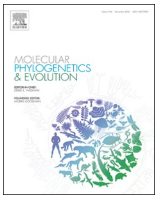
Molecular Phylogenetics and Evolution 150 (2020) 106857

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev



Mitogenomics reveals phylogenetic relationships of Arcoida (Mollusca, Bivalvia) and multiple independent expansions and contractions in mitochondrial genome size

Lingfeng Kong[a,b,1,*], Yuanning Li[c,d,1], Kevin M. Kocot[e], Yi Yang[a], Lu Qi[a], Qi Li[a,b], Kenneth M. Halanych[c,*]

[a] Key Laboratory of Mariculture, Ministry of Education, Ocean University of China, Qingdao 266003, China
[b] Laboratory for Marine Fisheries Science and Food Production Processes, Qingdao National Laboratory for Marine Science and Technology, Qingdao 266237, China
[c] Department of Biological Sciences & Molette Biology Laboratory for Environmental and Climate Change Studies, Auburn University, AL 36849, USA
[d] Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06520, USA
[e] Department of Biological Sciences and Alabama Museum of Natural History, The University of Alabama, Tuscaloosa, AL 35487, USA

ARTICLE INFO

Keywords:
Genome size
Mitogenome
Tandem repeats
Inverted repeats

ABSTRACT

Arcoida, comprising about 570 species of blood cockles, is an ecologically and economically important lineage of bivalve molluscs. Current classification of arcoids is largely based on morphology, which shows widespread homoplasy. Despite two recent studies employing multi-locus analyses with broad sampling of Arcoida, evolutionary relationships among major lineages remain controversial. Interestingly, mitochondrial genomes of several ark shell species are 2–3 times larger than those found in most bilaterians, and are among the largest bilaterian mitochondrial genomes reported to date. These results highlight the need of detailed phylogenetic study to explore evolutionary relationships within Arcoida so that the evolution of mitochondrial genome size can be understood. To this end, we sequenced 17 mitochondrial genomes and compared them with publicly available data, including those from other lineages of Arcoida with emphasis on the subclade Arcoidea species. Our phylogenetic analyses indicate that Noetiidae, Cucullaeidae and Glycymerididae are nested within a polyphyletic Arcidae. Moreover, we find multiple independent expansions and potential contractions of mitochondrial genome size, suggesting that the large mitochondrial genome is not a shared ancestral feature in Arcoida. We also examined tandem repeats and inverted repeats in non-coding regions and investigated the presence of such repeats with relation to genome size variation. Our results suggest that tandem repeats might facilitate intraspecific mitochondrial genome size variation, and that inverted repeats, which could be derived from transposons, might be responsible for mitochondrial genome expansions and contractions. We show that mitochondrial genome size in Arcoida is more dynamic than previously understood and provide insights into evolution of mitochondrial genome size variation in metazoans.

1. Introduction

Ark shells or blood cockles (Order Arcoida Gray, 1854) are a well-known, economically important group of bivalves (Fig. 1). The name blood cockle comes from the presence of hemoglobin, a feature found in all Arcoidea but not exclusive to them among bivalves (Oliver and Holmes, 2006). Ark shells have a long fossil record, dating back to the Lower Ordovician (~450 Mya; Cope, 1997). Today, ark shells are globally distributed with approximately 570 species (Huber, 2010). They occur predominantly in shallow tropical waters and warm temperate seas, and are most diverse in the Indo-West Pacific, where they play important roles in the ecosystem (Nakamura, 2005). Ark shell bivalves are perhaps best known for use as Akagai, red clam or surf clam sushi. World fishery production of clams, cockles, and ark shells is estimated to be 591,000 tons per year, which is worth nearly USD $600,000,000 (FAO, 2018). Production of ark shells in China alone was more than 360,000 tons in 2016 (Chinese National Department of Fisheries, 2017). Because of their economic and ecological importance,

---

* Corresponding authors at: Key Laboratory of Mariculture, Ministry of Education, Ocean University of China, Qingdao 266003, China (L. Kong). Department of Biological Sciences, Auburn University, AL 36849, USA (K.M. Halanych).

E-mail addresses: klfaly@ouc.edu.cn (L. Kong), ken@auburn.edu (K.M. Halanych).

[1] Contributed equally to this work.

https://doi.org/10.1016/j.ympev.2020.106857
Received 7 September 2019; Received in revised form 9 April 2020; Accepted 21 May 2020

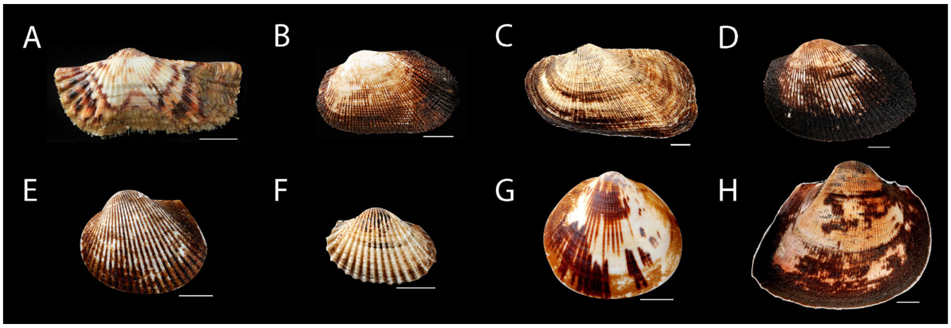


Fig. 1. Representatives of the groups of Arcoida included in the study. A. *Arca navicularis* Bruguière, 1789. B. *Barbatia lima* (Reeve, 1844). C. *Trisidos semitorta* (Lamarck, 1819) D. *Anadara crebricostata* (Reeve, 1844) E. *Scapharca globosa* (Reeve, 1844). F. *Tegillarca nodifera* (Martens, 1860). G. *Glycymeris formosana* (Yokoyama, 1928). H. *Cucullaea labiata* (Lightfoot, 1786). All scale bars 1 cm length.

ark shells have been the subject of considerable research efforts (Sturmer et al., 2009; Morton and Puljas, 2016; Li et al., 2016; Hou et al., 2016).

The taxonomy of Arcoida has primarily been based on shell morphology and palaeontological records. Currently, two superfamilies are recognized: Arcoidea and Limopsoidea. Arcoidea encompasses Arcidae, Noetiidae, Cucullaeidae and Glycymerididae, whereas Limopsidae and Philobryidae are included in Limopsoidea (Oliver and Holmes, 2006). However, current classification of families relies on a limited number of morphological characters. In addition, morphological characters such as ligament and byssus arrangement, which have been used within genera, are homoplastic (reviewed in Oliver and Holmes, 2006). Although Arcoida is well-established as a clade (e.g., Steiner and Hammer, 2000; Giribet and Wheeler, 2005; Matsumoto, 2003; Bieler et al., 2014; Combosch et al., 2017), its internal relationships remain controversial and poorly understood. Recent multi-locus studies (Feng et al., 2015; Combosch and Giribet, 2016) have improved our understanding of evolutionary relationships within Arcoida, but placement of some important lineages remains contentious. In particular, the precise phylogenetic position of Glycymerididae remains unresolved (Combosch and Giribet, 2016). Moreover, Arcidae, which is divided into two subfamilies Arcinae and Anadarinae, based on the strength of the byssus in the attached or free-living forms, e.g., epibyssate and endobyssate, respectively (Newell, 1969), remains a source of inconsistency in Arcoida classification and is in need of taxonomic revision (Feng et al., 2015; Combosch and Giribet, 2016).

Mitogenomics has proven useful in resolving phylogenetic relationships across a wide range of metazoans (e.g. Miya et al., 2001; Osigus et al., 2013; Li et al., 2015; Mikkelsen et al., 2018). Despite the evolutionary importance of Arcoida, genomic resources from this clade are still scarce in comparison to other major bivalve clades. Prior to this study, only seven complete mitochondrial genomes of ark shells were available (Liu et al., 2013; Sun et al., 2015a, 2015b, 2015c; Sun et al., 2016; Feng et al., 2017) and the size of ark shell mitochondrial genomes are unusual. For example, the mitochondrial genome of *Scapharca broughtonii* is 46,985 bp and *Anadara vellicata's* mitochondrial genome is 34,147 bp in length, 2–3 times the 15–17 kb size typically reported in other bilaterians (Boore, 1999). Much of this increase in size is due to the presence of non-coding regions, which have the potential to influence mitochondrial energetics and replication (Sun et al., 2016). Remarkably large mitochondrial genomes are also found in other bivalves (e.g. sea scallop *Placopecten magellanicus* 31–41 kb, Smith and Snyder, 2007; and *Bryopa lata*, Clavagellidae, which is at least 31,969 bp long, Williams et al., 2017). More than 50% of the mitochondrial genome of the scallop *Placopecten magellanicus* is non-coding regions that contain tandemly organized, perfectly repeated sequences and dispersed, imperfect members of repeat families. This indicates that transposition may lead to its uncharacteristically large mitochondrial genome (Smith and Snyder, 2007). However, the relationship between major types of non-coding regions (e.g. tandem repeats, inverted repeats) to genome size variation has not been investigated, and the mechanism of mitochondrial genome expansion of ark shells as well as other bivalves is still not clear.

In this study, we sequenced 17 mitochondrial genomes from representatives of major Arcoida lineages. This information combined with data available from GenBank can serve as a comparative framework for understanding mitochondrial genome evolution of ark shells. We examine three questions: (1) if the phylogenetic relationships inferred from mitochondrial genomes will be consistent with that of the (Feng et al., 2015; Combosch and Giribet, 2016) multi-locus datasets; (2) if the mitochondrial genome of arcoids experiences only expansion; and (3) if there is a correlation between the mitochondrial genome size and repeats. In addition, we estimated divergence times of major clades of Arcoida to date the mitochondrial genome expansion or contraction events in Arcoida molluscs.

2. Materials and methods

2.1. Specimen collection and mitochondrial genome sequencing

Specimen collection data are shown in Table 1. All specimens were preserved by freezing at −80 °C or in 80–100% non-denatured ethanol following collection. *Arca zebra*, *Anadara transversa*, and *Lunarca ovalis* were ordered from Gulf Specimen Supply in Panacea, Florida. These animals were collected locally by the company, but exact latitude and longitude are not available. Total genomic DNA was extracted using the DNeasy Blood & Tissue Kit (Qiagen) following manufacture's protocols. Sequencing libraries were constructed using Illumina TruSeq^{TM} DNA Sample Prep Kit (Illumina, San Diego, CA, USA). For *Anadara transversa* and *Lunarca ovalis*, sequencing of genomic DNA was performed by the Genomic Services Lab at the Hudson Alpha Institute in Huntsville, Alabama using 2 × 125 paired-end on an Illumina HiSeq 2500 platform (San Diego, California). For *Arcopsis adamsi* and *Arca zebra*, sequencing was conducted using 2 × 150 paired-end on the Illumina HiSeq X by Novogene Corporation (USA). All other samples were sequenced on a single lane of an Illumina HiSeq X using 2 × 150 paired-end by Beijing Novogene Technology Co., Ltd (China).

2.2. Mitochondrial genome assemblies and annotation

Raw sequence data for all 17 samples were trimmed using Trimmomatic 0.36 (Bolger et al., 2014) with the parameters "ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36". Resulting clean reads were assembled *de novo* using Ray 2.3.1 (Boisvert et al., 2010) with k-mer = 31 and SPAdes v3.11.1 with k-mer of 21, 33, 55, 77 and the –careful flag (Bankevich et al., 2012). To identify putative mitochondrial contigs, BLAST (Altschul et al., 1997) was performed on each assembly using the previously published *Scapharca broughtonii* mitochondrial genome (GenBank Accession AB729113, Liu et al., 2013). Assemblies of mitochondrial genomes from different assemblers were evaluated using Quast 4.5 (Gurevich et al., 2013) based on genome size, number of genes recovered, GC content, and level of completeness. For *Anadara crebricostata*, *Barbatia virescens*, *Scapharca inaequivalvis*, *Scapharca*

Table 1
New mitochondrial (mt) genomes analyzed in this study.

Species	Mitogenome size (bp)	Coverage depth	Genbank Accession no.	Locality	GPS coordinates
Order Arcoida					
Superfamily Arcoidea					
Family Arcidae					
Subfamily Arcinae					
Arca navicularis	18,004	2455x	MN326818	Beihai, Guangxi, China	N21°22′9″, E109°4′16″
Arca zebra	44,651	565x	MN366003	Panacea, Florida, USA	N/A
Barbatia lima	17,479	380x	MN366005	Beihai, Guangxi, China	N21°0′51″, E109°5′33″
Barbatia virescens	24,871	1386x	MN366006	Qinzhou, Guangxi, China	N21°37′2″, E108°45′9″
Trisidos semitorta	19,461	2042x	MN366015	Beihai, Guangxi, China	N21°22′9″, E109°4′16″
Subfamily Anadarinae					
Anadara crebricostata	36,671	1082x	MN316632	Beihai, Guangxi, China	N21°22′9″, E109°4′16″
Anadara transversa	18,780	256x	MN326817	Panacea, Florida, USA	N/A
Lunarca ovalis	19,620	164x	MN366010	Panacea, Florida, USA	N/A
Scapharca globosa	33,405	1720x	MN366011	Beihai, Guangxi, China	N21°1′34″, E109°6′48″
Scapharca inaequivalvis	45,859	1461x	MN366012	Beihai, Guangxi, China	N21°24′18″, E109°11′26″
Scapharca kagoshimensis	56,170	1827x	MN366013	Beihai, Guangxi, China	N21°24′18″, E109°11′26″
Tegillarca sp.	50,104	918x	MN366016	Fangchenggang, Guangxi, China	N21°29′44", E108°12′49"
Tegillarca nodifera	38,672	428x	MN366014	Fangchenggang, Guangxi, China	N21°29′44″, E108°12′49″
Family Noetiidae					
Arcopsis adamsi	18,716	823x	MN366004	Florida Keys, Florida, USA	N24°40′32″, W081°14′21″
Family Cucullaeididae					
Cucullaea labiata	20,481	755x	MN366007	Beihai, Guangxi, China	N21°22′9″, E109°4′16″
Family Glycymerididae					
Glycymeris formosana	19,027	774x	MN366008	Beihai, Guangxi, China	N21°1′34″, E109°6′48″
Glycymeris yessoensis	17,903	300x	MN366009	Lianyungang, Jiangsu, China	N34°57′11″, E119°12′50″

kagoshimensis, *Tegillarca* sp., however, two or three partial mtDNA contigs were recovered. In an attempt to join these partial contigs, Price 1.2 (Ruby et al., 2013) was employed to extend existing contigs by iteratively adding raw sequence reads to the contig ends as appropriate using default settings. In all cases, partial contigs were successfully bridged together into a single contig. For additional quality control, raw paired-end reads were mapped to the corresponding draft mitochondrial genome using Bowtie 2.3.4.1 (Langmead and Salzberg, 2012) and alignments were visualized in Integrative Genomics Viewer, IGV 2.4.10 (Robinson et al., 2011). Annotation of the protein-coding genes (PCGs), 2 ribosomal RNAs (rRNAs) and transfer RNAs (tRNAs) was conducted initially with the MITOS web server (revision 999, Bernt et al., 2013), with default settings and the invertebrate genetic code for mitochondria, followed by manual genome annotation in Artemis (Rutherford et al., 2000).

Transcriptomic data for *Arca noae*, *Anadara trapezia* and *Neocardia* sp. were download from NCBI SRA database and raw reads were assembled using Trinity v. 2.6.6 (Grabherr et al., 2011) with the “trimmomatic” flag. Mitochondrial protein-coding genes were identified by TBLASTX (Camacho et al., 2009) using the recovered Ark shell mitochondrial genomes above as queries. The full-length sequences of all 17 newly sequenced mitochondrial genomes can be accessed through GenBank (Table 1). Mitochondrial protein-coding genes derived from transcriptomic data acquired from NCBI SRA Database were deposited to figshare (https://doi.org/10.6084/m9.figshare.9746348.v1).

2.3. Identification of tandem repeats, inverted repeats and transposable elements

Tandem repeats, inverted repeats and transposable elements were searched for in the intergenic non-coding regions (NCRs) of new 17 mitochondrial genomes and four mitochondrial genomes from GenBank. Only one specimen was analyzed when multiple mitochondrial genomes were available in the same species. The intergenic regions of each mitochondrial genome were extracted using Artemis (Carver et al., 2012). Tandem repeats were identified using the TRF (tandem repeats finder) V 4.09 (Benson, 1999) with default parameters. Because the same repeat will be detected by TRF at various period sizes, only those with the best alignment scoring for each mitochondrial genome were chosen for further analyses. The Palindrome program from the EMBOSS package (Rice et al., 2000) was employed to search for closely spaced inverted repeats with identical sequences. The following parameters were used: minimum length of palindromes = 7; maximum length of palindromes = 100; maximum gap between elements = 10. Transposable elements are often the hallmark of nuclear-derived insertions (Alverson et al., 2010). With this in mind, each mitochondrial genome was searched against the Repbase repetitive element database (version 23.05, Kohany et al., 2006). Statistical tests were conducted in R v3.5.0 (R Development Core Team 2018). As the data of mitochondrial genomic features are not necessarily normally distributed, nonparametric Spearman’s ρ values were calculated for the correlation coefficient.

2.4. Phylogenetic analyses

Thirty-four taxa were included in the phylogenetic analyses of which data for eight additional Ark shell and six outgroup species were retrieved from GenBank (Tables 1 and 2). The outgroup species *Mizuhopecten yessoensis* and *Argopecten irradians* from Pectinidae, *Pinctada maxima* from Pteriidae, *Ostrea denselamellosa*, *Crassostrea hongkongensis* and *Crassostrea gigas* from Ostreidae were used based on data availability and current understanding of bivalve evolutionary history (Gonzalez et al. 2015; Lemer et al. 2016). Given that Ark shells date to the Ordovician (Cope, 1997), relationships were reconstructed based on amino acid sequences from twelve mitochondrial protein genes (*atp6, cox1, cox2, cox3, cob, nad1, nad2, nad3, nad4, nad4l, nad5* and *nad6*). Each gene was individually aligned using MAFFT (Katoh and Standley, 2013). Selected genes were then trimmed using the default setting in Gblocks (Talavera and Castresana, 2007) to remove ambiguously aligned regions with default parameters. Resulting alignments were concatenated into final supermatrix using FASconCAT (Kück and Meusemann, 2010) for downstream phylogenetic analysis.

Maximum-likelihood (ML) analyses were conducted using IQ-TREE 1.6.1 (Nguyen et al., 2015). Partition schemes and best-fit models were selected by PartitionFinder (Lanfear et al., 2012). Nodal support was assessed with 1000 replicates of ultrafast bootstrapping (-bb 1000).

Bayesian inference (BI) with the site-heterogeneous CAT-GTR substitution model (Lartillot and Philippe, 2004) was employed using

Table 2
Taxa from Genbank used in phylogenetic analysis. Superscript: 1 = *Trisidos kiyonoi* in Sun et al. (2016) and on GenBank; 2 = *Scapharca subcrenata* in Hou et al. (2016) and *Anadara sativa* on GenBank.

Species	Mitogenome size (bp)	Genbank/SRA Accession no.	Source
Family Arcidae			
Subfamily Arcinae			
Anadara trapezia	N/A	SRX323049	Prentis & Pavasovic (2014)
Arca noae	N/A	SRX687762/SRS690228	González *et al.* (2015)
Trisidos semitorta[1]	19,613	KU975161	Sun et al., 2016
Subfamily Anadarinae			
Anadara vellicata	34,147	KP954700	Sun et al., 2015c
Potiarca pilula	28,386	KU975162	Sun et al., 2016
Scapharca broughtonii (1)[2]	48,161	KF667521	Hou et al., 2016
Scapharca broughtonii (2)	46,985	AB729113	Liu et al., 2013
Scapharca kagoshimensis	46,713	KF750628	Sun et al., 2015a
Tegillarca granosa	31,589	KJ607173	Sun et al., 2015b
Family Cucullaeididae			
Cucullaea labiata	25,845	KP091889	Feng et al., 2017
Family Philobryidae			
Neocardia sp.	N/A	SRX701839/SRS701705	González *et al.*, 2015
Outgroups			
Family Pectinidae			
Mizuhopecten yessoensis	20,964	FJ595959.1	Wu et al., 2009
Argopecten irradians	16,221	NC_009687.1	N/A
Family Pteriidae			
Pinctada maxima	16,994	NC_018752.1	Wu *et al.*, 2012
Family Ostreidae			
Ostrea denselamellosa	16,277	NC_015231.1	Yu & Li, 2011
Crassostrea hongkongensis	16,475	EU266073.1	Yu *et al.*, 2008
Crassostrea gigas	18,224	AF177226.1	N/A

PhyloBayes MPI 1.6j (Lartillot et al., 2013). The concatenated AA dataset was run with four parallel Markov chain Monte Carlo (MCMC) chains for 20,000–30000 generations each. Burn-in was determined based on trace plots of log likelihood scores as viewed with plot_phylobayes_traces.R script (https://github.com/wrf/graphphylo). Chains were considered to have reached convergence when the maxdiff value was less than 0.1 as measured by bpcomp (Lartillot et al., 2013), the rel_diff value was less than 0.3, and the effective sample size was greater than 50 as measured by tracecomp (Lartillot et al., 2013). A 50% majority rule consensus tree was computed with bpcomp and nodal support was estimated by posterior probability.

2.5. Ancestral mitogenome size reconstruction

Ancestral mitogenome sizes (with 95% confidence interval) were reconstructed using maximum likelihood under a Brownian motion model by the anc.ML function of the R package Phytools (Revell 2012) on the phylogeny inferred with the molecular clock analyses (see below). The species *Arca noae*, *Anadara trapezia* and *Neocardia* sp. were removed from the phylogeny for this analysis using the drop.tip function in the R package Ape (Paradis et al., 2004) since the mitogenome size is not available.

2.6. Molecular clock analyses

A Bayesian relaxed-clock approach was employed to estimate divergence times of major clades of Arcoida using BEAST 2.5.0 (Bouckaert et al., 2014). The molecular clock was estimated using the uncorrelated lognormal distribution clock model, random starting trees and a Calibrated Yule model. Divergence times were calibrated with three timepoints from the fossil record. The age of the last common ancestor of Bivalvia was constrained with a uniform distribution prior between 520.5 and 530 Mya (Bieler et al., 2014). Arcida was constrained using a normal distribution prior spanning 471.8–488.6 Ma, based on *Glyptarca serrata* Cope, 1996 (Arenigian; Cope, 1997). The age of Northwestern Pacific *Anadara* species was constrained using a normal distribution prior with the means 138.3 Mya and standard deviations of 5 Mya, based on *Anadara ferruginea* (Jaccard, 1869; Huber, 2010). We performed two independent BEAST runs (different seeds) for 20 million generations and with a sampling frequency of 1000 generations. We combined both runs using LogCombiner and after checking for convergence of the runs with Tracer, the first 30% of generations were discarded as the burn-in period. The maximum clade credibility tree with median ages and the 95% credibility interval (CI) at each node was computed using TreeAnnotator.

3. Results

3.1. Mitochondrial genome assemblies and general features

Of the 17 new taxa, complete mitochondrial genomes were recovered from 15 (Table 1). The remaining two were nearly complete: *Arca zebra* was missing rrnS, *Barbatia virescens* was missing trnA, trnE and trnV. Mitochondrial genome sequences varied in size from 17,479 bp (*B. lima*) to 56,170 bp (*S. kagoshimensis*). Four mitochondrial genomes had sequence lengths more than 40 kb and three ranged from 30 kb to 40 kb (Table 1). The average coverage of mitochondrial genomes varied from 164-fold to > 1,000-fold coverage.

Nucleotide composition of the mitochondrial genome was biased toward A and T for all species (Table S1). The A + T content ranged from 54.06% in *B. virescens* to 67.84% in *S. kagoshimensis*. GC-skew and AT-skew for a given strand were calculated as (G - C)/(G + C) and (A - T)/(A + T), respectively, with negative values in skewness meaning the coding strand is enriched for C or T. We found the values of the AT-skew were mostly negative, while values of the GC-skew were all positive (Table S1).

Six different start codons were observed (Table S2 & S3) but most protein-coding genes (50.49%) start with the codon ATG. Similar results were observed in mitochondrial genomes of anomalodesmatan bivalves (Williams et al., 2017). In most genes, complete stop codons were identified (TAG or TAA), but a truncated stop codon T, which may be modified to a complete TAA stop codon via posttranscriptional polyadenylation (Ojala et al., 1981), was found in *cox1-3, cob* and *nad3* (Table S2).

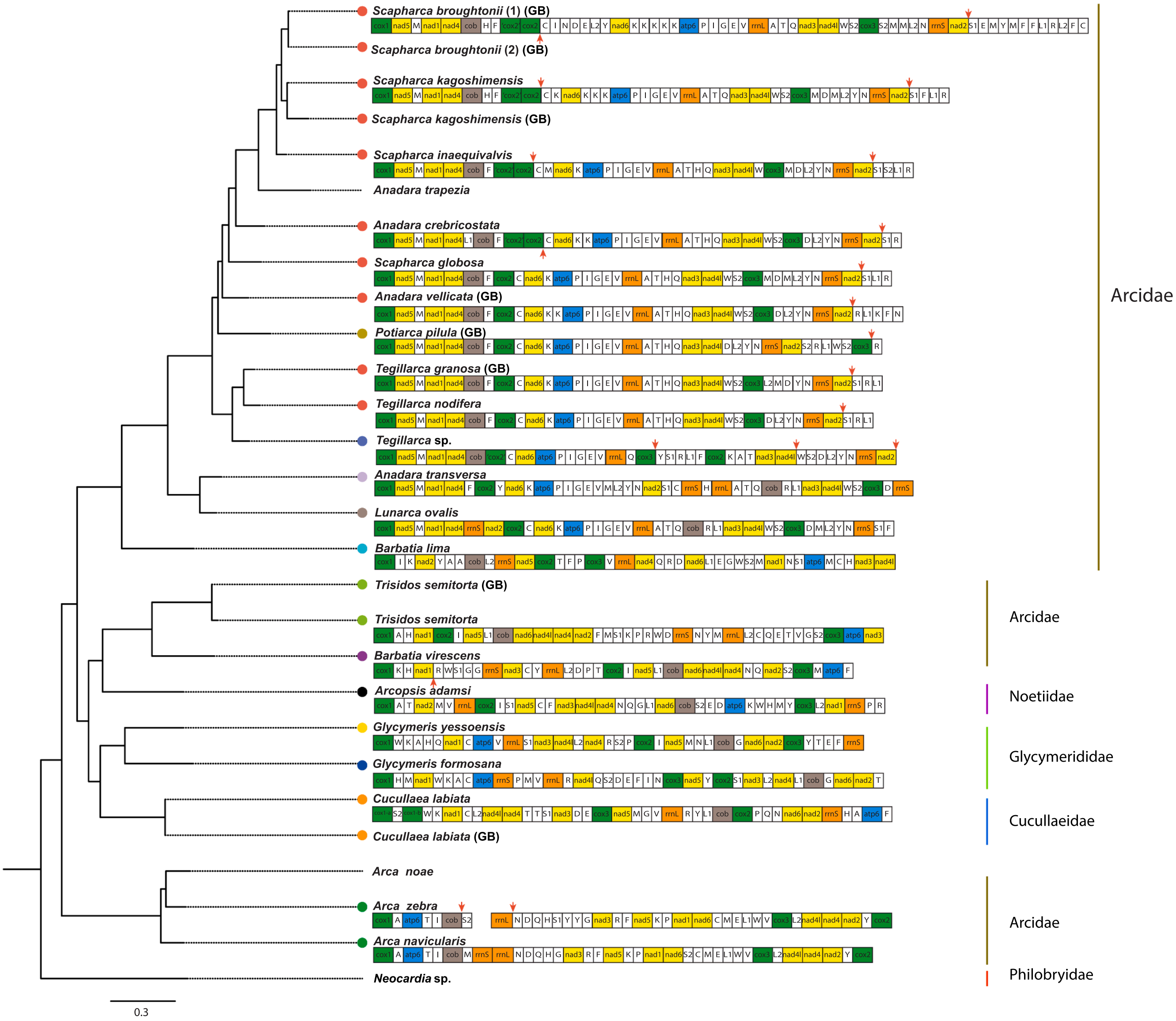


Fig. 2. Organization of mitochondrial genomes of arcoid bivalves sampled. All genes are encoded on the same strand. Phylogenetic relationships were recovered by Bayesian inference utilizing amino acid alignments (See details in Fig. 4). Taxon tips are colored to show gene order of protein-coding genes and ribosomal genes. The different colored circles indicate different gene order. GB means the mitogenome data were retrieved from GenBank. Where two specimens of the same species were sampled, the gene order of only one specimen is shown. Genes are denoted by standard abbreviations and are not scaled according to length. Possible cases of *cox2* duplication were marked with single quotes. Cox1-a and cox1-b represent the two exons of cox1 in the mitogenome of *Cucullaea labiata*. Small red arrows indicate where the major insertions (over 5 kb) occur. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

All genes are encoded on the same strand, as typically found in other bivalves (Williams et al., 2017). All complete mitochondrial genomes sequenced herein are composed of 12 protein-coding genes (all taxa lacked *atp8*), two ribosomal (rRNA) genes, and 21–29 tRNAs. Gene order of protein-coding genes and rRNA genes (Fig. 2) differed among taxa, but is generally conserved in closely related species. Arrangements of tRNAs are highly positional variable across the arcoid mitochondrial genomes, but there are several conserved gene blocks within different lineages. The tRNA complex trnP-trnI-trnG-trnE-trnV is present in all species of *Scapharca*, *Anadara*, *Potiarca*, *Tegillarca* and *Lunarca*. There are two tRNA complexes (trnN-trnD-trnQ-trnH and trnC-trnM-trnE-trnL1-trnW) shared by *A. zebra* and *A. navicularis*.

Of the protein-coding genes and rRNA genes, six cases of gene duplication were observed: *cox2* is duplicated in *Anadara crebricostata*, *S. inaequivalvis*, *S. kagoshimensis*, and *Tegillarca* sp. and *rrnS* is duplicated in *A. transversa* and *Lunarca ovalis* (Fig. 2). Two copies of *cox2* in *A. crebricostata*, *S. inaequivalvis* and *S. kagoshimensis* were sequentially organized and located downstream of *cob*, while one copy of *cox2* in *Tegillarca* sp. is located on downstream of *cob* and the other is downstream of *cox3*. Overall identity between the two *cox2* sequences in each species where it is duplicated is low. However, for *A. transversa* and *L. ovalis*, the sequences of the two copies of *rrnS* in each species are nearly identical. Consistent with a previous study (Feng et al., 2017), the newly sequenced mitochondrial genome of *Cucullaea labiata* revealed the presence of an intron within *cox1*. The intron was 606 bp long and no open reading frames were found. Only two nucleotide bp were found different of the overlap 606 bp when we aligned the intron sequences of *C. labiata* from this and previous study.

3.2. Non-coding regions

The total length of intergenic non-coding region ranged from

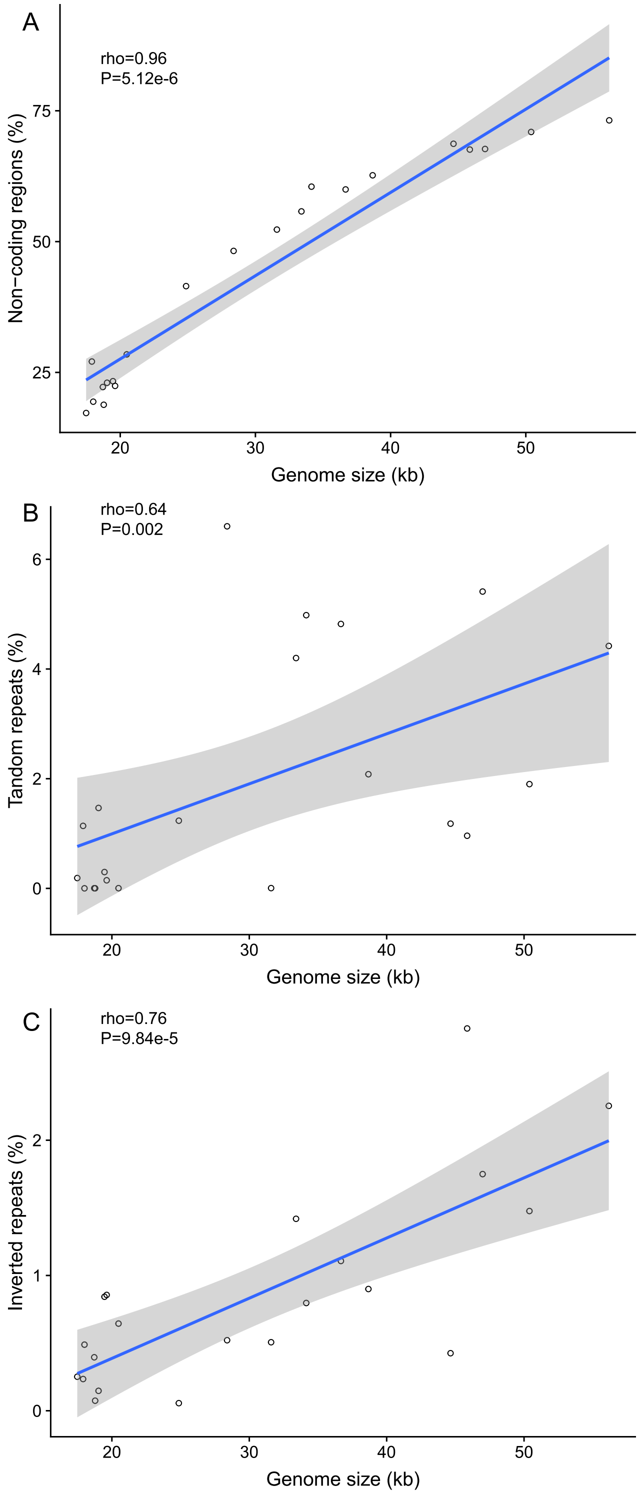


Fig. 3. Contribution of non-coding regions, tandem-, and invert-repeats to mitogenome variation among 21 ark shell species. (A–C) Scatter plots (with Spearman's correlation coefficient ρ and P value) of mitogenome size against the proportion of (A) non-coding regions, (B) tandem repeats, and (C) inverted repeats in each mitogenome.

3,013 bp (17.24% of the genome) in *B. lima* to 41,101 bp (73.17% of the genome) in *S. kagoshimensis* (Table S4). All sequences are AT-rich and the A + T content of the intergenic non-coding regions in each mitochondrial genome is higher than that of the whole mitochondrial genome except *B. virescens*, and the AT-skew for non-coding region had a negative value in all species and showed a large variation (Table S1). A significant and strong positive correlation is observed between mitochondrial genome size and proportion of non-coding region ($\rho = 0.96$, $P = 5.12\mathrm{e}-6$; Fig. 3A). Tandem repeats and inverted repeats in intergenic non-coding regions show significant variation in number and sequence lengths (Table S4). We found significantly positive correlations between mitochondrial genome size and proportion of tandem repeats and proportion of inverted repeats ($\rho = 0.64$, $P = 0.002$, Fig. 3B; $\rho = 0.76$, $P = 9.84\mathrm{e}-5$, Fig. 3C).

3.3. Phylogenetic analyses

The concatenated alignment of amino acid sequences from 34 taxa had a total length of 4,866 positions. Both ML and BI analyses of the two concatenated datasets yielded identical branching orders with high posterior probabilities (PP) and bootstrap support values (BS, Fig. 4). Arcidae was found to be polyphyletic with three well-supported lineages. Three *Arca* species formed a lineage; the second lineage included *Barbatia virescens* and *Trisidos* species; and the third enclosed the subfamily Anadarinae and the sister taxon *Barbatia lima*. Northwestern Pacific Anadarinae species which included *Anadara*, *Scapharca*, *Potiarca* and *Tegillarca* were sister to the Gulf of Mexico Anadarinae species *A. transversa* and *Lunarca ovalis*. The only representative of Noetiidae, *Arcopsis adamsi*, was found to nest within the polyphyletic Arcidae as the sister taxon of the *B. virescens/Trisidos* clade. Arcoidea, which encompasses Arcidae, Noetiidae, Cucullaeidae and Glycymerididae, forms a clade that is not well-supported in the ML analysis (PP = 1; BS = 56%) and *Neocardia* sp. (Limopsoiea) is the sister taxon of Arcoidea (PP = 1; BS = 100%). *C. labiata*, which is the only extant species of Cucullaeidae, was strongly supported as the sister taxon of Glycymerididae (PP = 1; BS = 92%).

3.4. Ancestral state reconstruction

Ancestral state reconstruction indicates that the evolution of the mitochondrial genome size has experienced different trends across different arcoid lineages. When considering a difference of > 4 kb from their respective recent ancestral mitogenomes, mitochondrial genome expansion is apparent in Northwestern Pacific Anadarinae species and *Arca zebra*, but the opposite trend of genome contraction occurs in *Arca navicularis*, *Barbatia lima*, and the Gulf of Mexico Anadarinae species (Fig. 5). Notably, both mitochondrial genome expansion and contraction were found within the Northwestern Pacific Anadarinae linage, e.g. *Tegillarca* sp. and a clade comprising *S. broughtonii*, *S. kagoshimensis* and *S. inaequivalvis* occur expansions, and *T. granosa*, *Potiarca pilula* experience contraction. A medium-sized mitochondrial genome (23.4 kb) was estimated to be the ancestral state of Arcoidea.

3.5. Divergence times

Bayesian inference with a relaxed molecular clock recovered the same topology as ML and BI analyses with strong support (Fig. 6), with the exception of the placement of *A. noae* and *A. navicularis*. The dated topology shows that *A. noae* is the sister taxon of the clade containing *A. navicularis* and *A. zebra*. Our analyses suggest size of mitochondrial genome expansions in Northwestern Pacific Anadarinae species (34.2 kb) primarily originated approximately 158 MYA (95% highest posterior density interval [HPD] 139.9–184.4), while mitochondrial genome contractions in the Gulf of Mexico Anadarinae species (20.5 kb) happened at 58.9 MYA (95% HPD 15.6–132.3). Within the Northwestern Pacific Anadarinae linage, a clade comprising *S. broughtonii*, *S. kagoshimensis* and *S. inaequivalvis* experienced expansions (48.1 kb) at 61 MYA (95% HPD 36.9–84.1). Moreover, mitochondrial genome expansion of *A. zebra* (44.7 kb) and contractions of *A. navicularis* (18.0 kb) occurred at 54.8 MYA (95% HPD 11.0–91.4).

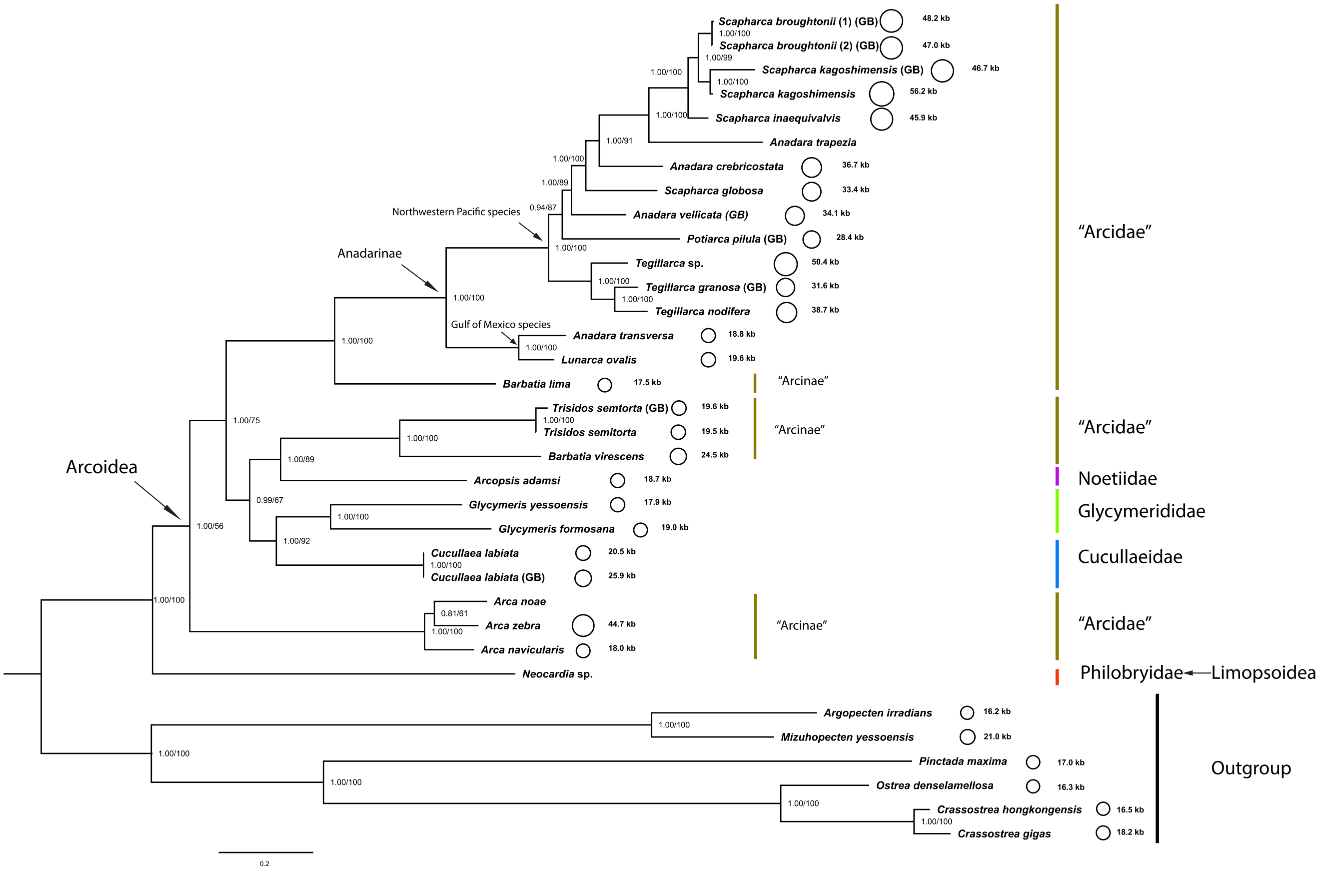


Fig. 4. Phylogenetic reconstruction of Arcoida based on concatenated amino acids of the 12 mitochondrial protein-coding genes. Majority rule (50%) consensus phylograms from the Bayesian analyses of the concatenated data matrices are shown. Values are shown next to nodes with posterior probabilities left and ML bootstrap support values right. Non-monophyletic Arcidae and Arcinae were marked with quotes. Mitogenome sizes are shown are circles to the right of taxon names are to scale.

4. Discussion

4.1. Arcoid phylogeny

Arcidae was recovered as a polyphyletic group, consistent with earlier studies (e.g. Marko, 2002; Matsumoto, 2003; Feng et al., 2015; Combosch and Giribet, 2016), whereas Anadarinae was recovered as monophyletic in both ML and BI analyses (Fig. 4), consistent with previous findings (Marko, 2002; Matsumoto, 2003; Feng et al., 2015; Combosch and Giribet, 2016). Within Anadarinae, *Anadara* and *Scapharca* were found to be polyphyletic which suggest that extensive taxonomy revisions in the genus level are needed. Interestingly, Anadarinae was further split into two subclades based on geographic locality, northwestern Pacific species and Gulf of Mexico species. By comparison, Arcinae was found to be non-monophyletic, with three well-supported lineages. The three *Arca* species were recovered in one clade, sister to all other Arcoidea species with high posterior probability (PP = 1). Morphologically, this clade corresponds to the "*noae*" morphotype within the genus *Arca* of Oliver and Holmes (2006). The second clade included *Barbatia virescens* and *Trisidos* species, shows *Trisidos* and *Barbatia* shared a common ancestor. The other *Barbatia* lineage, which was only represented by *B. lima*, was sister to the clade Anadarinae with high nodal support, consistent with Combosch and Giribet (2016). Results from the present study support raising these three lineages to the familial rank. However, confirmation from future studies including missing genera such as *Bathyarca* and *Bentharca* is needed. Moreover, reconsideration of morphological characters used in the traditional taxonomy of the group in light of these results is needed as it is evident that many of these characters exhibit homoplasy (Oliver and Holmes, 2006).

The phylogenetic position and taxonomic status of Glycymerididae has been contentious and nesting Glycymerididae within Arcoidea or Limopsoidea remains unresolved (see Feng et al., 2015; Combosch and Giribet, 2016 for details). In the present study, both ML and BI phylogenetic analyses support placement of Glycimerididae within Arcidea (Fig. 4), in agreement with previous studies based on combinations of nuclear 18S rRNA, 28S rRNA, histone H3, mitochondrial 12S and COI data (Feng et al., 2015), and corroborated the morphological analyses of Amler (1999) and Oliver and Holmes (2006). *Cucullaea labiata,* the only extant species within Cucullaeidae, was recovered as the sister taxon of a clade containing *Bathyarca glomerula* and *Arca boucardi* (Combosch and Giribet, 2016), however, the phylogenetic position of the clade containing the three species was not congruent among parsimony, maximum likelihood and Bayesian inference analyses (Combosch and Giribet, 2016). Although *B. glomerula* and *A. boucardi* were not included in our study, our analyses support the phylogenetic position of *C. labiata* as the sister lineage of Glycimerididae both ML and BI analysis. *Arcopsis adamsi* (Noetiidae) was recovered as the sister taxon of the *B. virescens/Trisidos* clade, consistent with that of Feng et al. (2015) and Combosch and Giribet (2016).

4.2. Genome size

We found evidence for multiple expansions and contractions of the mitochondrial genome (Fig. 5), which implies that mitochondrial genome size evolution could have been involved in the diversification process. Based on ancestral state reconstruction, the large mitochondrial genome found in previous studies (Liu et al., 2013; Sun et al.,

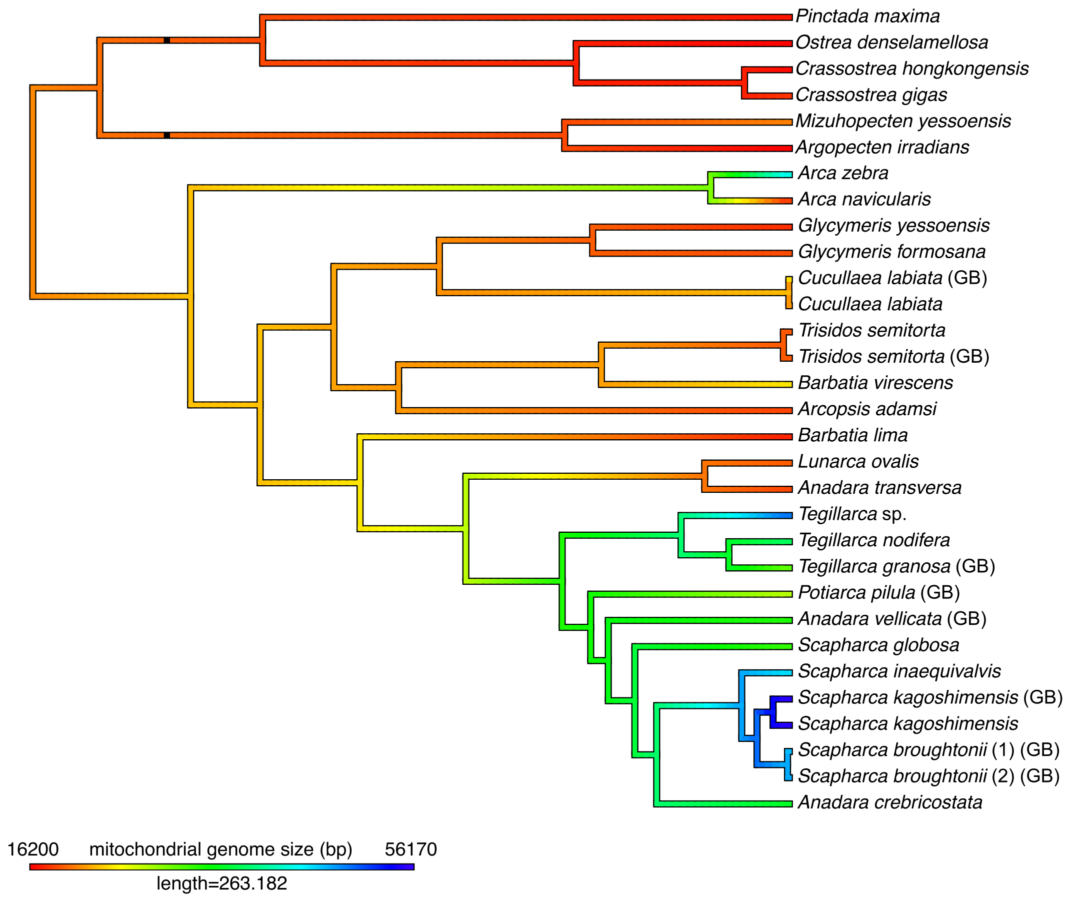


Fig. 5. Ancestral reconstructing the evolution of mitochondrial genome size across the time-calibrated phylogenetic tree of arcoids. Color coding of the branches represents mitochondrial genome size with blue being larger (up to 56,170 bp) and red being shorter (16,200 bp). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2015c) is not a shared ancestral feature of arcoids. Molecular clock analyses suggest mitochondrial genome expansions in Northwestern Pacific Anadarinae species primarily originated approximately 158 million years ago, but mitochondrial genome occurred contraction in the Gulf of Mexico Anadarinae species about 59 million years ago. Interestingly, *Arca zebra*, a Gulf of Mexico Arcinae species has undergone an expansion event while *A. navicularis*, a Northwestern Pacific Arcinae species underwent contraction about 55 million years ago (Fig. 5, Fig. 6), which indicates that mitochondrial genome expansion and contraction are independent evolutionary events. Despite the fact that the mitochondrial genome does not have recombination and therefore only represents the history of one marker, we stress that conclusions from mitochondrial sequences regarding divergence times are comparable with those from concatenated nuclear or nuclear and mitochondrial gene sequences data (Bieler et al., 2014; Combosch and Giribet, 2016). The profound climatic and ecological changes that occurred during this period (e.g., Cretaceous–Paleogene mass extinction and Paleocene–Eocene Thermal Maximum) may have contributed to the adaptive radiation of these arcoids to new niches, which then influence the mitochondrial genome evolution process. Animal mitochondrial genomes have purportedly been under selection for small genome size and are generally devoid of noncoding sequences to improve replication and translation efficiency (Rand, 1993). However, several studies show that low metabolic rates and limited locomotive ability are correlated with weak selective constraints on mitochondrial genes (e.g., Chong and Mueller, 2013; Sun et al., 2017). Up to now, all remarkably large mitochondrial genomes found in molluscs are from bivalves, a group that is primarily sedentary, except the limpet *Lottia digitalis*, of which the mitochondrial genome size is 26.8kb (Simison et al., 2006). We propose that the large mitochondrial genome size in these bivalves may be correlated with their metabolic rates (Strotz et al., 2018). If the metabolic rate is low, bivalves perhaps experienced weak purifying selection on their long intergenic regions.

As seen in Fig. 3A, the size variation of mitochondrial genomes is mainly due to the different lengths of the NCRs. Several characteristics were revealed when the NCRs were examined in detail. Firstly, the distribution of the NCRs was variable among different linages (Fig. 2). The two large non-coding regions in *A. zebra* located between *cob* and *rrnL*, *rrnL* and *nad3*, respectively. In the mitochondrial genomes of *S. broughtonii*, *S. kagoshimensis*, *S. inaequivalvis* and *Anadara crebricostata*, the two largest NCRs were observed between *cox2* and *nad6*, *nad2* and *cox1*. *Tegillarca* sp. has three main non-coding regions which located separately between *nad2* and *cox1*, *cox3* and *cox2*, *nad4l* and *rrnS*. However, *S. globosa*, *A. vellicata*, *T. nodifera* and *T. granosa* have only one large concentrated non-coding region, located between *nad2* and *cox1*. Although less remarkable in size, several cases of major NCRs have previously been reported in other bivalves mitogenomes. However, the intergenic location of these NCRs does not coincide with that found in ark shells. Major NCRs were described between *nad4* and *nad1* in three scallops (Pectinidae; Wu et al., 2009), and between *atp6* and *nad2* in *Crassostrea iredalei* (Ostreidae; Wu et al., 2010). The different distribution of NCRs in ark shells and close related taxa strongly suggests multiple independent insertions instead of a single ancestral event with subsequent loss in other lineages.

Secondly, among the 21 known mitochondrial genomes from ark shells, one to ten tandem repeats were detected in sixteen species (Table S4). Significant correlations were observed between mitochondrial genome size and proportion of each mitochondrial genome made up of

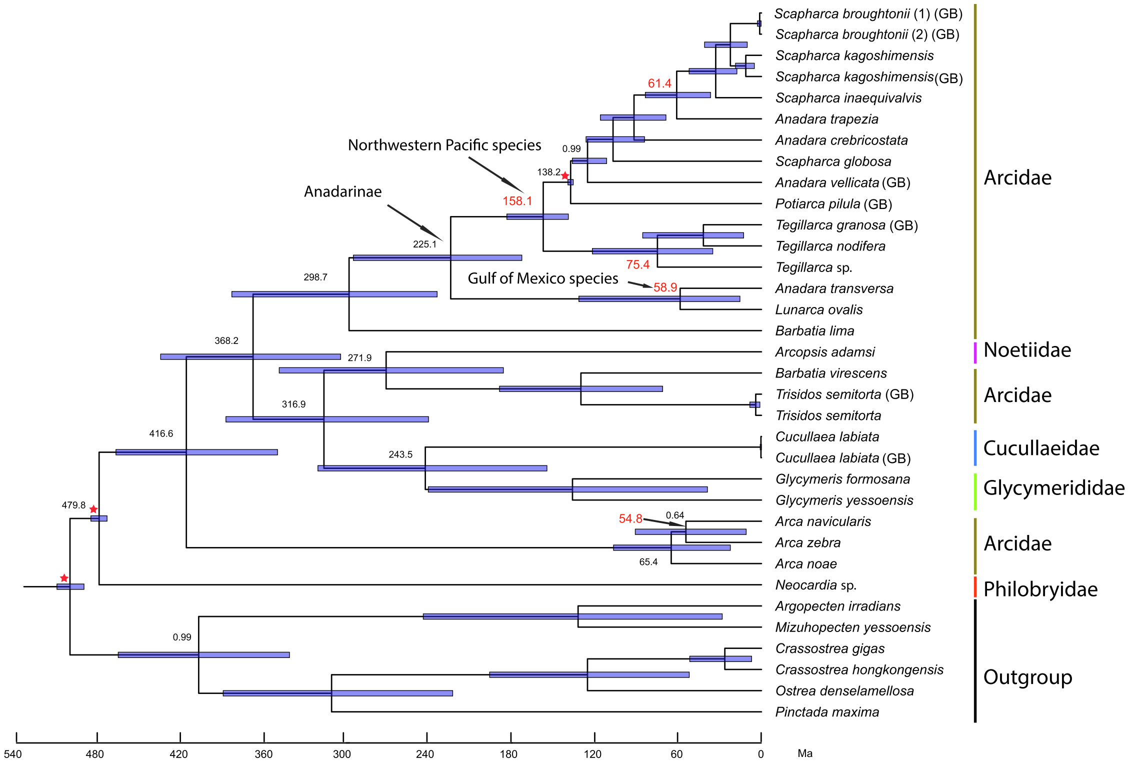


Fig. 6. Time calibrated Arcoida relationships inferred by Bayesian inference analyses with BEAST. Blue bars indicate 95% highest posterior density intervals (HPD). Posterior probabilities and median ages are shown adjacent to selected nodes of interest. Nodes that were used for calibration were marked with stars. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

tandem repeats ($\rho = 0.64$, $P = 0.002$; Fig. 3B). Indeed, the tandem repeats are potentially a main factor leading to the variation of intraspecific mitochondrial genome size. Comparing our newly sequenced mitochondrial genome of *Cucullaea labiata* to that of previous study (Feng et al., 2017), we found that the difference of mitochondrial genome size between two individuals was due to the number of copies of a 659 bp repeat (one copy vs. 8.9 copies). Similar results have been found in other bivalves, including the sea scallop *Placopecten magellanicus* whose mitochondrial genome ranges in size from 30.7 kb to 41 kb among different individuals, depending on the number of copies of a 1.4 kb repeat (Smith and Snyder, 2007). The mitochondrial genomes of *S. broughtonii* ranges in size from about 47 kb to ~ 50 kb because of variation in the number of tandem repeat arrays (Liu et al., 2013).

Thirdly, a systematic search detected inverted repeats (ranging from 1 to 83 copies) in all sampled arcoid species. Strong correlations between presence of inverted repeats and the genome size is observed ($\rho = 0.76$, $P = 9.84e\text{-}5$; Fig. 3C), which indicates the inverted repeats might be the major elements responsible for mitochondrial genome size variation, especially accounting for the interspecific mitochondrial genome size. Inverted repeats are not common in bilaterian mitochondrial genomes, but they have been found previously in fungi, algae, plants and demosponges (see Lavrov, 2010 and references therein). This rapid rate of proliferation combined with the broad phylogenetic distribution of hairpin elements was suggested as a defining force in the evolution of mitochondrial genomes of demosponges (Lavrov, 2010). However, the origin and propagation mechanism is still largely unknown. Lavrov (2010) suggested that the occurrence of stem-loop elements in the mitochondrial genome of the sponge *Lubomirskia baicalensis* may be linked to the activity of the Baikalum-1 transposon in its nuclear genome. Notably, we found numerous fragments of transposable elements in the non-coding region of arcoid mitochondrial genomes when they were blasted against the Repbase repetitive element database (Table S4). Previously, two transposon-like noncanonical open reading frames were also found to integrate into the mitochondrial genome of a deep-sea anemone *Bolocera* sp. (Zhang et al., 2017). Further studies are needed to test this association and the molecular mechanisms involved.

4.3. Genome composition and structure

Similar to other bivalves, arcoids show highly variable mitochondrial gene order, even within the same genus. All arcoid mitochondrial genomes sampled to date possess substantially different mitochondrial gene orders (Fig. 2). When tRNAs were not considered, there are still twelve different gene orders for 21 species. Coding on both strands has been assumed to restrain the gene arrangement of mitochondrial genomes, because rearranging a genome with dual-strand coding may be more complicated than if all genes are on the same strand (Ren et al., 2010). Perhaps coding on only one strand may be one of the factors behind the extensive rearrangement of bivalve mitochondrial genomes. Although variability in gene arrangement is high, gene order of protein-coding genes and ribosomal genes is generally conserved within each phylogenetic clade. Interestingly, *Tegillarca granosa/T. nodifera* clade share same mitochondrial gene order with *A. vellicata/S. globosa/A. crebricostata/S. inaequivalvis/S. kagoshimensis/S. broughtonii* clade, which might imply synapomorphy of identical PCG/rRNA clusters.

Six putative cases of gene duplication, especially in *cox2* and *rrn*, were observed in the 17 mitochondrial genome orders found in ark shells. By revisiting the annotation of mtDNA genomes of *Scapharca broughtonii* (AB729113, Liu et al., 2013), *A. sativa* (=*S. broughtonii*, KF667521, Hou et al., 2016) and *S. kagoshimensis* (KF750628, Sun et al., 2015a), we found the existence of a duplication of the *cox2* gene. Duplication of *cox2* was detected in other bivalve taxa such as *Meretrix*, *Loripes* and *Venerupis* within Heterodonta, hence duplicated *cox2* was considered to be synapomorphic of Heterodonta (Stöger and Schrödl, 2013). In addition, *cox2* is duplicated in tandem in the mitochondrial genomes of *Ruditapes philippinarum* and *Musculista senhousia*, two bivalves with doubly uniparental inheritance (DUI) of mtDNA (Breton et al., 2014). However, to our knowledge, the blood cockles do not have DUI, hence their duplicate *cox2* cannot be at all related to this mechanism and its functionality is still under study. Duplicated rrnS has been also reported as a synapomorphy for all *Crassostrea* species except *C. virginica*, which is the sister taxon to other *Crassostrea* species (Wu et al., 2010).

To the best of our knowledge, *C. labiata* is the only mollusc species to contain an intron in a mitochondrial protein-coding gene. Introns rarely occur in animal mitochondrial genomes. However, Group I introns have been found in Placozoa, Anthozoa (corals, soft corals, sea anemones) and Porifera (sponges), while Group II introns have been found in Placozoa and Annelida (Vallès et al., 2008; Huchon et al., 2015; Bernardino et al., 2017). Although ORFs were not found in the *cox1* introns of *C. labiata*, the 606 bp-intron region could possibly be derived from ancient transposable elements that have since lost any function (Bernardino et al., 2017). Notably, based on the result of Palindrome program from the EMBOSS package, 18 bp palindromic motifs were found in the intron. The palindrome is thought to provide the site for DNA-binding proteins involved in the transcriptional machinery (Arunkumar and Nagaraju, 2006). In addition, the conventional splicing signals (GT-AG) of a nuclear intron (Breathnach et al., 1978) was found at the *cox1* exon–intron boundaries by aligning mitochondrial genome of *C. labiata* with the GenBank published one (KP091889.1, Feng et al., 2017). Thus, we speculate that the GT-AG motif in the 5′ and 3′ end of the intron may represent a splicing signal involved the transcription process of the *cox1*.

Here, we show that mitochondrial genomes of some ark shells have large insertions of non-coding regions that are unusual for bilaterian mitochondrial genomes, leading several ark shell species to have some of the largest animal mitochondrial genomes observed to date (~50 kb). Moreover, our results suggest that the large variation of mitochondrial genome size are the result of multiple independent events. Tandem repeats likely facilitate mitochondrial genome size variation, and inverted repeats which could related to the transposition events might be responsible elements for interspecific mitochondrial genome expansions and contraction in ark shells.

Acknowledgements

We thank Dr. Rudiger Bieler from Field Museum of Natural History for kindly providing the sample of *Arcopsis adamsi*. Two anonymous reviewers provided feedback that helped to improve this article. This study was supported by awards from the National Natural Science Foundation of China (31772414, 41276138) and the Fundamental Research Funds for the Central Universities (201964001) to L.K., the Ocean University of China and Auburn University foundation/integrated grant program to K.M.H. and L.K. Lingfeng Kong was also supported by a one-year scholarship from the China Scholarship Council (CSC) as a visiting scholar at Auburn University. This is Molette Biology Laboratory contribution #102 and Auburn University Marine Biology Program contribution #201.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2020.106857.

References

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402. https://doi.org/10.1093/nar/25.17.3389.

Alverson, A.J., Wei, X., Rice, D.W., Stern, D.B., Barry, K., Palmer, J.D., 2010. Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). Mol. Biol. Evol. 27, 1436–1448. https://doi.org/10.1093/molbev/msq029.

Amler, M.R.W., 1999. Synoptical classification of fossil and recent Bivalvia. Geol. Palaeontol. 33, 237–248.

Arunkumar, K.P., Nagaraju, J., 2006. Unusually long palindromes are abundant in mitochondrial control regions of insects and nematodes. PLoS One 1, e110. https://doi.org/10.1371/journal.pone.0000110.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19, 455–477. https://doi.org/10.1089/cmb.2012.0021.

Benson, G., 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res 27, 573–580. https://doi.org/10.1093/nar/27.2.573.

Bernardino, A.F., Li, Y., Smith, C.R., Halanych, K.M., 2017. Multiple introns in a deep-sea Annelid (Decemunciger: Ampharetidae) mitochondrial genome. Sci. Rep. 7, 4295. https://doi.org/10.1038/s41598-017-04094-w.

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., Pütz, J., Middendorf, M., Stadler, P.F., 2013. MITOS: Improved de novo metazoan mitochondrial genome annotation. Mol. Phylogenet. Evol. 69, 313–319. https://doi.org/10.1016/j.ympev.2012.08.023.

Bieler, R., Mikkelsen, P.M., Collins, T.M., Glover, E.A., González, V.L., Graf, D.L., Harper, E.M., Healy, J., Kawauchi, G.Y., Sharma, P.P., Staubach, S., Strong, E.E., Taylor, J.D., Tëmkin, I., Zardus, J.D., Clark, S., Guzmán, A., McIntyre, E., Sharp, P., Giribet, G., 2014. Investigating the Bivalve Tree of Life - An exemplar-based approach combining molecular and novel morphological characters. Invertebr. Syst. 28, 32–115. https://doi.org/10.1071/IS13010.

Boisvert, S., Laviolette, F., Corbeil, J., 2010. Ray: simultaneous assembly of reads from a mix of high-throughput sequencing technologies. J. Comput. Biol. 17, 1519–1533. https://doi.org/10.1089/cmb.2009.0238.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Boore, J.L., 1999. Animal mitochondrial genomes. Nucleic Acids Res. 27, 1767–1780. https://doi.org/10.1093/nar/27.8.1767.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J., 2014. BEAST 2: A software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 10, 1–6. https://doi.org/10.1371/journal.pcbi.1003537.

Breathnach, R., Benoist, C., O'Hare, K., Gannon, F., Chambon, P., 1978. Ovalbumin gene: evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries. Proc. Natl. Acad. Sci. 75, 4853–4857. https://doi.org/10.1073/pnas.75.10.4853.

Breton, S., Milani, L., Ghiselli, F., Guerra, D., Stewart, D.T., Passamonti, M., 2014. A resourceful genome: updating the functional repertoire and evolutionary role of animal mitochondrial DNAs. Trends Genet. 30, 555–564. https://doi.org/10.1016/j.tig.2014.09.002.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: Architecture and applications. BMC Bioinformat. 10, 1–9. https://doi.org/10.1186/1471-2105-10-421.

Carver, T., Harris, S.R., Berriman, M., Parkhill, J., McQuillan, J.A., 2012. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. Bioinformatics 28, 464–469. https://doi.org/10.1093/bioinformatics/btr703.

Chinese National Department of Fisheries, 2017. China fisheries statistic yearbook 2016 (in Chinese). China Agriculture Press, Beijing.

Chong, R.A., Mueller, R.L., 2013. Low metabolic rates in salamanders are correlated with weak selective constraints on mitochondrial genes. Evolution (N. Y) 67, 894–899. https://doi.org/10.1111/j.1558-5646.2012.01830.x.

Combosch, D.J., Collins, T.M., Glover, E.A., Graf, D.L., Harper, E.M., Healy, J.M., Kawauchi, G.Y., Lemer, S., McIntyre, E., Strong, E.E., Taylor, J.D., Zardus, J.D., Mikkelsen, P.M., Giribet, G., Bieler, R., 2017. A family-level Tree of Life for bivalves based on a Sanger-sequencing approach. Mol. Phylogenet. Evol. 107, 191–208. https://doi.org/10.1016/j.ympev.2016.11.003.

Combosch, D.J., Giribet, G., 2016. Clarifying phylogenetic relationships and the evolutionary history of the bivalve order Arcida (Mollusca: Bivalvia: Pteriomorphia). Mol. Phylogenet. Evol. 94, 298–312. https://doi.org/10.1016/j.ympev.2015.09.016.

Cope, J.C.W., 1997. The early phylogeny of the class Bivalvia. Palaeontology 40, 713–746.

FAO, 2018. The State of World Fisheries and Aquaculture. Food and Agriculture Organization of the United Nations, Rome, Italy.

Feng, Y., Li, Q., Kong, L., 2015. Molecular phylogeny of Arcoidea with emphasis on Arcidae species (Bivalvia: Pteriomorphia) along the coast of China: Challenges to current classification of arcoids. Mol. Phylogenet. Evol. 85, 189–196. https://doi.org/10.1016/j.ympev.2015.02.006.

Feng, Y., Li, Q., Yu, H., Kong, L., 2017. Complete mitochondrial genome sequence of *Cucullaea labiata* (Arcoida: Cucullaeidae) and phylogenetic implications. Genes Genomics 39, 867–875. https://doi.org/10.1007/s13258-017-0548-1.

Giribet, G., Wheeler, W., 2005. On bivalve phylogeny: a high-level analysis of the Bivalvia (Mollusca) based on combined morphology and DNA sequence data. Invertebr. Biol. 121, 271–324. https://doi.org/10.1111/j.1744-7410.2002.tb00132.x.

Gonzalez, V.L., Andrade, S.C.S., Bieler, R., Collins, T.M., Dunn, C.W., Mikkelsen, P.M., Taylor, J.D., Giribet, G., 2015. A phylogenetic backbone for Bivalvia: an RNA-seq approach. Proc. R. Soc. B Biol. Sci. 282, 20142332. https://doi.org/10.1098/rspb.2014.2332.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., Di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29, 644–652. https://doi.org/10.1038/nbt.1883.

Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G., 2013. QUAST: Quality assessment tool for genome assemblies. Bioinformatics. 29, 1072–1075. https://doi.org/10.1093/bioinformatics/btt086.

Hou, Y., Wu, B., Liu, Z.H., Yang, A.G., Ren, J.F., Zhou, L.Q., Dong, C.G., Tian, J.T., 2016. Complete mitochondrial genome of Ark shell *Scapharca subcrenata*. Mitochondrial DNA 27, 939–940. https://doi.org/10.3109/19401736.2014.926495.

Huber, M., 2010. Compendium of bivalves. A full-color guide to 3,300 of the world marine bivalves. A status on Bivalvia after 250 years of research. Hackenheim: ConchBooks.

Huchon, D., Szitenberg, A., Shefer, S., Ilan, M., Feldstein, T., 2015. Mitochondrial group I and group II introns in the sponge orders Agelasida and Axinellida. BMC Evol. Biol. 15, 1–14. https://doi.org/10.1186/s12862-015-0556-1.

Jaccard, A., 1869. Description géologique du Jura Vaudois et Neuchatelois. Matériaux pour la Cart. géologique la. Suisse 6, 1–340.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. 30, 772–780. https://doi.org/10.1093/molbev/mst010.

Kohany, O., Gentles, A.J., Hankus, L., Jurka, J., 2006. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. BMC Bioinformatics 7, 1–7. https://doi.org/10.1186/1471-2105-7-474.

Kück, P., Meusemann, K., 2010. FASconCAT: Convenient handling of data matrices. Mol. Phylogenet. Evol. 56, 1115–1118. https://doi.org/10.1016/j.ympev.2010.04.024.

Lanfear, R., Calcott, B., Ho, S.Y.W., Guindon, S., 2012. PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. 29, 1695–1701. https://doi.org/10.1093/molbev/mss020.

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359. https://doi.org/10.1038/nmeth.1923.

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. https://doi.org/10.1093/molbev/msh112.

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. https://doi.org/10.1093/sysbio/syt022.

Lavrov, D.V., 2010. Rapid proliferation of repetitive palindromic elements in mtDNA of the endemic Baikalian sponge *Lubomirskia baicalensis*. Mol. Biol. Evol. 27, 757–760. https://doi.org/10.1093/molbev/msp317.

Lemer, S., González, V.L., Bieler, R., Giribet, G., 2016. Cementing mussels to oysters in the pteriomorphian tree: a phylogenomic approach. Proc. R. Soc. B Biol. Sci. 283, 20160857. https://doi.org/10.1098/rspb.2016.0857.

Li, G. yao, Zhang, L., Liu, J.Z., Chen, S.G., Xiao, T.W., Liu, G.Z., Wang, J.X., Wang, L.X., Hou, M., 2016. Marine drug Haishengsu increases chemosensitivity to conventional chemotherapy and improves quality of life in patients with acute leukemia. Biomed. Pharmacother. 81, 160–165. https://doi.org/10.1016/j.biopha.2016.04.005.

Li, Y., Kocot, K.M., Schander, C., Santos, S.R., Thornhill, D.J., Halanych, K.M., 2015. Mitogenomics reveals phylogeny and repeated motifs in control regions of the deep-sea family Siboglinidae (Annelida). Mol. Phylogenet. Evol. 85, 221–229. https://doi.org/10.1016/j.ympev.2015.02.008.

Liu, Y.G., Kurokawa, T., Sekino, M., Tanabe, T., Watanabe, K., 2013. Complete mitochondrial DNA sequence of the ark shell *Scapharca broughtonii*: An ultra-large metazoan mitochondrial genome. Comp. Biochem. Physiol. - Part D Genomics Proteomics 8, 72–81. https://doi.org/10.1016/j.cbd.2012.12.003.

Marko, P.B., 2002. Fossil calibration of molecular clocks and the divergence times of geminate species pairs separated by the Isthmus of Panama. Mol. Biol. Evol. 19, 2005–2021. https://doi.org/10.1093/oxfordjournals.molbev.a004024.

Matsumoto, M., 2003. Phylogenetic analysis of the subclass Pteriomorphia (Bivalvia) from mtDNA COI sequences. Mol. Phylogenet. Evol. 27, 429–440. https://doi.org/10.1016/S1055-7903(03)00013-7.

Mikkelsen, N.T., Kocot, K.M., Halanych, K.M., 2018. Mitogenomics reveals phylogenetic relationships of caudofoveate aplacophoran molluscs. Mol. Phylogenet. Evol. 127, 429–436. https://doi.org/10.1016/j.ympev.2018.04.031.

Miya, M., Kawaguchi, A., Nishida, M., 2001. Mitogenomic exploration of higher teleostean phylogenies: A case study for moderate-scale evolutionary genomics with 38 newly determined complete mitochondrial DNA sequences. Mol. Biol. Evol. 18, 1993–2009. https://doi.org/10.1093/oxfordjournals.molbev.a003741.

Morton, B., Puljas, S., 2016. The ectopic compound ommatidium-like pallial eyes of three species of Mediterranean (Adriatic Sea) *Glycymeris* (Bivalvia: Arcoida). Decreasing visual acuity with increasing depth? Acta Zool. 97, 464–474. https://doi.org/10.1111/azo.12140.

Nakamura, Y., 2005. Suspension feeding of the ark shell *Scapharca subcrenata* as a function of environmental and biological variables. Fish. Sci. 71, 875–883. https://doi.org/10.1111/j.1444-2906.2005.01040.x.

Newell, N.D., 1969. Classification of Bivalvia. In: In: Moore, R. (Ed.), Treatise on Invertebrate Paleontology, Part N, Mollusca 6, vol. 1. Bivalvia. Geological Society of America & University of Kansas, Boulder-Lawrence, pp. N205–N244.

Nguyen, L.T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32, 268–274. https://doi.org/10.1093/molbev/msu300.

Ojala, D., Montoya, J., Attardi, G., 1981. TRNA punctuation model of RNA processing in human mitochondria. Nature 290, 470–474. https://doi.org/10.1038/290470a0.

Oliver, P.G., Holmes, A.M., 2006. The Arcoidea (Mollusca: Bivalvia): A review of the current phenetic-based systematics. Zool. J. Linn. Soc. 148, 237–251. https://doi.org/10.1111/j.1096-3642.2006.00256.x.

Osigus, H.J., Eitel, M., Bernt, M., Donath, A., Schierwater, B., 2013. Mitogenomics at the base of Metazoa. Mol. Phylogenet. Evol. 69, 339–351. https://doi.org/10.1016/j.ympev.2013.07.016.

Paradis, E., Claude, J., Strimmer, K., 2004. APE: Analyses of phylogenetics and evolution in R language. Bioinformatics 20, 289–290. https://doi.org/10.1093/bioinformatics/btg412.

Development Core Team, R., 2018. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna (Austria).

Rand, D.M., 1993. Endotherms, ectotherms, and mitochondrial genome-size variation. J. Mol. Evol. 37, 281–295. https://doi.org/10.1007/BF00175505.

Ren, J., Liu, X., Jiang, F., Guo, X., Liu, B., 2010. Unusual conservation of mitochondrial gene order in *Crassostrea* oysters: evidence for recent speciation in Asia. BMC Evol. Biol. 10, 394. https://doi.org/10.1186/1471-2148-10-394.

Revell, L.J., 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. 3, 217–223. https://doi.org/10.1111/j.2041-210X.2011.00169.x.

Rice, P., Longden, I., Bleasby, A., 2000. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. 16, 276–277. https://doi.org/10.1016/S0168-9525(00)02024-2.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., Mesirov, J.P., 2011. Integrative genomics viewer. Nat. Biotechnol. 29, 24–26. https://doi.org/10.1038/nbt.1754.

Ruby, J. Graham, Bellare, Priya, DeRisi, Joseph L., 2013. PRICE: Software for the Targeted Assembly of Components of (Meta) Genomic Sequence Data. G3 3 (5), 865–880. https://doi.org/10.1534/g3.113.005967.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A., Barrell, B., 2000. Artemis: sequence visualization and annotation. Bioinformatics 16, 944–945. https://doi.org/10.1093/bioinformatics/16.10.944.

Simison, W.B., Lindberg, D.R., Boore, J.L., 2006. Rolling circle amplification of metazoan mitochondrial genomes. Mol. Phylogenet. Evol. 39, 562–567. https://doi.org/10.1016/j.ympev.2005.11.006.

Smith, D.R., Snyder, M., 2007. Complete mitochondrial DNA sequence of the scallop *Placopecten magellanicus*: Evidence of transposition leading to an uncharacteristically large mitochondrial genome. J. Mol. Evol. 65, 380–391. https://doi.org/10.1007/s00239-007-9016-x.

Steiner, G., Hammer, S., 2000. Molecular phylogeny of the Bivalvia inferred from 18S rDNA sequences with particular reference to the Pteriomorphia. Geol. Soc. London Spec. Publ. 177, 11–29. https://doi.org/10.1144/GSL.SP.2000.177.01.02.

Stöger, I., Schrödl, M., 2013. Mitogenomics does not resolve deep molluscan relationships (yet?). Mol. Phylogenet. Evol. 69, 376–392. https://doi.org/10.1016/j.ympev.2012.11.017.

Strotz, L.C., Saupe, E.E., Kimmig, J., Lieberman, B.S., 2018. Metabolic rates, climate and macroevolution: a case study using Neogene molluscs. Proc. R. Soc. B Biol. Sci. 285, 20181292. https://doi.org/10.1098/rspb.2018.1292.

Sturmer, L., Nunez, J., Creswell, R., Baker, S., 2009. The potential of blood ark and ponderous ark aquaculture in Flordia: results of spawning, larval rearing, nursery and growout trials. SeaGrant, Florida TP-169. Shellfish Aquaculture. Extension Program, University of Florida.

Sun, S., Kong, L., Yu, H., Li, Q., 2015a. The complete mitochondrial genome of *Scapharca kagoshimensis* (Bivalvia: Arcidae). Mitochondrial DNA 26, 957–958. https://doi.org/10.3109/19401736.2013.865174.

Sun, S., Kong, L., Yu, H., Li, Q., 2015b. The complete mitochondrial DNA of *Tegillarca granosa* and comparative mitogenomic analyses of three Arcidae species. Gene 557, 61–70. https://doi.org/10.1016/j.gene.2014.12.011.

Sun, S., Kong, L., Yu, H., Li, Q., 2015c. Complete mitochondrial genome of *Anadara vellicata* (Bivalvia: Arcidae): A unique gene order and large atypical non-coding region. Comp. Biochem. Physiol. - Part D Genomics Proteomics 16, 73–82. https://doi.org/10.1016/j.cbd.2015.08.001.

Sun, S., Li, Q., Kong, L., Yu, H., 2016. Complete mitochondrial genomes of *Trisidos kiyoni* and *Potiarca pilula*: Varied mitochondrial genome size and highly rearranged gene order in Arcidae. Sci. Rep. 6, 33794. https://doi.org/10.1038/srep33794.

Sun, S., Li, Q., Kong, L., Yu, H., 2017. Limited locomotive ability relaxed selective constraints on molluscs mitochondrial genomes. Sci. Rep. 7, 10628. https://doi.org/10.1038/s41598-017-11117-z.

Talavera, G., Castresana, J., 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst. Biol. 56, 564–577. https://doi.org/10.1080/10635150701472164.

Vallès, Y., Halanych, K.M., Boore, J.L., 2008. Group II introns break new boundaries: Presence in a bilaterian's genome. PLoS One 3, e1488. https://doi.org/10.1371/journal.pone.0001488.

Williams, S.T., Foster, P.G., Hughes, C., Harper, E.M., Taylor, J.D., Littlewood, D.T.J., Dyal, P., Hopkins, K.P., Briscoe, A.G., 2017. Curious bivalves: Systematic utility and unusual properties of anomalodesmatan mitochondrial genomes. Mol. Phylogenet. Evol. 110, 60–72. https://doi.org/10.1016/j.ympev.2017.03.004.

Wu, X., Xu, X., Yu, Z., Kong, X., 2009. Comparative mitogenomic analyses of three scallops (Bivalvia: Pectinidae) reveal high level variation of genomic organization and a diversity of transfer RNA gene sets. BMC Res. Notes 2, 1–7. https://doi.org/10.1186/1756-0500-2-69.

Wu, X., Xu, X., Yu, Z., Wei, Z., Xia, J., 2010. Comparison of seven *Crassostrea* mitogenomes and phylogenetic analyses. Mol. Phylogenet. Evol. 57, 448–454. https://doi.org/10.1016/j.ympev.2010.05.029.

Zhang, B., Zhang, Y.H., Wang, X., Zhang, H.X., Lin, Q., 2017. The mitochondrial genome of a sea anemone *Bolocera* sp. exhibits novel genetic structures potentially involved in adaptation to the deep-sea environment. Ecol. Evol. 7, 4951–4962. https://doi.org/10.1002/ece3.3067.