Gene 557 (2015) 61–70

Contents lists available at ScienceDirect

Gene

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

The complete mitochondrial DNA of *Tegillarca granosa* and comparative mitogenomic analyses of three Arcidae species

Shao'e Sun, Lingfeng Kong *, Hong Yu, Qi Li

Key Laboratory of Mariculture, Ministry of Education, Ocean University of China, Qingdao 266003, China

ARTICLE INFO

Article history:
Received 25 June 2014
Received in revised form 5 December 2014
Accepted 6 December 2014
Available online 9 December 2014

Keywords:
Arcidae
Mitochondrial genome
Comparative analysis
Phylogenetic relationships

ABSTRACT

To better understand the characteristics and the evolutionary dynamics of mt genomes in Arcidae, the complete mitochondrial genome of *Tegillarca granosa* was firstly determined and compared with other two Arcidae species (*Scapharca broughtonii* and *Scapharca kagoshimensis*). The complete mitochondrial genome of *T. granosa* was 31,589 bp in length, including 12 protein-coding genes, 2 rRNA genes and 23 tRNA genes, and a major non-coding region. Three tandem repeat fragments were identified in the major non-coding region and the tandem repeat motifs of these fragments can be folded into stem–loop structures. The mitochondrial genome of the three species has several common features such as the AT content, the arrangement of the protein-coding genes, the codon usage of the protein-coding genes and AT/GC skew. However, a high level of variability is presented in the size of the genome, the number of tRNA genes and the length of non-coding sequences in the three mitogenomes. According to the phylogenetic analyses, these mitogenome-level characters are correlated with their phylogenetic relationships. It is the absence of the duplicated tRNAs and large non-coding sequences that are responsible for the length divergence of mitogenomes between *T. granosa* and other two Arcidae species. The phylogenetic analyses were conducted based on 12 partitioned protein genes, which support the relationship at the family level: (((Pectinidae + Ostreidae) + Mytilidae) + Arcidae).

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

The typical metazoan mitochondrial (mt) genome is a covalently closed circular molecule ranging in size from 14 to 20 kb, encoding 37 genes: 13 protein genes, 22 transfer RNAs and two ribosomal RNAs (Boore, 1999). However, as more and more animal mt genome sequences are determined, an increasing number of animal mtDNAs seem to deviate from this model substantially, and many of them are much larger than 20 kb. At present, many studies have found atypically large mtDNA chains of varying size in the newts (Wallis, 1986), fishes (Gach and Brown, 1997; Richardson and Gold, 1991), lizards (Moritz and Brown, 1986), insects (Boyce et al., 1989) and nematode *Romanomermis culicivorax* (Hyman et al., 1988; Hyman and Slater, 1990; Powers et al., 1986). These molecules gained their large size in various ways. Tandem duplications of control region sequences were found to be the main reason of size increases in mitogenomes of newts and fishes, whereas the duplications of coding sequences are responsible for the large size of lizard mitogenomes. The large size of mtDNA of the nematode, *R. culicivorax*, is due to many dispersed repeats in the molecular. In bivalves, the mitochondrial genomes of relatively large size (greater than 20 kb) have been found in the deep sea scallop *Placopecten magellanicus* (up to 40,725 bp) (Smith and Snyder, 2007) and the Zhikong scallop *Chlamys farreri* (21,695 bp) (Xu et al., 2011). The repeated sequences, transposition involving tRNAs or tRNA-like sequences and duplications in the non-coding regions are responsible for the large size of these mitogenomes (Liu et al., 2013; Smith and Snyder, 2007). It was also observed that the mitogenome of Manila clam *Venerupis philippinarum* (AB065375) is larger than 20 kb. A peculiar way of mitochondrial inheritance, doubly uniparental inheritance (DUI), has been found in *V. philippinarum*, which may also influence the genome size in M and F mitochondrial DNA (Passamonti and Scali, 2001; Breton et al., 2007; Ghiselli et al., 2013; Zouros, 2013). The female type is 22,676 bp (GenBank accession number AB065375) and the male type is 21,441 bp (GenBank accession number AB065374).

The Arcidae have a high species richness in the tropical shallow waters and warm temperate seas, comprising about 260 species (Oliver and Holmes, 2006). Two complete mitochondrial genomes were recently made available for this family: *Scapharca broughtonii* (AB729113) and *Scapharca kagoshimensis* (KF750628). Interestingly, the mt genomes of these two Arcidae species deviate the most dramatically from the size of typical metazoan mtDNAs, with a length of 46,985 bp (Liu et al.,

Abbreviations: atp6, ATPase subunit 6 genes; *Cytb*, cytochrome b gene; *cox1–3*, cytochrome *c* oxidase subunits I–III genes; NCR, non-coding region; *nad1–6* and *nad4L*, NADH dehydrogenase subunits 1–6 and 4L genes; ML, maximum likelihood; rRNA, ribosomal RNA; *rrnL* and *rrnS*, large and small subunits of ribosomal RNA genes; tRNA, transfer RNA; PCG, protein coding gene; K2P, Kimura two-parameter distance; TDRL, tandem duplication and random loss; mtDNA, mitochondrial DNA; mt genome, mitochondrial genome; MNR, major non-coding region; Sb, *Scapharca broughtonii*; Sk, *Scapharca kagoshimensis*; Tg, *Tegillarca granosa*.

* Corresponding author.
E-mail address: klfaly@ouc.edu.cn (L. Kong).

http://dx.doi.org/10.1016/j.gene.2014.12.011
0378-1119/© 2014 Elsevier B.V. All rights reserved.

Table 1
Partitioned data for phylogenetic analyses and DNA substitution models applied to each (Bayesian analyses only, RAxML only applies GTR + G for any partition).

Partition	Alignment length (bp)	DNA substitution models
cox1	1373	GTR + I + G
nad5	814	GTR + I + G
nad4	674	GTR + I + G
nad1	646	GTR + I + G
Cytb	860	GTR + I + G
cox2	459	GTR + I + G
nad6	252	GTR + I + G
atp6	375	HKY + G
nad3	108	HKY + G
nad4L	203	GTR + I + G
cox3	437	HKY + G
nad2	281	GTR + I + G

2013) and 46,713 bp (Sun et al., 2014), respectively. Several special features have been evidenced in the two mitochondrial genomes: (i) *S. broughtonii* and *S. kagoshimensis* have the largest mtDNAs in all metazoan mt genomes sequenced to date; (ii) more than 40 tRNA genes are identified from the two mt genomes, being the largest number of tRNA discovered in metazoan mtDNAs; and (iii) the percentage of non-coding sequences of these two mtDNAs is much larger than that of other metazoan mtDNAs. The sequence duplications in the non-coding regions have been considered one of the factors responsible for the large size of the two mt genomes. Considering the unusual features of the two mt genomes in Arcidae, we attempt to validate whether these characteristics are shared by other species of Arcidae. In the present study we have determined the third complete mt genome of the Arcidae species, *Tegillarca granosa*, and compared it with those of *S. broughtonii* and *S. kagoshimensis*. This species is phylogenetically close to *S. broughtonii* and *S. kagoshimensis* (Oliver and Holmes, 2006; Matsumoto, 2003). The investigation of *T. granosa* mitochondrial genome supplies the gene feature requirements for the comparative analysis of mitochondrial genome of Arcidae. Moreover, this study provided effective information to understand the characteristics and the evolutionary scenarios of mt genomes in Arcidae, explaining the cause of the length differences among Arcidae mt genomes based on the genomic structure. Additionally, in this study, we inferred the phylogenetic topologies with the concatenated nucleotide sequences of 12 protein-coding gene datasets to get insights into the phylogenetic relationship among Arcidae species and the phylogenetic position of Arcidae within Bivalvia.

2. Materials and methods

2.1. Specimens, DNA extraction, PCR amplification and sequencing

Two live *T. granosa* individuals were collected from the coastal water of Wenzhou, Zhejiang Province, China. Total genomic DNA of *T. granosa* was extracted from adductor muscle by a modification of standard phenol–chloroform procedure as described by Li et al. (2002). Short fragments of the genes *cox1* were amplified by PCR with primers LCO-1490/HCO-2198 (Folmer et al., 1994). Other short fragments, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *cob* and *cox2* were obtained from ESTs of *T. granosa* (Bao and Lin, 2010). Based on the obtained partial sequences, long-PCR primers were designed to amplify the entire mitochondrial genomes in five long PCR reactions (Supplementary Table 1). PCR were carried out in 50 μl reaction mixtures containing 8 μl 2.5 mM dNTPs, 5 μl 10× reaction buffer (Mg^{2+} plus), 1 μl genomic DNA, 1 μl 10 μM each primer and 0.5 μl LA-Taq polymerase. PCR procedures for all the long fragments were: 94 °C for 3 min followed by 35 cycles of 94 °C for 30 s, 62 °C for 30 s and 68 °C for 10 min. A final extension step of 72 °C for 10 min was added. PCR products were purified using EZ-10 spin column DNA gel extraction kit (Sangon Biotech), and then directly sequenced with the primer walking method. The sequencing was conducted on an ABI PRISM 3730 (Applied Biosystems) automatic sequencer.

After sequencing the regions between *nd4l* and *cox3*, we used the genome-walking based method to obtain the whole mitogenome sequence. The primary PCR procedures were performed following the methods described in the BD GenomeWalker Universal Kit User Manual (BD Biosciences). The primer sequences used for genome-walking are presented in Supplementary Table 2. All the PCR amplicons were sequenced using the methods described above.

2.2. Analysis of sequence data and gene annotation

All sequence data were analyzed and arranged to create the full genome using the Seqman program from DNASTAR (http://www.DNASTAR.com). Manual examinations were applied to ensure correct assembly. To define protein coding genes, ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html), BLASTx (http://blast.st-va.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastx&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome) and HMMER web server (Finn et al., 2011; Eddy, 2011) were used. Signal peptides (SPs) were identified using InterProScan 5 (http://www.ebi.ac.uk/Tools/pfa/iprscan5/, Zdobnov and Apweiler, 2001) and Phobius (http://phobius.sbc.su.se/, Käll et al., 2004). TMpred (http://www.ch.embnet.org/software/TMPRED_form.html, Hofmann and Stoffel, 1993) and Phobius (http://phobius.sbc.su.se/, Käll et al., 2004) were used to localize putative transmembrane helices (TM-helices). @TOME 2 (http://atome.cbs.cnrs.fr/AT2B/meta.html, Pons and Labesse, 2009) was used to find similarities with known proteins. The tandem repeat sequences were searched by Tandem Repeats Finder 4.0 (Benson, 1999). Prediction of potential secondary structure was performed by the online version of the mfold software, version 3.2 (Zuker, 2003). The tRNA genes were identified by tRNA-scan SE Search Server (Lowe and Eddy, 1997) and ARWEN (Laslett and Canback, 2008), using the mito/chloroplast genetic code and the default search mode. Codon usage analysis was performed by MEGA 5 (Tamura et al., 2011). The A + T content values and nucleotide frequencies were computed using Editseq program from DNASTAR. The GC and AT skews

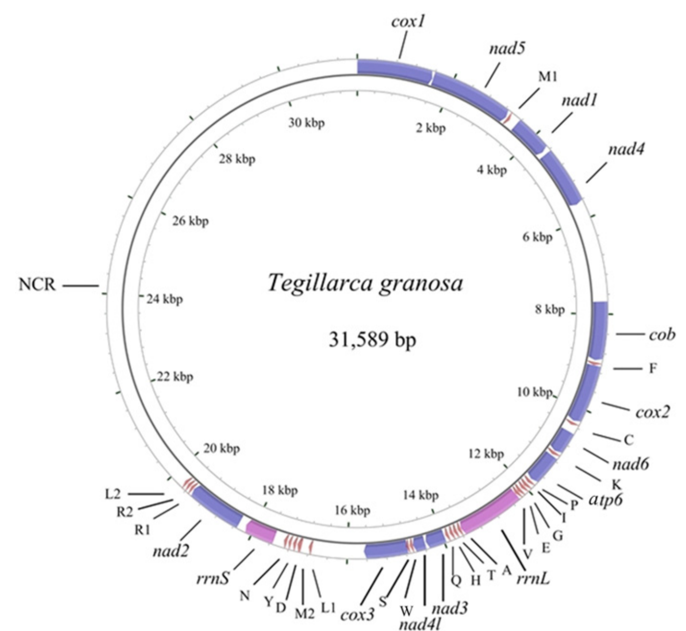

Fig. 1. Gene map of the mitochondrial genome of *Tegillarca granosa*. All the 37 genes transcribed from the plus strand. The largest non-coding region is designated as MNR.

Table 2
Organization of the mitochondrial genome of *Tegillarca granosa*.

Gene	Location	Size (bp)	Start codon	Stop codon	Intergenic nucleotides
cox1	1–1578	1578	ATG	TAG	15
nad5	1594–3309	1716	ATA	TAA	27
trnM1	3337–3403	67			63
nad1	3567–4361	795	ATA	TAG	71
nad4	4433–5659	1227	ATG	TAA	2083
Cytb	7743–8963	1221	ATA	TAG	40
trnF	9004–9069	66			1
cox2	9071–10,297	1227	ATG	TAA	32
trnC	10,330–10,393	64			146
nad6	10,540–11,031	492	ATG	TAA	18
trnK	11,050–11,120	71			24
atp6	11,145–11,792	648	ATG	TAG	34
trnP	11,827–11,893	65			28
trnI	11,922–11,992	71			10
trnG	12,003–12,071	69			25
trnE	12,097–12,167	71			2
trnV	12,170–12,237	68			2
rrnL	12,240–13,582	1343			4
trnA	13,587–13,654	68			23
trnT	13,678–13,748	71			17
trnH	13,766–13,834	69			32
trnQ	13,867–13,933	67			27
nad3	13,961–14,362	402	ATG	TAG	22
nad4L	14,385–14,636	252	ATA	TAG	0
trnW	14,637–14,702	66			13
trnS	14,716–14,777	62			−19
cox3	14,757–15,659	903	ATA	TAA	76
trnL	16,736–16,805	70			148
trnM2	16,954–17,046	93			23
trnD	17,070–17,138	69			37
trnY	17,176–17,239	64			26
trnN	17,266–17,335	70			241
rrnS	17,577–18,217	641			107
nad2	18,325–19,494	1170	ATG	TAG	14
trnR1	19,509–19,574	66			16
trnR2	19,591–19,657	67			17
trnL	19,675–19,741	67			10,848

were calculated according to the formulae by Perna and Kocher (1995): AT skew = (A − T) / (A + T); GC skew = (G − C) / (G + C). The mitochondrial genome was initially generated with the program CGView (Stothard and Wishart, 2005). The complete mtDNA sequence of *T. granosa* was deposited in the GenBank database under accession number (KJ607173).

2.3. Phylogenetic analyses

A total of 16 taxa were selected for phylogenetic analysis, including the complete mt sequences of *T. granosa* (this study). The representative species we chose for each taxonomic group were: *S. broughtonii* (AB729113) and *S. kagoshimensis* (KF750628) (Arcidae), *Crassostrea virginica* (AY905542), *Crassostrea iredalei* (FJ841967) and *Saccostrea mordax* (FJ841968) (Ostreidae), *Mizuhopecten yessoensis* (AB271769), *C. farreri* (EU715252), *Argopecten irradians* (EU023915), *Mimachlamys nobilis* (FJ415225) and *P. magellanicus* (DQ088274) (Pectinidae), *Mytilus edulis* (AY484747), *Mytilus galloprovincialis* (AY497292), *Mytilus trossulus* (GU936625) and *Mytilus californianus* (GQ527172) (Mytilidae). *Hyriopsis cumingii* (HM347668) from the subclass Palaeoheterodonta was used as outgroup.

The twelve-partitioned nucleotide sequences of protein coding genes were aligned with MAFFT (Katoh et al., 2005). Areas of dubious alignment were isolated using Gblocks (Castresana, 2000; Talavera and Castresana, 2007) (default setting) and excluded from the analyses. The best-fit nucleotide substitution models for each data partitions were selected by jModelTest (Posada, 2008). Because MrBayes 3.1 (Ronquist and Huelsenbeck, 2003) only implements 1, 2, and 6 substitution rate models, it was often not possible to implement the preferred model as selected by the AIC, and we consequently choose nearest parameterized models. We employed ML in RAxML Black-Box webserver (http://phylobench.vital-it.ch/raxml-bb/index.php; Stamatakis et al., 2008) with GTR + G substitution model to each partition. The information of alignment length and DNA substitution models applied to each partition (Bayesian analyses only) were listed in Table 1. The partitioned analyses were run, implementing separate nucleotide substitution models for each partition.

The phylogenetic relationship was built by two methods: maximum likelihood (ML) and Bayesian inference (BI). For the RaxML analyses, the dataset was partitioned and bootstrapped with 100 replicates. BI was performed using MrBayes 3.1.2. The Bayesian analyses were run with default priors and unlinked partitions, allowing each partition to have its own set of parameters. All partitioned analyses accommodated among-partition rate variation by using the ratepr parameter of the prset command (prset ratepr = variable) in MrBayes. The Markov chain Monte Carlo (MCMC) was run for 2,000,000 generations (sampling every 100 generations) to allow adequate time for convergence. Parameter convergence was achieved within two million generations and the standard deviation of split frequencies was less than 0.01. All parameters were checked with Tracer v 1.5 (Drummond and Rambaut, 2007). After omitting the first 10,000 "burn-in" trees, the

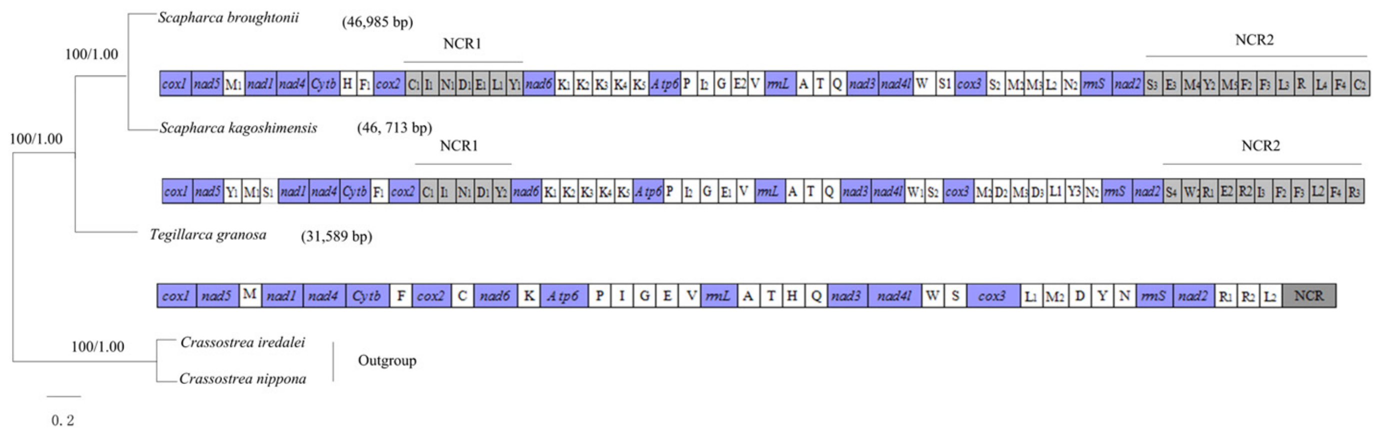

Fig. 2. Phylogenetic trees of the three Arcidae species based on nucleotide sequences of 12 mitochondrial protein-coding genes using Maximum Likelihood and Bayesian analyses. The first number at each node is the bootstrap probability of ML analyses and the second number is Bayesian posterior probability. Linear mitochondrial genomic organizations are shown under each species name. Each protein-coding gene (PCG), rRNA genes and tRNA genes are depicted, and major non-coding regions are marked.

Table 3
List of total size, AT content, AT and GC skew for *S. broughtonii*, *S. kagoshimensis* and *T. granosa*, with length of genes and NCR.

Gene	*S. broughtonii*	*S. kagoshimensis*	*T. granosa*
Total size	46,985	46,713	31,589
%AT	67.89	62.75	60.18
AT skew	−0.1657	−0.0990	−0.1353
GC skew	0.3574	0.1066	0.4127
NCR	31,658	32,982	16,394
rrnL	1324	1332	1343
rrnS	673	675	641
cox1	1584(ATG/TAG)	1575(ATA/TAA)	1578(ATG/TAG)
nad5	1707(ATA/TAA)	1002(ATG/TAA)	1716(ATA/TAA)
nad1	924(ATA/TAG)	747(ATA/TAG)	795(ATA/TAG)
nad4	1296(ATA/TAA)	804(ATA/TAA)	1227(ATG/TAA)
Cytb	1278(GTG/TAA)	1080(ATA/TAA)	1221(ATA/TAG)
cox2	666(ATG/TAA)	666(ATG/TAG)	1227(ATG/TAA)
nad6	468(ATA/TAA)	408(ATG/TAG)	492(ATG/TAA)
atp6	525(ATA/TAA)	675(ATG/TAA)	648(ATG/TAG)
nad3	360(ATA/TAA)	246(ATG/TAG)	402(ATG/TAG)
nad4L	285(ATG/TAA)	252(ATA/TAA)	252(ATA/TAG)
cox3	759(GTG/TAA)	537(ATG/TAA)	903(ATA/TAA)
nad2	567(ATA/TAG)	801(ATA/TAA)	1170(ATG/TAG)

remaining 10,000 sampled trees were used to estimate the 50% majority rule consensus trees and the Bayesian posterior probabilities.

3. Result and discussion

3.1. Genome organization and rearrangement in the three Arcidaes

The mt genome of *T. granosa* contains 12 protein-coding genes, two rRNAs, 23 tRNAs and non-coding regions, with all the 37 genes transcribed from the plus strand (Fig. 1, Table 2). The arrangement of the 12 protein-coding genes in *T. granosa* is identical to that of *S. broughtonii* and *S. kagoshimensis*, in the order *cox1–nad5–nad1–nad4–cob–cox2–nad6–atp6–nad3–nad4L–cox3–nad2*, corresponding evidently to their closer genetic relationships. However, the location and the number of the tRNA genes of *T. granosa* are different from those of the other two Arcidae genomes (Fig. 2). Mt-gene order appears to be dramatically variable in the major groups of bivalves. Pectinidae species seem to be a good example to prove this. Comparing gene orders of *C. farreri*, *A. irradians*, *M. nobilis* and *P. magellanicus*, even after excluding the tRNA genes from the comparison, the four mt genomes still show no identical gene arrangement (Xu et al., 2011). In bivalves, it is also common that species belonging to the same genus have different gene orders. For example, in *Crassostrea* congeners *C. virginica* and the six Asian *Crassostrea*, only protein-coding gene is arranged in an identical order, but tRNAs are extensively rearranged (Wu et al., 2010). The relatively conserved gene order in the three Arcidaes suggests that the divergence of the species may have occurred recently. The complete mitochondrial genome of *T. granosa* is 31,589 bp in length, and is far less than those of the other two Arcidae species, *S. broughtonii* (46,985 bp) and *S. kagoshimensis* (46,713 bp) (Table 3). However, it is still an uncharacteristically large mitochondrial genome. Such differences may be mainly resulted from the variable number and size of non-coding regions (NCRs). The length of the non-coding regions found in the mt genome of *T. granosa* (16,394 bp) is markedly shorter than that of *S. broughtonii* (31,658 bp) and *S. kagoshimensis* (32,982 bp). Furthermore, in the mt genomes of *S. broughtonii* and *S. kagoshimensis*, most of the non-coding DNAs were observed within two distinct zones. One lies between *cox2* and *nad6*, and the other lies between *nad2* and *cox1*. Unlike *S. broughtonii* and *S. kagoshimensis*, *T. granosa* has only one concentrated non-coding region, located between *nad2* and *cox1*, the same position of one of the two non-coding segments in *S. broughtonii* and *S. kagoshimensis* (Fig. 2). Even so, some protein-coding genes in *T. granosa*, such as *cox2*, *nad3*, *cox3* and *nad2*, are much longer than in *Scapharca* species (Table 3). There are 3877 codons for the 12 protein-coding genes of *T. granosa* with the stop codons excluded, compared with 3473 and 2931 for *S. broughtonii* and *S. kagoshimensis*. We attempt to explore whether the divergence of the mitogenomes among Arcidae species is phylogenetically based or owing to species-specific mitogenome maintenance mechanisms (Xu et al., 2012). According to the phylogenetic framework, we argue that genome reorganization among these congeneric species is not random, but correlated with their phylogenetic relationships (Fig. 2).

The base compositions and the AT contents of the mitogenomes of the three Arcidae species have been shown in Table 3. The overall AT content of *T. granosa* mt genome is 60.18%, the lowest among the three Arcidae species (*S. broughtonii*, 67.89% and *S. kagoshimensis*, 62.75%). The nucleotide compositions are all strongly skewed away from C in favor of G (the GC-skews are from 0.107 to 0.413) and from A in favor of T (the AT-skews are from −0.166 to −0.100).

3.2. Protein-coding genes

In total, 12 protein-coding genes (PCGs) were identified in the mitogenome of *T. granosa*. The *atp8* gene was not identified, as apparently is the case of most marine bivalve species studied so far (Yu et al., 2008). However, several exceptions have been found. For example, *Mimachlamys senatoria* (Wu et al., 2013), *Meretrix lusoria* (Wang et al., 2010), and *Hiatella arctica* (Dreyer and Steiner, 2006) show an *atp8* gene in their mtDNA. All these 12 PCGs start with conventional invertebrate start codons, with 7 PCGs using ATG and the other 5 PCGs employing ATA. Compared the initiation codon usage of the three Arcidaes, we found that almost all PCGs of the Arcidae mitogenomes initiated with ATG and ATA, and only one alternate initiation codon, GTG, which was detected in *cox3* gene of *S. broughtonii*. The alternate initiation codon GTG was also discovered in mitochondrial genomes of other bivalves, such as *Paphia euglypta* (*nad4L* and *cox3*: GTG) (Xu et al., 2012) and *Solecurtus divaricatus* (*nad4L* and *Cytb*: GTG) (Yuan et al., 2012). Two PCGs contain the same initiation codons in all three

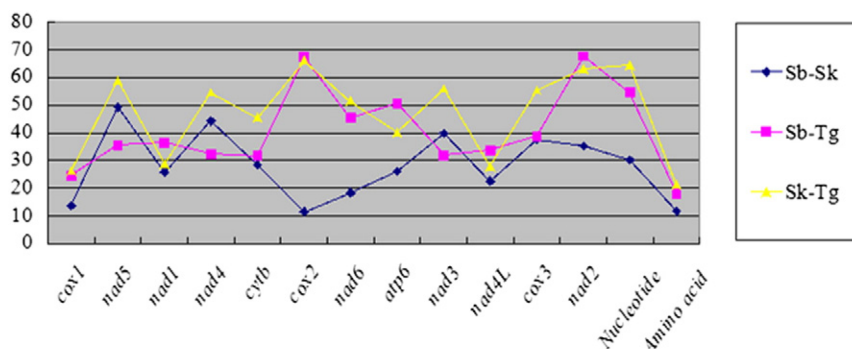

Fig. 3. Pairwise genetic divergence among the three Arcidaes based on separate and concatenated protein coding genes.

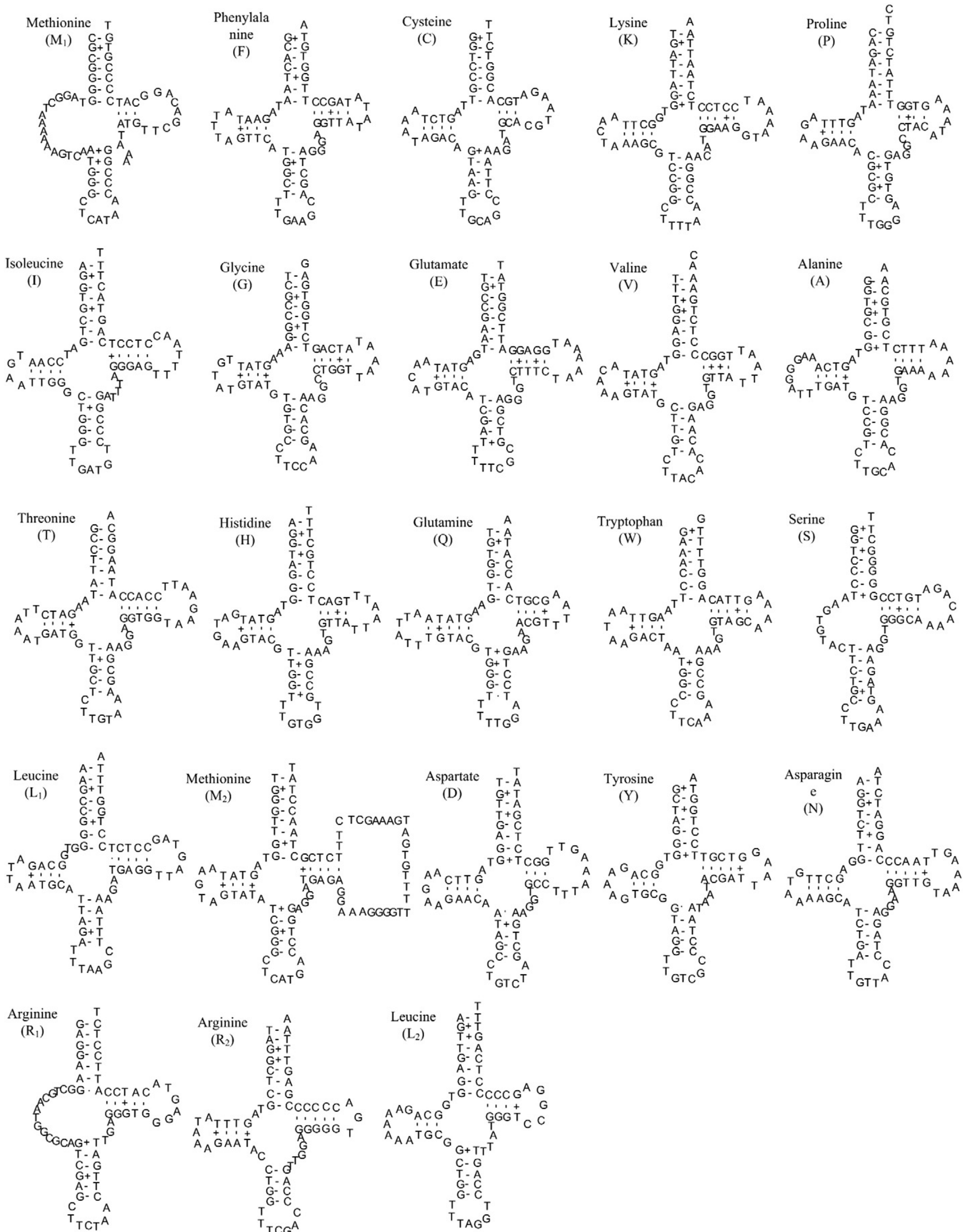

Fig. 4. Putative secondary structure of the 23 tRNA genes predicted based on the *Tegillarca. granosa* mitochondrial genome sequence.

Arcidae species (*nad1*: ATA; *cox2*: ATG) and four genes used the same termination codon (*nad5*: TAA; *nad1*: TAG; *nad4*: TAA; *cox3*: TAA). Comparing the usage of start and stop codon, we found that they were random cases unrelated to the phylogenetic relationships of Arcidae. Similar to the other two Arcidae species, UUU is the most frequently used codon in *T. granosa* (7.9%).

Pairwise divergence among the three Arcidae mitogenomes was calculated based on the separate 12 protein-coding genes (Fig. 3). The *cox1* gene showed the smallest divergence among the 12 protein-coding genes (average divergence: 0.212, Kimura two-parameter distance; K2P), while *nad2* gene revealed the largest divergence (average divergence: 0.553, K2P), which supported the existence of different mutation constraints among genes. In most metazoans, the genes coding for the three subunits of the cytochrome *c* oxidase and cytochrome *b* had a higher degree of conservation (>0.750) than the NADH dehydrogenase genes (Saccone et al., 1999). In this study, this tendency was also observed, but the divergences of *cox2* and *cox3* were similar to that of NADH dehydrogenase genes.

Pairwise divergence among the three Arcidae mitogenomes was also calculated, taking into account nucleotide sequences and amino acid sequences of the concatenated 12 protein-coding genes (Fig. 3). Nucleotide divergence between *S. broughtonii* and *S. kagoshimensis* was the lowest, at 0.302 (K2P), confirming the close relationship between *S. broughtonii* and *S. kagoshimensis*. The highest divergence was revealed between *S. kagoshimensis* and *T. granosa*, at 0.643 (K2P), corresponding to the relatively distant relationship in the phylogenetic analysis. The pairwise divergence values calculated by amino acid sequences among the three Arcidae mitogenomes were lower than that calculated based on nucleotide sequences, implying that synonymous substitutions were more than nonsynonymous substitutions in the protein-coding genes among the three Arcidae mitogenomes.

3.3. Transfer RNA (tRNA) genes and ribosomal RNA (rRNA) genes

There were 23 tRNA genes in the mt genome of *T. granosa*, which ranged from 62 to 93 bp. Most of them can be folded into typical cloverleaf secondary structures (Fig. 4). Compared with the standard complement of 22 tRNA genes, only one *trnS* has been detected, but an additional copy of *trnM* and *trnR* has been identified in the mt genome of *T. granosa*. Differences in tRNA gene composition across the three Arcidae species have been detected. The mt genomes of *S. broughtonii* and *S. kagoshimensis* contained as many as 42 tRNAs, and 10 of the 42 genes were present more than once in the mt genomes of *S. broughtonii* and *S. kagoshimensis*, respectively. Five copies of *trnM* have been found in the mt genomes of *S. broughtonii* and three copies of *trnM* existed in *S. kagoshimensis*. According to the available data, most bivalve mitogenomes contain more than one *trnM* genes, such as *Crassostrea*, *Mytilus* and *Venerupis* (Xu et al., 2011). Alignments of *trnM* genes in the three Arcidae mt genomes indicate that the *trnM1* (located between *nad5* and *nad1*) from *T. granosa* has the highest similarity (77.6%, 79.4%) with *trnM1* in *S. broughtonii* and *trnM1* in *S. kagoshimensis*. Furthermore, these three *trnM* genes are similar in the secondary structure. These results may imply that these three *trnM* genes have a common ancestor. It has been reported that the existence of two *trnM* genes per mitogenome may originate from a gene duplication event, and a higher similarity may imply that the duplication event happened more recently (Xu et al., 2012). The similarity of the two *trnM* genes in the mt genome of *T. granosa* was 52.2%, which is similar to that of oysters, scallops and sea mussels (53.14% in average) (Xu et al., 2012). This result may indicate that these different lineages have similar abilities to maintain the double *trnM* gene during the evolutionary history.

The majority of the speculated tRNA clover-leaf secondary structures include 7 bp aminoacyl acceptor stems, 4 bp DHU stem, 5 bp in the anticodon stem and 3 to 5 bp in the TψC stem. However, part of the tRNA genes in the mt genome of *T. granosa* has unusual secondary structures. The *trnS* gene lacks the DHU-arm, and this feature is common in metazoan mtDNAs (Wolstenholme, 1992). The *trnM1* and *trnR1* also have the DHU-arms replaced by loops. In addition, 6 bp or 4 bp is found in the anticodon stems of *trnS*, *trnR1*, *trnW* and *trnR2* and the DHU stems of *trnQ* and *trnL* have 5 or 3 bp. Furthermore, 13 tRNA (*trnK*, *trnP*, *trnI*, *trnG*, *trnE*, *trnV*, *trnH*, *trnL*, *trnM2*, *trnD*, *trnN*, *trnR1*, *trnR2*) structures include an acceptor stem with eight base pairs.

The *rrnL* and *rrnS* genes of *T. granosa* in this study were identified by sequence comparison with those of *S. broughtonii* and *S. kagoshimensis*. The *rrnL* is located between *trnV* and *trnA*, and the *rrnS* located between *trnN* and *nad2*. The sizes of the *rrnL* and *rrnS* genes are 1343 bp and 641 bp. The lengths of the two genes are similar to those of *S. broughtonii* (1324 bp and 673 bp) and *S. kagoshimensis* (1332 bp and 675 bp). The A + T contents of the *rrnL* and *rrnS* for *T. granosa*, *S. broughtonii* and *S. kagoshimensis* are 59.79% and 53.04%, 63.82% and

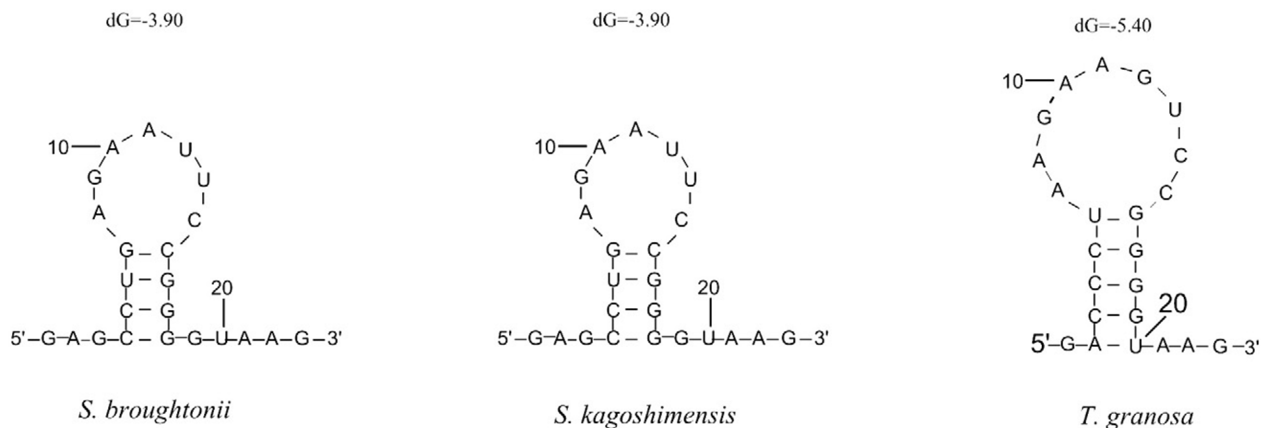

Fig. 5. A) Alignment of the 3′ region sequences of *rrnS*. The numbers indicate the 5′ and 3′ boundary positions of *rrnS* published in NCBI. The numbers in the brackets indicate the 3′ boundary positions used in this study. The green shade indicates the positions involved in the stem–loop structures. B) Stem–loop structures of the 3′ boundary of rrnS folded by the green shaded sequences.

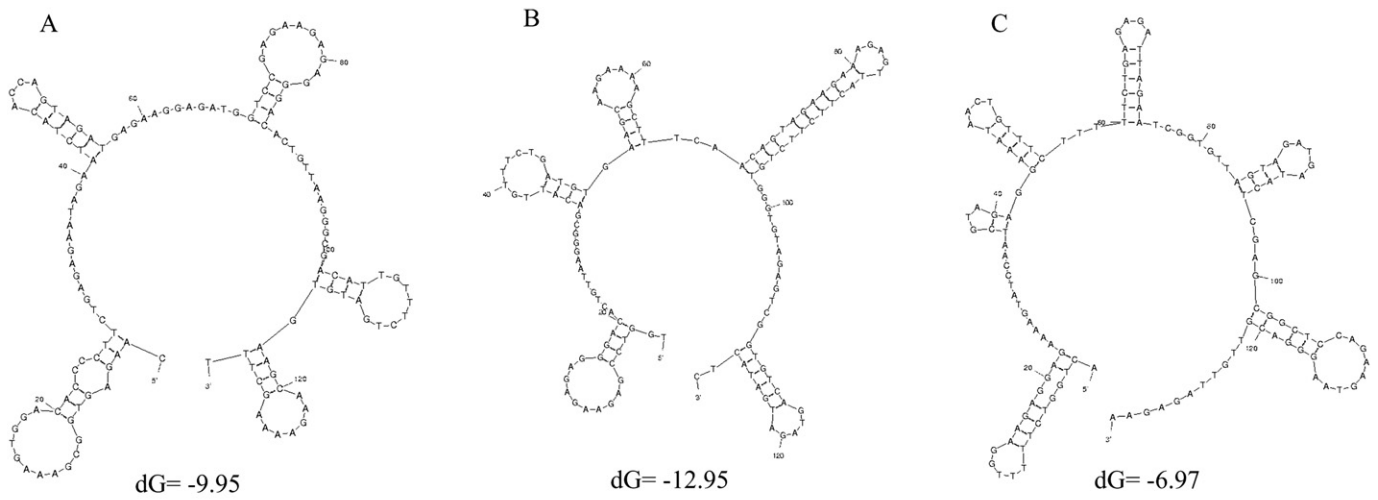

Fig. 6. Stem–loop structures of the tandem repeat motif in the MNR of *T. granosa* mt genome. A) The first stem–loop structure of the tandem repeat motif with 132 bp. B) The second stem–loop structure of the tandem repeat motif with 129 bp. C) The last stem–loop structure of the tandem repeat motif with 132 bp.

53.64%, 63.74% and 52.74%, respectively. These results showed that the *rrnL* and *rrnS* are stable among Arcidae groups.

In *T. granosa* 3′ end boundary of *rrnS* identification, a 19 bp-long sequence was used, which can be folded into a small stem–loop structure (5-bp stem, 9-bp loop). Furthermore, this structure was similar to the 3′ terminus stem–loops (4-bp stem, 7-bp loop) found in *S. broughtonii* and *S. kagoshimensis rrnS* (Fig. 5). As for the *rrnL*, its boundaries were tentatively defined immediately adjacent to the genes, *tRNA-Val* and *tRNA-Ala*. Boundaries were identified by sequence similarity to *S. broughtonii* and *S. kagoshimensis rrnS*. In a wide range of organisms, a heptamer TGGCAGA$(N)_5$G box is well-conserved near the 3′ end of *rrnL*, which is regarded as the signal of the mitochondrial rRNA transcription termination (Valverde et al., 1994). In bivalves, this motif is present in the 5′ end of the *atp8* in *Meretrix lamarckii*, which is adjacent to the *rrnL* (Wang et al., 2011). The same situation also occurs in *M. lusoria* (Wang et al., 2010) and *V. philippinarum* (AB065375). But this motif was not found in the mt genomes of Arcidae species. Although the putative boundaries of *T. granosa rrnS* and *rrnL* have been found, these cannot be precisely determined until transcript mapping is carried out.

3.4. Non-coding regions

As mentioned above, the variable lengths of the non-coding regions (NCRs) in the three mt genomes are the source of their different mt genome sizes. In total, 16,394 bp of the *T. granosa* mt genome was predicted to be non-coding sequence, accounting for 51.9% of the entire mtDNA. This percentage of non-coding mtDNA is much lower than those of *S. broughtonii* (67.4%) and *S. kagoshimensis* (70.6%). Within these non-coding sequences, an 11,848 bp nucleotide segment was putatively identified as the major non-coding region (MNR), which contained three distinct tandem repeat units (24,693–25,037, 26,235–26,953 and 30,292–30,877). They were 344 bp, 718 bp and 585 bp in length, respectively. The first repeat family contained two identical copies and a third copy with a 60% length of a 132-base sequence. The second one had a 129-bp repeat motif and five identical copies and a sixth copy with a 57% length of the repeat motif were found. The last one comprised four identical copies and a fifth copy with a 50% length of a 130 bp sequence. Furthermore, each tandem repeat motif of the MNR formed five secondary structures with stem–loop when the sequence is folded to minimize the free energy of the structure (Fig. 6), supporting the view that the stem–loop structures play an important part in sequence duplications in animal mtDNA (Stanton et al., 1994; Wilkinson and Chapman, 1991). A large number of reports have described tandem repeats in the non-coding regions of mitogenomes of metazoan (Lunt et al., 1998; Rand, 1993; Wilkinson et al., 1997; Yokobori et al., 2004). Tandem repeat units have also been regarded as a common feature for mitogenomes of bivalves. For example, they have been described in the mt genome of *C. farreri* (Xu et al., 2011) *S. divaricatus* (Yuan et al., 2012) and *Coelomactra antiquata* (Meng et al., 2012). Smith and Snyder (2007) have also found repeated sequences in the mt genome of *P. magellanicus* and the repeat units of the non-coding regions are associated with tRNA or pseudo-tRNA structures. This feature was also

Table 4
Signal peptide and transmembrane helix prediction in the ORFs.

	Signal peptides		Transmembrane helices	
ORF	InterProScan	Phobius	TMpred	Phobius
orf01	–	–	–	–
orf02	1–31	–	27–44/55–72	27–48/60–79
orf03	–	–	26–45/55–73	20–44/51–73
orf04	1–31	–	19–41/55–74	24–46/58–76
orf05	–	1–34	4–21/80–97	81–99
orf06	–	–	35–51	33–51
orf07	–	–	25–40	25–42
orf08	–	–	37–61	41–65
orf09	–	–	7–22	–
orf10	–	–	22–37	21–40
orf11	–	–	18–35	16–35/42–59
orf12	1–21	1–21	1–17	–
orf13	1–50	1–50	–	–
orf14	–	–	23–52	23–53
orf15	–	–	–	–
orf16	–	–	13–37/32–52	12–30
orf17	–	–	1–18/11–41	12–33
orf18	–	–	–	–
orf19	–	–	26–41	24–41
orf20	1–22	1–22	3–23	–
orf21	–	–	–	–
orf22	1–21	1–21	1–17	–
orf23	–	–	15–30	12–33
orf24	–	–	13–29	6–28
orf25	–	–	–	–
orf26	–	–	–	–
orf27	1–23	1–23	–	–
orf28	–	–	–	–
orf29	–	–	–	–
orf30	–	–	6–22	6–32
orf31	–	–	–	–
orf32	–	–	7–32	6–32

Note—Signal peptides: Only signal peptides statistically supported (Phobius posterior label probability > 0.5, significance test not provided by InterProScan) are shown; Transmembrane helices: Only transmembrane helices considered significant (TMpred score > 500; Phobius posterior label probability > 0.5) are shown.

reported in the mt genome of brachiopod *Lingula anatina* (Endo et al., 2005). However, this characteristic was not found in *T. granosa*. Among the large non-coding sequences of *T. granosa*, a total of 32 open reading frames for polypeptides were found, ranging in size from 34 to 1641 amino acids. No significant amino acid sequence similarity was detected between the *T. granosa* ORFs and known proteins or any of the ORFs found in other mtDNAs. Of the 32 ORFs, 6 had both signal peptide (SP) and transmembrane helix (TM-helix), 17 had either of them, and the remaining nine contained neither SP nor TM (Table 4). In the 6 ORFs, SPs and TM-helices shared the same region. The signal peptides (SPs) are peptide chains of hydrophobic amino acids, so it's difficult for software to discern them from TM-helices (Käll et al., 2004).

Dramatic change was shown when comparing the non-coding regions of *T. granosa* to those of the other two Arcidae species, which was discussed above. Such structural divergence is reasonable and consistent with the previous finding that the non-coding sequence was the most changeable segment, followed by tRNA genes and PCGs (Cunha et al., 2009; Shao et al., 2006). We favor the hypothesis that this structural divergence may be generated by sequence duplications. The emergence of two large non-coding segments in *S. broughtonii* and *S. kagoshimensis* could be explained reasonably if we suppose that the ancestor of the two *Scapharca* species had the segment as *T. granosa* currently does and then the gene duplication event happened in the mitochondrial evolution. This is also consistent with the report that size variants of mitochondrial DNA are generated by sequence duplications of a variable length during replication (Hale and Singh, 1986). However, at present, the available mitochondrial sequence data is still insufficient to acknowledge the population level variation to validate this hypothesis. Therefore, more sequences need to be determined for comparative analysis in further studies, in order to reveal this mechanism.

3.5. Phylogenetic relationships

Using the combined nucleotide sequences of 12 partitioned protein-coding genes, the phylogenetic relationships of the three Arcidae species and the phylogenetic position of Arcidae within Bivalvia were analyzed. The topological structure of the trees inferred by two different methods (ML, BI) with different building strategies was essentially uniform (Fig. 7), except the small difference in bootstrap values. The phylogenetic tree in the study shows that *S. broughtonii* and *S. kagoshimensis* form a good tight group, with *T. granosa* being the sister taxon at the Arcidae level. This result is consistent with the phenetic-based systematics (Oliver and Holmes, 2006), and supports the previous finding revealed by *cox1*-based DNA barcoding (Feng et al., 2011).

There are some disagreements about the phylogenetic position of Arcidae among bivalve taxa. A possible explanation for the bivalve phylogenetic relationship raised by Waller (1998) is that Ostreidae and Pectinidae had the closest relationship, as the sister group of Arcidae. Then Mytilidae clustered with the group of (Arcidae + (Ostreidae + Pectinidae)). However, Steiner and Hammer (2000) hypothesized that Pectinidae is first clustered with Arcidae and then united with the clade of Ostreidae and Mytilidae, which was compatible with the study using *cox1* sequences (Matsumoto, 2003). The phylogenetic analysis of four mitochondrial markers (*rrnL*, *rrnS*, *cox1*, *cob*) was conducted by Plazzi et al. (2011), which proposed that Ostreidae and Arcidae formed a single group, being the sister group of Pectinidae clade. Mytilidae was found at the base of the (Pectinidae + (Ostreidae + Arcidae)) with 100% support. Later, Sharma et al. (2012) evaluated the basal relationships of Bivalvia using four nuclear protein-encoding genes (ATP synthase β, elongation factor-1 α, myosin heavy chain type II, and RNA polymerase II), supporting that Arcidae to be sister species of Mytilidae and that

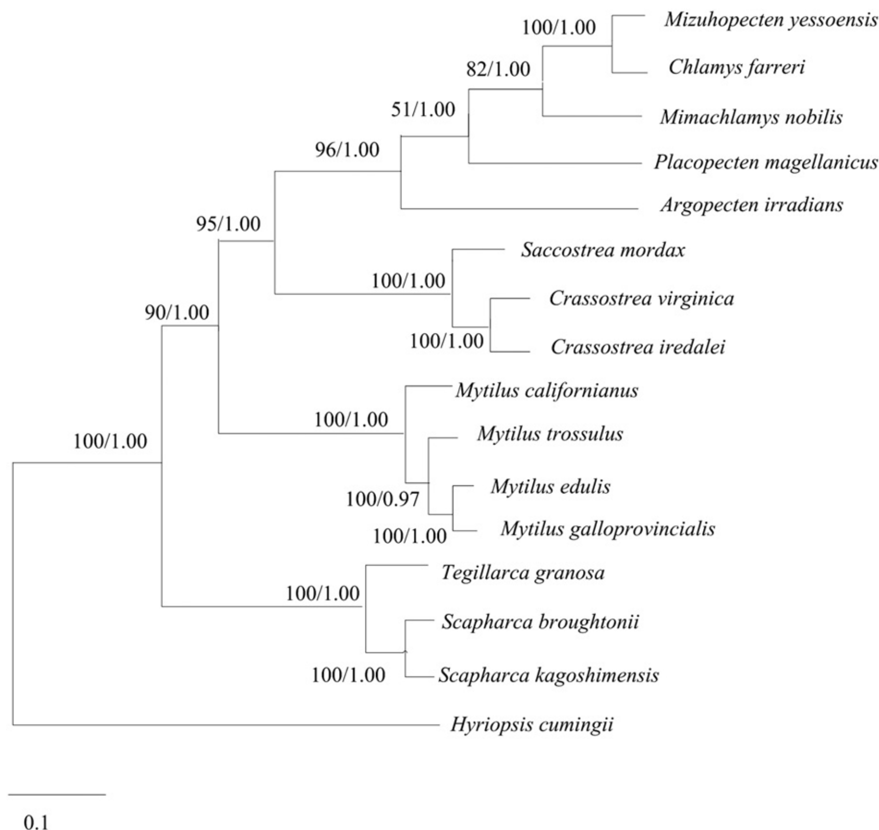

Fig. 7. Phylogenetic trees derived from Maximum Likelihood (ML) and Bayesian analyses based on partitioned nucleotide sequences of 12 mitochondrial protein-coding genes. The first number at each node is the bootstrap probability of ML analyses and the second number is Bayesian posterior probability.

Pectinidae appears as sister to this group. In our study, according to the phylogenetic tree, we can clearly see that Ostreidae and Pectinidae have the closest relationship, with Mytilidae being the sister taxon. Arcidae is well supported to be the sister group to the ((Pectinidae + Ostreidae) + Mytilidae) clade. This result suggests that the family Arcidae diverges early and is positioned at the base of the ((Pectinidae + Ostreidae) + Mytilidae) clade, which is incompatible with the previous viewpoints based on short fragments of nuclear gene or mtDNA.

Acknowledgments

This study was supported by research grants from the National Natural Science Foundation of China (41276138), and the National Marine Public Welfare Research Program (201305005).

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gene.2014.12.011.

References

Bao, Y., Lin, Z., 2010. Generation, annotation, and analysis of ESTs from hemocyte of the bloody clam, *Tegillarca granosa*. Fish Shellfish Immunol. 29, 740–746.
Benson, G., 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 27, 573–580.
Boore, J., 1999. Animal mitochondrial genomes. Nucleic Acids Res. 27, 1767–1780.
Boyce, T.M., Zwick, M.E., Aquadro, C.F., 1989. Mitochondrial DNA in the bark weevils: size, structure and heteroplasmy. Genetics 123, 825–836.
Breton, S., Beaupré, H.D., Stewart, D.T., Hoeh, W.R., Blier, P.U., 2007. The unusual system of doubly uniparental inheritance of mtDNA: isn't one enough? Trends Genet. 23, 465–474.
Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17, 540–552.
Cunha, R.L., Grande, C., Zardoya, R., 2009. Neogastropod phylogenetic relationships based on entire mitochondrial genomes. BMC Evol. Biol. 9, 210.
Dreyer, H., Steiner, G., 2006. The complete sequences and gene organisation of the mitochondrial genomes of the heterodont bivalves *Acanthocardia tuberculata* and *Hiatella arctica*—and the first record for a putative Atpase subunit 8 gene in marine bivalves. Front. Zool. 3, 13.
Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7, 214.
Eddy, S.R., 2011. Accelerated profile HMM searches. PLoS Comput. Biol. 7, e1002195.
Endo, K., Noguchi, Y., Ueshima, R., Jacobs, H.T., 2005. Novel repetitive structures, deviant protein-encoding sequences and unidentified ORFs in the mitochondrial genome of the brachiopod *Lingula anatina*. J. Mol. Evol. 61, 36–53.
Feng, Y.W., Li, Q., Kong, L.F., Zheng, X.D., 2011. COI-based DNA barcoding of *Arcoida* species (Bivalvia: Pteriomorphia) along the coast of China. Mol. Ecol. Resour. 11, 435–441.
Finn, R.D., Clements, J., Eddy, S.R., 2011. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 39, 29–37.
Folmer, O., Black, M., Hoeh, W., Lutz, R., Vrijenhoek, R., 1994. DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan invertebrates. Mol. Mar. Biol. Biotechnol. 3, 294–299.
Gach, M.H., Brown, W.M., 1997. Characteristics and distribution of large tandem duplications in brook stickleback *Culaea inconstants* mitochondrial DNA. Genetics 145, 383–394.
Ghiselli, F., Milani, L., Guerra, D., Chang, P.L., Breton, S., Nuzhdin, S.V., Passamonti, M., 2013. Structure, transcription, and variability of metazoan mitochondrial genome: perspectives from an unusual mitochondrial inheritance system. Genome Biol. Evol. 5, 1535–1554.
Hale, L.R., Singh, R.S., 1986. Extensive variation and heteroplasmy in size of mitochondrial DNA among geographic populations of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. U. S. A. 83, 8813–8817.
Hofmann, K., Stoffel, W., 1993. TMbase—a database of membrane spanning proteins segments. Biol. Chem. Hoppe-Seyler 374, 166.
Hyman, B.C., Slater, T.M., 1990. Recent appearance and molecular characterization of mitochondrial DNA deletions within a defined nematode pedigree. Genetics 124, 845–853.
Hyman, B.C., Beck, J.L., Weiss, K.C., 1988. Sequence amplification and gene rearrangement in parasitic nematode mitochondrial DNA. Genetics 120, 707–712.
Käll, L., Krogh, A., Sonnhammer, E.L.L., 2004. A combined transmembrane topology and signal peptide prediction method. J. Mol. Biol. 338, 1027–1036.
Katoh, K., Kuma, K., Toh, H., Miyata, T., 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res. 33, 511–518.
Laslett, D., Canback, B., 2008. ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. Bioinformatics 24, 172–175.
Li, Q., Park, C., Kijima, A., 2002. Isolation and characterization of microsatellite loci in the Pacific abalone, *Haliotis discus hannai*. J. Shellfish Res. 21, 811–815.
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. D Genomics Proteomics 8, 72–81.
Lowe, T.M., Eddy, S.R., 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25, 955–964.
Lunt, D.H., Whipple, L.E., Hyman, B.C., 1998. Mitochondrial DNA variable number tandem repeats (VNTRs): utility and problems in molecular ecology. Mol. Ecol. 7, 1441–1455.
Matsumoto, M., 2003. Phylogenetic analysis of the subclass Pteriomorphia (Bivalvia) from mtDNA COI sequences. Mol. Phylogenet. Evol. 27, 429–440.
Meng, X., Zhao, N., Shen, X., Hao, J., Liang, M., Zhu, X., Cheng, H.L., Yan, B.L., Liu, Z., 2012. Complete mitochondrial genome of *Coelomactra antiquata* (Mollusca: Bivalvia): the first representative from the family Mactridae with novel gene order and unusual tandem repeats. Comp. Biochem. Physiol. D Genomics Proteomics 7, 175–179.
Moritz, C., Brown, W.M., 1986. Tandem duplication of D-loop and ribosomal RNA sequences in lizard mitochondrial DNA. Science 233, 1425–1427.
Oliver, P.G., Holmes, A.M., 2006. The Arcoidea (Mollusca: Bivalvia): a review of the current phenetic-based systematics. Zool. J. Linnean Soc. 148, 237–251.
Passamonti, M., Scali, V., 2001. Gender-associated mitochondrial DNA heteroplasmy in the venerid clam *Tapes philippinarum* (Mollusca Bivalvia). Curr. Genet. 39, 117–124.
Perna, N.T., Kocher, T.D., 1995. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. J. Mol. Evol. 41, 353–358.
Plazzi, F., Ceregato, A., Taviani, M., Passamonti, M., 2011. A molecular phylogeny of bivalve mollusks: ancient radiations and divergences as revealed by mitochondrial genes. PLoS One 6, e27147.
Pons, J.L., Labesse, G., 2009. @TOME-2: a new pipeline for comparative modeling of protein–ligand complexes. Nucleic Acids Res. 37, W485–W491.
Posada, D., 2008. jModelTest: phylogenetic model averaging. Mol. Biol. Evol. 25, 1253–1256.
Powers, T.O., Platzer, E.G., Hyman, B.C., 1986. Large mitochondrial genome and mitochondrial DNA size polymorphism in the mosquito parasite, *Romanomermis culicivorax*. Curr. Genet. 11, 71–77.
Rand, D.M., 1993. Endotherms, ectotherms, and mitochondrial genome-size variation. J. Mol. Evol. 37, 281–295.
Richardson, L.R., Gold, J.R., 1991. A tandem duplication in the mitochondrial DNA of the red shiner, *Cyprinella lutrensis*. Copeia 1991, 842–845.
Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.
Saccone, C., De Giorgi, C., Gissi, C., Pesole, G., Reyes, A., 1999. Evolutionary genomics in Metazoa: the mitochondrial DNA as a model system. Gene 238, 195–209.
Shao, R., Barker, S.C., Mitani, H., Takahashi, M., Fukunaga, M., 2006. Molecular mechanisms for the variation of mitochondrial gene content and gene arrangement among chigger mites of the genus *Leptotrombidium* (Acari: Acariformes). J. Mol. Evol. 63, 251–261.
Sharma, P.P., González, V.L., Kawauchi, G.Y., Andrade, S.C.S., Guzmán, A., Collins, T.M., Glover, E.A., Harper, E.M., Healy, J.M., Mikkelsen, P.M., Taylor, J.D., Bieler, R., Giribet, G., 2012. Phylogenetic analysis of four nuclear protein-encoding genes largely corroborates the traditional classification of Bivalvia (Mollusca). Mol. Phylogenet. Evol. 65, 64–74.
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.
Stamatakis, A., Hoover, P., Rougemont, J., 2008. A rapid bootstrap algorithm for the RAxML Web servers. Syst. Biol. 57, 758–771.
Stanton, D.J., Daehler, L.L., Moritz, C.C., Brown, W.M., 1994. Sequences with the potential to form stem-and-loop structures are associated with coding-region duplications in animal mitochondrial DNA. Genetics 137, 233–241.
Steiner, G., Hammer, S., 2000. Molecular phylogeny of the Bivalvia inferred from 18S rDNA sequences with particular reference to the Pteriomorphia. Geol. Soc. Lond., Spec. Publ. 177, 11–29.
Stothard, P., Wishart, D.S., 2005. Circular genome visualization and exploration using CGView. Bioinformatics 21, 537–539.
Sun, S.E., Kong, L.F., Yu, H., Li, Q., 2014. The complete mitochondrial genome of *Scapharca kagoshimensis* (Bivalvia: Arcidae). Mitochondrial DNA 1–2.
Talavera, G., Castresana, J., 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst. Biol. 56, 564–577.
Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731–2739.
Valverde, J.R., Marco, R., Garesse, R., 1994. A conserved heptamer motif for ribosomal RNA transcription termination in animal mitochondria. Proc. Natl. Acad. Sci. U. S. A. 91, 5368–5371.
Waller, T.R., 1998. Origin of the molluscan class Bivalvia and a phylogeny of major groups. Bivalves, An Eon of Evolution. University of Calgary Press, Calgary, pp. 1–45.
Wallis, G.P., 1986. Mitochondrial DNA insertion polymorphism and germ line heteroplasmy in the *Triturus cristatus* complex. Heredity 58, 229–238.
Wang, H., Zhang, S., Xiao, G., Liu, B., 2010. Complete mtDNA of the *Meretrix lusoria* (Bivalvia: Veneridae) reveals the presence of an *atp8* gene, length variation and heteroplasmy in the control region. Comp. Biochem. Physiol. D Genomics Proteomics 5, 256–264.
Wang, H., Zhang, S., Xiao, G., Liu, B., 2011. Complete mtDNA of the *Meretrix lamarckii* (Bivalvia: Veneridae) and molecular identification of suspected *M. lamarckii* based on the whole mitochondrial genome. Mar. Genomics 4, 263–271.
Wilkinson, G.S., Chapman, A.M., 1991. Length and sequence variation in evening bat D-loop mtDNA. Genetics 128, 607–617.
Wilkinson, G.S., Mayer, F., Kerth, G., Petri, B., 1997. Evolution of repeated sequence arrays in the D-Loop region of bat mitochondrial DNA. Genetics 146, 1035–1048.

Wolstenholme, D.R., 1992. Genetic novelties in mitochondrial genomes of multicellular animals. Curr. Opin. Genet. Dev. 2, 918–925.
Wu, X.Y., Xu, X.D., Yu, Z.N., Wei, Z.P., Xia, J.J., 2010. Comparison of seven *Crassostrea* mitogenomes and phylogenetic analyses. Mol. Phylogenet. Evol. 57, 448–454.
Wu, X.Y., Li, X.L., Yu, Z.N., 2013. The mitochondrial genome of the scallop *Mimachlamys senatoria* (Bivalvia, Pectinidae). Mitochondrial DNA 1–3.
Xu, K.F., Kanno, M., Yu, H., Li, Q., Kijima, A., 2011. Complete mitochondrial DNA sequence and phylogenetic analysis of Zhikong scallop *Chlamys farreri* (Bivalvia: Pectinidae). Mol. Biol. Rep. 38, 3067–3074.
Xu, X.D., Wu, X.Y., Yu, Z.N., 2012. Comparative studies of the complete mitochondrial genomes of four *Paphia* clams and reconsideration of subgenus *Neotapes* (Bivalvia: Veneridae). Gene 494, 17–23.
Yokobori, S.I., Fukuda, N., Nakamura, M., Aoyama, T., Oshima, T., 2004. Long-term conservation of six duplicated structural genes in cephalopod mitochondrial genomes. Mol. Biol. Evol. 21, 2034–2046.
Yu, Z.N., Wei, Z.P., Kong, X.Y., Shi, W., 2008. Complete mitochondrial DNA sequence of oyster *Crassostrea hongkongensis*—a case of tandem duplication-random loss for genome rearrangement in *Crassostrea*? BMC Genomics 9, 477.
Yuan, Y., Li, Q., Yu, H., Kong, L.F., 2012. The complete mitochondrial genomes of six heterodont bivalves (Tellinoidea and Solenoidea): variable gene arrangements and phylogenetic implications. PLoS ONE 7, e32353.
Zdobnov, E.M., Apweiler, R., 2001. InterProScan—an integration platform for the signature-recognition methods in InterPro. Bioinformatics 17, 847–848.
Zouros, E., 2013. Biparental inheritance through uniparental transmission: the doubly uniparental inheritance (DUI) of mitochondrial DNA. Evol. Biol. 40, 1–31.
Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31, 3406–3415.