Complete mitochondrial genomes of *Trisidos kiyoni* and *Potiarca pilula*: Varied mitochondrial genome size and highly rearranged gene order in Arcidae

Shao’e Sun, Qi Li, Lingfeng Kong & Hong Yu

We present the complete mitochondrial genomes (mitogenomes) of *Trisidos kiyoni* and *Potiarca pilula*, both important species from the family Arcidae (Arcoida: Arcacea). Typical bivalve mtDNA features were described, such as the relatively conserved gene number (36 and 37), a high A + T content (62.73% and 61.16%), the preference for A + T-rich codons, and the evidence of non-optimal codon usage. The mitogenomes of Arcidae species are exceptional for their extraordinarily large and variable sizes and substantial gene rearrangements. The mitogenome of *T. kiyoni* (19,614 bp) and *P. pilula* (28,470 bp) are the two smallest Arcidae mitogenomes. The compact mitogenomes are weakly associated with gene number and primarily reflect shrinkage of the non-coding regions. The varied size in Arcidae mitogenomes reflect a dynamic history of expansion. A significant positive correlation is observed between mitogenome size and the combined length of *cox1-3*, the lengths of *Cytb*, and the combined length of rRNAs (*rrnS* and *rrnL*) (*P* < 0.001). Both protein coding genes (PCGs) and tRNA rearrangements is observed in *P. pilula* and *T. kiyoni* mitogenomes. This analysis imply that the complicated gene rearrangement in mitochondrial genome could be considered as one of key characters in inferring higher-level phylogenetic relationship of Arcidae.

Mitochondrial DNA (mtDNA) is the only extranuclear genome in animal cytoplasm. Most metazoan mitochondrial genomes (mitogenomes) are covalently closed circular molecules which range from 14 to 42 kb in length, but see some Arcidae bivalves. The mitogenome generally encodes for 37 genes: 13 protein-coding genes (PCGs), 2 ribosomal RNA genes (rRNAs), and 22 transfer RNA genes (tRNAs). In general, there are few intergenic nucleotides except for a single large non-coding region generally thought to contain elements that control the initiation of replication and transcription of the mitogenome. Owing to the abundance of mitochondria in cells, maternal inheritance, absence of introns, and higher evolutionary rates, the mitogenomes of Metazoa are good model systems for comparative and evolutionary genomic studies. Several features, such as genome size, gene arrangement, gene number and structure, can be easily and systematically investigated in the small mitogenome.

Mollusca is a megadiverse phylum originated in the PreCambrian/Cambrian border, according to the fossil record, during which organisms conquered a great variety of habitats, food habits and evolved a variety of body sizes. Bivalve mollusks present one of the most variable mitogenomes described among metazoans with low conservation of genome size, gene arrangement, strand assignment, gene duplications and losses, nucleotide composition, and more. The mitogenomes of relatively large size (greater than 20 kb) have been found in bivalves such as the deep sea scallop *Placopecten magellanicus* (up to 40,725 bp), the Zhikong scallop *Chlamys farrei* (21,693 bp), the Manila clam *Venerupis philippinarum* (22,676 bp in female type, 21,441 bp in male type), and the Arcidae species, *i.e.* *Scapharca brouhontii* (46,985 bp), *Scapharca kagoshimensis* (46,713 bp), *Tegillarca granosa* (31,589 bp) and *Anadara vellicata* (34,147 bp). Size variation in bivalves mitogenomes is usually due to the different length of the non-coding regions. A peculiar way of mitochondrial inheritance, doubly uniparental
in inheritance (DUI), has been found in *V. philippinarum*, which may also influence the genome size in male and female mitochondrial DNA

Gene arrangements appear to be dramatically variable in the major groups of bivalves, even with differences in the same family or genus. Pectinidae species seem to be a good example to prove this. Comparing gene orders of *C. farreri*, *Argopesten iradians*, *Mimachlamys nobilis* and *P. magellanicus*, even after excluding the tRNA genes from the comparison, the four mitogenomes still show no identical gene arrangement. 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. Gene order of the mitogenome can be used to investigate evolution of organisms and of their genomes by providing (1) characters that can be used in phylogenetic analysis of ancient lineages and (2) information that can be used to develop models for the mechanisms involved in gene rearrangement, replication, and regulation.

The family Arcidae belongs to the superfamily Arcacea in the order Arcoidea. The species of Arcidae are globally distributed, predominantly in the tropical shallow waters and warm temperate seas, containing approximately 260 species and 31 genera. Arcidae is subdivided in two subfamilies, Anadarinae and Arcinae, based on the strength of the byssus. Arcinae contains some of the best-known and most widely distributed genera, like *Arca* (Linnaeus, 1758) and *Barbatia* (Gray, 1842). In Anadarinae, several species have significant economic value. For example, *Tegillarca granosa* is cultivated on wide mudflats in South-East Asia (China, Taiwan, Korea, Malaysia and Thailand) and has been consumed by humans for centuries. *Scapharca* species are harvested in Japan and China. At present, four complete mitogenomes were available from this family, *i.e.*, *S. broughtoni* (GenBank: AB729113), *S. kagoshimensis* (GenBank: KF750628), *T. granosa* (GenBank: KJ607173) and *A. vellicata* (GenBank: KP954700). Obvious differences in mitogenome organization of the Arcidae species were observed: (1) the sizes of three mitogenomes are distinct from each other, *i.e.*, 46,985 bp for *S. kagoshimensis*, 46,713 bp for *S. kagoshimensis*, 31,589 bp for *T. granosa* and 34,147 bp for *A. vellicata*; (2) the genomes show distinct gene arrangement patterns, namely unique rearrangements involving the tRNA genes. These variation of genomic organization provide a good system to understand the evolutionary history of the mitochondrial genomes.

In this work, we present the complete mitochondrial genomes of *T. kiyoni* and *P. pilula*, both important species from the family Arcidae (Arcoida: Arcacea). The characterization of the evolution and structural organization of *T. kiyoni* and *P. pilula* mitochondrial genomes were analyzed and compared with other Arcidae mitogenomes. We discussed our findings with particular reference to the variations in genome size and gene arrangement in the family Arcidae. We reconstructed the phylogenetic relationships of six Arcidae species based on twelve protein-coding genes, which allows for the understanding of ancestral organization of the Arcidae mitogenomes.

**Results and Discussion**

**Genome organization, structure and composition.** The complete mtDNA sequences of *T. kiyoni* and *P. pilula* are 19,614 bp and 28,470 bp in size, respectively, and their structural organization are depicted in Fig. 1, Tables 1 and 2. The mitochondrial genome sequence of *T. kiyoni* is the smallest in all Arcidae mitochondrial genomes available in the GenBank. The *T. kiyoni* mitogenome contains 12 PCGs, 22 tRNA genes, 2 rRNA genes and non-coding regions. Unlike *T. kiyoni*, *P. pilula* mitogenome contains 23 tRNA genes (the standard 22 tRNAs and an extra *trnR*). No *atp8* coding sequence was detected in both *T. kiyoni* and *P. pilula* mitogenomes. All the genes are transcribed from the (+) strand of the molecules. The two mitogenomes exhibit different gene arrangements for both tRNA and protein-coding genes with other Arcidae mitogenomes.

In the *T. kiyoni* mitogenome, there are 36 non-coding regions with a total of 5,369 bp long varying from 2 bp to 1009 bp. The longest non-coding region is situated between *trnF* and *trnS* (GenBank: AB729113).

The complete mitochondrial genomes of *T. kiyoni* and *P. pilula*. Genes for proteins and rRNA (*rrnS* and *rrnL*) are listed under abbreviations. Transfer-RNAs are represented by their one-letter amino acid code.

**Figure 1.**
34 non-coding regions with a total of 13,642 bp with various lengths of 0–7,408 bp. The two long non-coding regions are located between *cox3* and *trnR2* (7,408 bp), *trnR2* and *cox1* (2,435 bp), respectively. In *P. pilula*, the overlaps occur two times and involve a total of 45 bp, which located between *cox1* and *nad5* (31 bp), *trnS2* and *cox3* (14 bp). There is no overlapping gene in the mitochondrial genome of *T. kiyoni*. The overlap between *cox1* and *nad5* was also observed in the mitogenome of *S. broughtonii* and *A. vellicata* 3,15.

The A + T content, AT-skew, and GC-skew are three parameters, which were usually used in the investigation of the nucleotide-compositional behavior of mitochondrial genomes 27,28. The nucleotide compositions of the complete mtDNA sequence for both of the Arcidae species are biased toward A and T (Table 3). The A + T content is 62.73% in *T. kiyoni* and 61.16% in *P. pilula*. The non-coding region (NCR) show the highest A + T content (67.15% and 63.09%, respectively). In order to evaluate the base bias in the mitogenomes, we measured skewness in different gene regions of *T. kiyoni* and *P. pilula* mitochondrial genomes, and found the values of the AT-skew were mostly negative, as well as values of the GC-skew were all positive (Table 3).

### Table 1. Organization of the mitochondrial genome of *Trisidos kiyoni.*

| Gene   | Strand | Position From | Position To | Size nt | Codon Start | Codon Stop | Intergenic nucleotides |
|--------|--------|---------------|-------------|---------|-------------|-------------|------------------------|
| *cox1* | +      | 1 1560        | 1560        | 520     | ATG         | TAG         | 501                    |
| *trnA* | +      | 1594 1660     | 1660        | 67      |             |             | 33                     |
| *trnH* | +      | 1685 1751     | 1751        | 67      |             |             | 24                     |
| *nad1* | +      | 1922 2716     | 2716        | 795     | 265         | ATG         | TAG 170                |
| *cox2* | +      | 2806 3444     | 3444        | 639     | 213         | ATA         | TAG 89                 |
| *trnI* | +      | 3454 3527     | 3527        | 74      |             |             | 9                      |
| *nad5* | +      | 3562 5250     | 5250        | 1689    | 563         | ATA         | TAG 34                 |
| *trnLCCUA* | + | 5301 5370   | 5370        | 70      |             |             | 50                     |
| *Cytb* | +      | 5692 6630     | 6630        | 939     | 313         | ATG         | TAA 321                |
| *nad6* | +      | 6719 7024     | 7024        | 306     | 102         | ATG         | TAA 88                 |
| *nad4* | +      | 7292 7564     | 7564        | 273     | 91          | ATA         | TAG 267                |
| *nad2* | +      | 7591 8913     | 8913        | 1323    | 441         | ATA         | TAA 26                 |
| *trnF* | +      | 9011 10168    | 10168       | 1158    | 386         | ATG         | TAG 97                 |
| *trnSAGA* | + | 11254 11320 | 11320        | 67      |             |             | 1009                   |
| *trnK* | +      | 11348 11418   | 11418       | 71      |             |             | 27                     |
| *trnP* | +      | 11442 11512   | 11512       | 71      |             |             | 23                     |
| *trnR* | +      | 11525 11593   | 11593       | 69      |             |             | 12                     |
| *trnW* | +      | 11598 11666   | 11666       | 69      |             |             | 4                      |
| *trnD* | +      | 11729 11794   | 11794       | 66      |             |             | 62                     |
| *trnS* | +      | 12084 12793   | 12793       | 710     |             |             | 289                    |
| *trnN* | +      | 12983 13053   | 13053       | 71      |             |             | 189                    |
| *trnT* | +      | 13074 13140   | 13140       | 67      |             |             | 20                     |
| *trnM* | +      | 13177 13247   | 13247       | 71      |             |             | 36                     |
| *trnL* | +      | 13335 14813   | 14813       | 479     |             |             | 87                     |
| *trnLCCUA* | + | 14833 14900 | 14900        | 68      |             |             | 19                     |
| *trnC* | +      | 14930 15000   | 15000       | 71      |             |             | 29                     |
| *trnQ* | +      | 15046 15112   | 15112       | 67      |             |             | 45                     |
| *trnE* | +      | 15168 15235   | 15235       | 68      |             |             | 55                     |
| *trnT* | +      | 15258 15324   | 15324       | 67      |             |             | 22                     |
| *trnV* | +      | 15327 15395   | 15395       | 69      |             |             | 2                      |
| *trnG* | +      | 15458 15526   | 15526       | 69      |             |             | 62                     |
| *trnLCCUA* | + | 16525 16588 | 16588        | 64      |             |             | 998                    |
| *cox3* | +      | 16679 17425   | 17425       | 747     | 249         | ATG         | TAG 90                 |
| *atp6* | +      | 17901 18602   | 18602       | 702     | 234         | ATA         | TAG 475                |
| *nad3* | +      | 18698 19111   | 19111       | 414     | 138         | ATG         | TAG 95                 |

34 non-coding regions with a total of 13,642 bp with various lengths of 0–7,408 bp. The two long non-coding regions are located between *cox3* and *trnR2* (7,408 bp), *trnR2* and *cox1* (2,435 bp), respectively. In *P. pilula*, the overlaps occur two times and involve a total of 45 bp, which located between *cox1* and *nad5* (31 bp), *trnS2* and *cox3* (14 bp). There is no overlapping gene in the mitochondrial genome of *T. kiyoni*. The overlap between *cox1* and *nad5* was also observed in the mitogenome of *S. broughtonii* and *A. vellicata* 3,15.

The A + T content, AT-skew, and GC-skew are three parameters, which were usually used in the investigation of the nucleotide-compositional behavior of mitochondrial genomes 27,28. The nucleotide compositions of the complete mtDNA sequence for both of the Arcidae species are biased toward A and T (Table 3). The A + T content is 62.73% in *T. kiyoni* and 61.16% in *P. pilula*. The non-coding region (NCR) show the highest A + T content (67.15% and 63.09%, respectively). In order to evaluate the base bias in the mitogenomes, we measured skewness in different gene regions of *T. kiyoni* and *P. pilula* mitochondrial genomes, and found the values of the AT-skew were mostly negative, as well as values of the GC-skew were all positive (Table 3).

**Protein-coding genes and ribosomal RNA genes.** The entire length of the PCGs of *T. kiyoni* was 10,545 bp, while that of *P. pilula* was 11,151 bp. The overall A + T content of the 12 PCGs was 61.63% in the *T. kiyoni* mitogenome, ranging from 59.91% (*nad3*) to 64.72% (*atp6*). In *P. pilula* mitogenome, the A + T content of the 12 PCGs was 60.36%, ranging from 57.01% (*co2*) to 63.58% (*nad2*).

In the mitochondrial genomes of *T. kiyoni* and *P. pilula*, all of the 12 PCGs have complete start codons e.g. ATG and ATA (Table 1). In the *T. kiyoni* mitochondrial genome, eight and four PCGs started with ATG and ATA, respec-
tively, while in the *P. pilula* mitochondrial genome, five and seven PCGs started with ATG and ATA, respectively. All the 12 PCGs genes have complete stop codons, e.g. TAA, TAG.

A total of 3515 and 3717 amino acids are encoded in *T. kiyoni* and *P. pilula* mitogenomes, respectively. The codon usage of *T. kiyoni* and *P. pilula* mitochondrial genomes (Table 4) are similar to that of other Arcidae species. All codons are used in both of the mitogenomes but with different frequencies. Amino acids coded by A + T-rich codon families (e.g. Phe, Tyr and Lys) are more frequent than amino acids coded by G + C-rich codon families (e.g. Pro and Arg). The ratio G + C/A + T-rich codons was 0.43 in *T. kiyoni* mitogenome, which is lower than that of *P. pilula* mitogenome. In both of the mitogenomes, G-ending codons are most abundant in NNY codon families and T-ending codons are most abundant in NNR and NNN codon families, and consequently, the (+) strands are T and G-rich, outlines another bias of Arcidae codon usage. There are 1.7 times more G than C and 2.7 times more T than A in the neutral sites of *T. kiyoni*. In the case of *P. pilula*, 3.4 times more G than C and 1.6 times more T than A were found at at the strand neutral sites. Codon usage bias was also observed in the vertebrate mitochondrial genomes, in which the two strands are exposed to different mutational pressures during replication, leading to an increased frequency of A and C in the (+) strand (or L-strand, in case of vertebrates)27,29,30. However, the Arcidae mtDNA showed the accumulation of T and G in the (+) strand, suggesting that a reversal of strand asymmetry have occurred in the members of these taxa.

The nonsynonymous (Ka) and synonymous (Ks) substitution rates reflect the evolutionary dynamics of protein-coding sequences across closely related species31,32. In order to detect the influence of selection pressure in Arcidae mitochondrial genomes, the number of Ka and Ks, and the ratio of Ka/Ks, were calculated for all

| Gene | Strand | Position | Size | Codon | Intergenic nucleotides |
|------|--------|----------|------|-------|------------------------|
| coxl | +      | 1        | 1464 | ATA   | TAA                    |
| nad5 | +      | 1434     | 3161 | ATG   | TAA                    |
| trnM | +      | 3172     | 3236 | 65    |                        |
| nad3 | +      | 3240     | 4190 | 592   | 317                    |
| nad4 | +      | 4217     | 5497 | 1281  | 427                    |
| Cytb | +      | 5609     | 6820 | 1212  | 404                    |
| trnP | +      | 6864     | 6931 | 68    |                        |
| cox2 | +      | 7067     | 8008 | 942   | 314                    |
| trnC | +      | 8652     | 8716 | 65    | 643                    |
| nad6 | +      | 9050     | 9370 | 321   | 107                    |
| trnK | +      | 9387     | 9458 | 72    | 16                     |
| atp8 | +      | 9534     | 10178| 645   | 215                    |
| trnP | +      | 10193    | 10259| 67    |                        |
| rnl  | +      | 10301    | 10367| 67    | 41                     |
| trnG | +      | 10375    | 10445| 71    | 7                      |
| trnE | +      | 11007    | 11075| 69    | 561                    |
| trnV | +      | 11586    | 11651| 66    | 510                    |
| trnL | +      | 11689    | 13032| 1344  | 37                     |
| trnA | +      | 13033    | 13100| 68    | 0                      |
| trnT | +      | 13121    | 13189| 69    | 20                     |
| trnH | +      | 13202    | 13270| 69    | 12                     |
| trnQ | +      | 13283    | 13348| 66    | 12                     |
| nad3 | +      | 13495    | 13848| 354   | 118                    |
| nad4 | +      | 13853    | 14122| 270   | 90                     |
| trnD | +      | 14478    | 14546| 69    | 355                    |
| trnL^LUA^ | + | 14632 | 14698 | 67 | 85 |
| trnV | +      | 14769    | 14834| 66    | 70                     |
| trnN | +      | 14852    | 14921| 70    | 17                     |
| trnS | +      | 15175    | 15847| 673   | 253                    |
| nad2 | +      | 16003    | 17070| 1068  | 356                    |
| trnS^AGA^ | + | 17107 | 17174 | 68 | 36 |
| trnk1 | + | 17179 | 17244 | 66 | 4 |
| trnL^LUA^ | + | 17268 | 17335 | 68 | 23 |
| trnW | +      | 17416    | 17488| 73    | 80                     |
| trnS^AGA^ | + | 17496 | 17560 | 65 | 7 |
| cox3 | +      | 17547    | 18479| 933   | 311                    |
| trnR2 | +    | 25888    | 25951| 64    | 7408                   |

Table 2. Organization of the mitochondrial genome of *Potiarca pilula*. 

...
them can fold into canonical clover-leaf secondary structures except the secondary structures of tRNAs in two Arcidae mitogenomes were shown in Supplementary Figs 1 and 2. Most of the trnSAGA gene was also observed in vertebrate mitochondrial genomes, e.g. fishes, and may be associated with the distance from the origin of replication and the highest Ka/Ks ratio compared to the other mitochondrial coding genes. This pattern was also observed in vertebrate mitochondrial genomes, e.g. fishes, and may be associated with the distance from the origin of replication. Interestingly, nad2 showed an exceptionally high relative proportion of nonsynonymous changes and higher Ka/Ks ratio compared to the other mitochondrial coding genes. This pattern was also observed in vertebrate mitochondrial genomes, e.g. fishes, and may be associated with the distance from the origin of replication. Thus, during replication, it will exposed as single-stranded for longer time compared to the other genes, rendering it more likely to accumulate mutations in the highly mutagenic environment of the mitochondrial DNA. Although all ratios less than 1 is consistent with purifying selection, the Ka/Ks ratio close to 1 is unusual for mt genes, positive selection cannot be ruled out entirely in nad2 gene.

Identification of the rrnL genes in T. kiyoni and P. pilula were accomplished by comparison with other Arcidae mtDNAs. A conserved 23 bp-long sequence ‘AGGAGTACGGGAAACGTGCTCCT’ was used to identify the 3′ end of rrn5 gene in T. kiyoni and P. pilula. This motif was conserved and also reported as the basis to infer the 3′ end of rrn5 in other Arcidae mitogenomes. The length of rrnL is 1,479 bp, and the rrn5 is 710 bp in T. kiyoni. They are the largest rRNA genes yet reported in the family Arcidae. In P. pilula, the rrnL and rrn5 are 1,344 bp and 673 bp in length, respectively (Table 1).

Transfer RNA genes and anticodons. The complete set of 22 tRNA genes typical of metazoan mitochondrial genomes were present in T. kiyoni: two tRNAs for each of serine and leucine, and one tRNA for each of the other 18 amino acids. The P. pilula mitogenome contained 23 tRNAs, including the standard 22 tRNAs set and an extra trnR. All tRNAs were interspersed between the tRNAs and the protein-coding genes with the ranges from 64 bp (trnLSCG) to 74 bp (trnL) in T. kiyoni and ranged from 64 bp (trnR2) to 73 bp (trnW) in P. pilula. The predicted secondary structures of tRNAs in two Arcidae mitogenomes were shown in Supplementary Figs 1 and 2. Most of them can fold into canonical clover-leaf secondary structures except trnE and trnSAGA in T. kiyoni, trnRSCG and trnSAMGA in P. pilula, whose paired “DHU” arm were missing, simplifying down to a loop. A modified DHU-arms of trnE in T. kiyoni is unique among molluscs. The modified DHU-arms of trnRSCG is present in only few mitochondrial genomes. However, missing of the “DHU” arm in the secondary structure of the trnS gene- trnSRSC and trnSAMGA is common for molluscs. To work in a similar way as usual tRNAs, these aberrant tRNA genes may require coevolved interacting factors or post-transcriptional RNA editing.

In vertebrate mtDNAs, the most used codon in a degenerate codon family perfectly matches the anticodon of the corresponding tRNA, which is called codon-anticodon adaptation (also known as optimal codon usage). Different from the vertebrate mitochondrial genomes, non-optimal codon usage was the characteristic of Arcidae mtDNAs, and presumably other bivalves, where the most used codon does not perfectly match the corresponding tRNA anticodon in the 22 degenerate codon families (Table 4). This codon usage bias may disrupted by the A + T mutation pressure of the mitogenomes. In addition, the mitogenomes of T. kiyoni and P. pilula shared the same tRNA anticodons with vertebrate (Table 4) suggesting the anticodon evolution in metazoan mitochondrial genomes could be under the same operational forces. This result is not consistent with the hypothesis that the biased codon usage drives the evolution of tRNA anticodons in the vertebrate mitogenome.

| Feature | (A + T)% | AT skew | GC skew |
|---------|----------|---------|---------|
| T. kiyoni | P. pilula | T. kiyoni | P. pilula |
| Whole genome | 62.73 | 61.16 | −0.30 | −0.15 | 0.45 | 0.42 |
| Protein-coding genes | 61.63 | 60.36 | −0.39 | −0.36 | 0.43 | 0.42 |
| cox1 | 60.13 | 60.93 | −0.37 | −0.32 | 0.29 | 0.28 |
| cox2 | 61.50 | 57.01 | −0.27 | −0.09 | 0.36 | 0.29 |
| cox3 | 60.11 | 59.16 | −0.46 | −0.28 | 0.38 | 0.42 |
| Cyb | 63.26 | 62.29 | −0.39 | −0.31 | 0.37 | 0.31 |
| nad1 | 60.51 | 58.52 | −0.38 | −0.29 | 0.44 | 0.41 |
| nad2 | 61.92 | 63.58 | −0.36 | −0.25 | 0.49 | 0.35 |
| nad3 | 59.91 | 58.47 | −0.44 | −0.44 | 0.48 | 0.39 |
| nad4 | 61.30 | 61.99 | −0.43 | −0.33 | 0.48 | 0.49 |
| nad5L | 60.44 | 62.22 | −0.28 | −0.25 | 0.69 | 0.63 |
| nad5 | 62.58 | 58.80 | −0.40 | −0.19 | 0.47 | 0.52 |
| nad6 | 62.75 | 61.68 | −0.53 | −0.28 | 0.58 | 0.66 |
| atp6 | 64.72 | 60.78 | −0.41 | −0.26 | 0.51 | 0.43 |
| tRNAs | 58.38 | 53.45 | −0.19 | −0.07 | 0.35 | 0.30 |
| rrnS | 55.36 | 54.08 | 0.01 | 0.05 | 0.22 | 0.20 |
| rrnL | 62.61 | 61.46 | −0.09 | 0.05 | 0.39 | 0.32 |
| NCR | 67.15 | 63.09 | −0.25 | −0.08 | 0.59 | 0.46 |

Table 3. AT-content, AT-skew and GC-skew for mitochondrial genes of Trisidos kiyoni and Potia rca pilula.
Non-coding regions. 36 non-coding regions, totaling 5,369 bp, were interspersed throughout the T. kiyoni mitogenome; the corresponding values were 34 and 13,642 bp for P. pilula. The non-coding sequences are

Table 4. Codon usage of Trisidos kiyoni and Potiarca pilula PCGs. N: number of occurrence of the codon. RSCU: relative synonymous codon usage.

| Amino acid Code | N(RSCU) T. kiyoni | N(RSCU) P. pilula | Amino acid Code | N(RSCU) T. kiyoni | N(RSCU) P. pilula |
|-----------------|-------------------|-------------------|-----------------|-------------------|-------------------|
| T. kiyoni       | P. pilula         |
| F (gaa)         | 349(1.59)         | 281(1.71)         | Y (gta)         | 165(1.63)         | 129(1.61)         |
| TTC             | 89(0.41)          | 47(0.29)          | TAT             | 38(0.37)          | 31(0.39)          |
| L (taa)         | 76(1.82)          | 158(1.88)         | Stop            | 64(0.85)          | 86(1.19)          |
| TTG             | 105(2.51)         | 160(1.90)         | TAG             | 87(1.15)          | 59(0.81)          |
| L (tag)         | 46(1.10)          | 79(0.94)          | H (gtg)         | 17(1.21)          | 65(1.48)          |
| CTC             | 17(0.41)          | 20(0.24)          | CAT             | 11(0.79)          | 32(0.52)          |
| CTA             | 3(0.07)           | 50(0.59)          | Q (tgq)         | 7(0.70)           | 29(0.89)          |
| CTG             | 4(0.10)           | 38(0.45)          | CAG             | 13(1.30)          | 36(1.11)          |
| I (gat)         | 90(1.68)          | 121(1.66)         | N (gtt)         | 106(1.64)         | 49(1.61)          |
| ATC             | 17(0.32)          | 25(0.34)          | AAT             | 23(0.36)          | 12(0.39)          |
| M (cat)         | 21(0.86)          | 71(0.84)          | K (ttt)         | 35(0.80)          | 49(0.95)          |
| ATG             | 28(1.14)          | 99(1.16)          | AAA             | 52(1.20)          | 54(1.05)          |
| Y (tyc)         | 208(2.26)         | 130(1.55)         | D (gtc)         | 87(1.67)          | 52(1.46)          |
| GTC             | 34(0.37)          | 29(0.35)          | GAC             | 17(0.33)          | 19(0.54)          |
| GTA             | 56(0.61)          | 66(0.79)          | E (tcg)         | 26(0.93)          | 65(0.95)          |
| GTG             | 70(0.76)          | 111(1.32)         | GAG             | 30(1.07)          | 72(1.05)          |
| S (tga)         | 65(1.22)          | 55(1.48)          | C (gca)         | 185(1.50)         | 83(1.63)          |
| TCT             | 51(0.96)          | 19(0.51)          | TGT             | 61(0.50)          | 19(0.37)          |
| TCA             | 75(1.41)          | 32(0.86)          | TGA             | 90(0.66)          | 90(0.88)          |
| TCG             | 44(0.83)          | 16(0.43)          | TGG             | 181(1.34)         | 115(1.12)         |
| P (tgg)         | 11(1.05)          | 46(1.67)          | R (tgg)         | 10(0.83)          | 42(1.68)          |
| CCC             | 18(1.71)          | 21(0.76)          | CGG             | 70(0.58)          | 80(0.32)          |
| CCA             | 10(0.95)          | 25(0.91)          | CGA             | 6(0.50)           | 15(0.60)          |
| CGG             | 3(0.29)           | 18(0.65)          | CGG             | 25(2.08)          | 35(1.40)          |
| T (tgt)         | 9(0.88)           | 41(1.91)          | S (tct)         | 60(1.13)          | 43(1.15)          |
| ACC             | 11(1.07)          | 8(0.37)           | AGT             | 34(0.64)          | 13(0.35)          |
| ACA             | 14(1.37)          | 24(1.12)          | AGA             | 35(0.66)          | 41(1.10)          |
| ACG             | 7(0.68)           | 13(0.60)          | AGG             | 61(1.15)          | 79(2.12)          |
| A (tgc)         | 39(1.46)          | 58(1.97)          | G (tcc)         | 139(1.62)         | 136(1.42)         |
| GCC             | 28(1.05)          | 10(0.34)          | GGC             | 59(0.69)          | 18(0.19)          |
| GCA             | 25(0.93)          | 22(0.75)          | GGA             | 51(0.59)          | 70(0.73)          |
| GCG             | 15(0.56)          | 28(0.95)          | GGG             | 95(1.10)          | 159(1.66)         |

Figure 2. The ratios of nonsynonymous and synonymous substitution (Ka/Ks) estimated in all twelve protein coding genes of six Arcidae species.
generally rare and characterized by fewer nucleotides in *T. kiyoni*. However, it is important to notice the presence of a relatively large non-coding region in the *P. pilula* mtDNA. The A + T content of non-coding regions in *T. kiyoni* and *P. pilula* are 67.15% and 63.09%, respectively, with both negative AT-skew (−0.25 and −0.08) and positive GC-skew (0.59 and 0.46).

The largest non-coding region with increased A + T composition is considered as the control region as it usually contains the signals for replication and transcription. It shows a higher size variation than the other regions of the mitogenome due to both length variation with tandem repeat units (TRs) and differences in their copy numbers. In the *T. kiyoni* mitogenome, one 770-bp tandem repeat (10,333–11,102), comprising three nearly identical motifs was found in the largest non-coding region (1,009 bp) between *trnF* and *trnSUCU*. Most of the non-coding sequences (9,843 bp) were observed within one segment in the *P. pilula* mitogenome, within this segment all of the sequence, except *trnRCGA*, were predicted to be non-coding DNA. The large concentrated non-coding region of *P. pilula*, contained two distinct tandem repeat units (19,804–20,936 and 25,978–26,585), which were 1,133 bp and 668 bp in length, respectively. The first repeat family contained four nearly identical motifs. The second one had a three identical copies and a third copy with a 40% length of a 180-base sequence. Tandem repeat units within non-coding regions seem a common feature in Arcidae mitogenomes, despite different length and copy number in the repeat units. The tandem repeat region was also found in other molluscs. Stem-loop structures were detected in the tandem repeat region of *T. kiyoni* and *P. pilula* (Supplementary Fig. 3). It has been demonstrated that the potential stem-loop structures in repeated units and its flanking part may cause an increase in slipped-strand mispairing frequency.

**Varied genome size of Arcidae species.** Arcidae mitochondrial genomes are exceptional for their extraordinarily large and highly variable sizes. They house by far the largest known metazoan mitochondrial genomes, with sizes ranging from 19.6 to 47 kb among the four genomes sequenced to date (https://www.ncbi.nlm.nih.gov/). Arcidae mitogenomes possess an average length of 34.5 kb, whereas *T. kiyoni* and *P. pilula* showed the length of 19,614 bp and 27,895 bp, respectively, which was the smallest characterized mitochondrial genomes in Arcidae. The smallest genome-size is weakly associated with gene number and primarily reflect shrinkage of the non-coding regions. Genomic coverage by mitochondrial non-coding regions are only 27.37% for *T. kiyoni* and 40.84% for *P. pilula*, which were much lower than that of other Arcidae.

The early diverging phylogenetic positions of *T. kiyoni* within the Arcidae is such that this species provides an important insight into the historical information of Arcidae mitochondrial genomes (Fig. 3). Although it is difficult to reconstruct with the limited genomes, however, the diversity of mitogenome size among the species appears to reflect a dynamic history of expansion. The common ancestor of Arcidae, like *T. kiyoni*, might have possessed a relatively compact mitochondrial genome, with a series of independent expansions leading to the large genomes in other species. However, the sources underlying major expansions in mitochondrial genome size are unknown.

Already in 1991 it was reported that the length of mitochondrial rRNAs is correlated with the size of their corresponding organellar genomes in seven species. Highly significant positive correlations were detected between mitochondrial genome size and the combined length of *cox1*-3, the length of *Cytb*, and the combined
length of rRNAs (rrnS and rrnL) in 278 eukaryotes and 11 a-proteobacteria. The six mitochondrial genes are essential for oxidative phosphorylation, which in most species are refractory to nuclear transfer54. We presented here an analysis of this observation for 256 molluscs using six mitochondrial genes (cox1-3, Cytb, rrnS and rrnL). They have rarely been transferred to nucleus and are therefore well suited to test the hypothesis on the evolution of gene length in mitochondria54. A significant positive correlations are observed between the size of their mitochondrial genome and the combined length of cox1-3, the lengths of Cytb, and the combined length of rRNAs (rrnS and rrnL), which is consistent with former reports (Fig. 4, Supplementary Table 2). In many mitochondrial genomes, redox reactions produce oxygen free radicals during respiration, making a higher mutation rate than their corresponding nuclear DNA55. Müller’s ratchet states that these deleterious mutations can accumulate and lead to a mutational meltdown if recombination (either within or between organelles) never occurs56. Müller’s ratchet explains that the shorter genes may accumulate slightly deleterious mutations slower54. Further, the replication advantage hypothesis states that a smaller mitochondrial genome would be selected in intracellular

Figure 4. Covariation of complete mitochondrial genome size with gene lengths. Gene lengths of cox1-3 (A), Cytb (B), and rRNA (rrnS + rrnL) (C) of different species are plotted against the corresponding complete mitochondrial genome size. The lines have been fitted by linear regressions. For species included in the analysis see supporting table in the Supplementary Table 2.
stored at −20 °C.

**Gene arrangement as a novel structure.** The mitogenomes of Bivalvia show substantially gene rearrangements, having no obvious common pattern in the arrangement of PCGs and rRNA. Species sequenced in Bivalvia belong to three five subclasses: Palaeoheterodonta, Heterodonta, Pteriomorphia, Anomalodesmata and Protobranchia. Gene arrangement in Unionoida (Palaeoheterodonta) is relatively conserved, except for the translocation of several tRNAs, and protein-coding genes nad2 and nad3. In addition, only one mitochondrial genome is available in both Anomalodesmata (Laternula elliptica, KF534717) and Protobranchia (Solemya venum, JQ728447), and more sequences are need for further analyze. The mitochondrial gene order of the remaining bivalves is frequently rearranged, but see oysters especially for the family Pectinidae.

Although variability in gene arrangement is high, there are some conserved gene blocks within these groups. Arcidae seems to represent another example, as T. kiyoni mitogenome shared no gene block with any other five Arcidae species, despite being number of the same family, suggesting that gene rearrangements occurred dramatically among lineages in this family (Fig. 5). There were both PCGs and tRNA rearrangement in P. pilula mitogenome. In terms of gene arrangement, it is clear that P. pilula is more similar to A. vellicata and T. granosa than to S. broughtonii and S. kagoshimensis. They share three identical gene blocks: two large blocks cox1-nad5-trnM-nad1-nad4-Cytb-trnf-coc-trnC-nad6-trnK and atp6-trnP-trnL-trnG-trnV-trnL-trnA-trnI-trnH-trnQ-nad3-nad4L, and one small block trnY-trnN-trnS-nad2. If the tRNA genes are not considered, the gene order in P. pilula is nearly identical to that of A. vellicata, T. granosa, S. broughtonii and S. kagoshimensis, except for the translocation of gene cox3. For generation of the gene arrangement of mt DNA during evolution, a model involving slipped-strand mispairing of two homologous regions and random gene loss was proposed. We suspect that the asymmetric gene replication and transcription accelerate this phenomenon in the evolutionary process. This hypothesis is supported by the fact that all of the mt genes of marine bivalves are encoded on same strand and gene length. It is also supported by the shape of the gene length-genome size relationship investigated here, which is strongly asymptotic for all gene.

Gene arrangement comparisons may be a useful tool for phylogenetic studies. This is based on the hypothesis that gene arrangements are likely to be shared only as a result of common ancestry since it is highly unlikely that the same gene order would arise independently in separate lineages. We present a schematic representation of mitochondrial gene arrangements in Arcidae species on the phylogenetic trees inferred from the nucleotide dataset of 12 PCGs (Fig. 5). The comparative analysis of mt gene rearrangements in Arcidae reinforces the validity of our ML-tree and contributes new information on Arcidae phylogenetic relationships. As shown in Fig. 5, T. kiyoni was in a separate, more ancestral branch in the phylogenetic tree. Its gene order may represent the pleiomorphic gene arrangement in Arcidae. Hence, our analysis imply that the complicated gene rearrangement in mitochondrial genome could be considered as one of key characters in inferring higher-level phylogenetic relationships of Arcidae.

**Materials and Methods**

**Sample collection and DNA extraction.** Specimens of T. kiyoni and P. pilula were collected from the coastal water of Fujian Province, China. These samples were stored at −80 °C and deposited as voucher specimens (specimen number: TK01 and PP01) in Fisheries College, Ocean University of China. Each of the two Arcidae complete mitogenome sequenced was obtained from a single specimen. Total genomic DNA was extracted from adductor muscle by a modification of standard phenol-chloroform procedure as described by Li et al. and then stored at −20 °C.
Determination of partial sequences. In order to design gene-specific primers, we first obtained partial \( \text{cox}1 \) sequences for both \( T. \) \( \text{kiyoni} \) and \( P. \) \( \text{pilula} \), with the universal primers of LCO1490/HCO2198\(^{62} \). Another short fragment, \( \text{rrnS} \) genes, were obtained from NCBI data base (GenBank accession no. JN974675 for \( T. \) \( \text{kiyoni} \) and JN974660 for \( P. \) \( \text{pilula} \)).

Construction of BD GenomeWalker DNA libraries, PCR amplification and sequencing. Four BD GenomeWalker DNA libraries were constructed with the BD GenomeWalker Universal Kit (BD Biosciences, San Jose, CA, USA) following the manufacturer’s protocols.

The complete mitogenome of \( T. \) \( \text{kiyoni} \) and \( P. \) \( \text{pilula} \) were amplified using genome-walking based method, which involves two nested PCR reactions with a touch-down program modified from the BD GenomeWalker Universal Kit User Manual. The partial sequences of \( \text{cox}1 \) and \( \text{rrnS} \) were used to design the initial sets of gene-specific primers, one (GSP1) for original PCR and the other (GSP2) for nested PCR, which were used for genome-walking to amplify both of the Arcidae mitogenome. The primer sequences used for genome-walking are presented in Supplementary Tables 3 and 4.

PCR was performed in a total volume of 50 \( \mu \)l including 2 U \( Taq \) DNA polymerase (TaKaRa, Dalian, China), about 100 ng template DNA, 1 \( \mu \)l forward and reverse primers, 200 \( \mu \)M of each dNTP, 1 × PCR buffer and 2 mM \( \text{MgCl}_2 \). The original PCR were carried out as follows using the outer adaptor primer1 (AP1) and outer gene-specific primer1 (GSP1): 10 s initial denaturation at 94 °C, 7 cycles of 30 s at 94 °C, 3 min at 72 °C, 32 cycles 30 s at 94 °C, 3 min at 67 °C, and 67 °C for an additional 7 min after the final cycle. A 1-\( \mu \)l sample of the original PCR was diluted in 59 \( \mu \)l of distilled water as the template for nested PCR amplification. The nested PCR were carried out as follows using the outer adaptor primer2 (AP2) and the outer, gene-specific primer2 (GSP2): 10 s initial denaturation at 94 °C, 5 cycles of 30 s at 94 °C, 3 min at 72 °C, 25 cycles 30 s at 94 °C, 3 min at 67 °C, and 67 °C for an additional 7 min after the final cycle. This procedure generally produces a single, major PCR product (100 bp–5000 bp) in at least one of the four libraries, which begins in a known sequence at the 5’ end of GSP2 and extends into the unknown adjacent genomic DNA.

PCR products were purified with 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 in Beijing Genomics Institute (BGI) using standard Sanger sequencing chemistry.

Sequencing assembling and annotation. All sequence data were analysed and arranged to create the full genomes using the Seqman program from DNASTAR (http://www.DNASTAR.com). The protein coding genes were analyzed with ORF Finder (http://www.ncbi.nlm.nih.gov/orf/orf.html) and BLASTx using the invertebrate mitochondrial genetic code. The tRNA genes were identified by ARWEN\(^{63} \) and DOGMA\(^{64} \) using the mito/chloroplast or invertebrate genetic code and the default search mode. The rRNA genes were identified by their similarity to published gene sequences and by using BLAST searches (http://www.ncbi.nlm.nih.gov/BLAST/).

The base composition and skewness analyses were performed and compared between \( T. \) \( \text{kiyoni} \) and \( P. \) \( \text{pilula} \) genomes, as well as the other four Arcidae genomes (\( S. \) \( \text{broughtonii} \) (46,985 bp), \( S. \) \( \text{kagoshimensis} \) (46,713 bp), \( T. \) \( \text{granosa} \) (31,589 bp)\(^{14} \) and \( A. \) \( \text{vellicata} \) (34,147 bp)\(^{13} \)). The \( A + T \) content values were computed using Editseq program from DNASTAR. The GC and AT skews described strand bias were calculated according to the formulae by Perna and Kocher\(^{65} \), AT skew = \((A−T)/(A+T)\); GC skew = \((G−C)/(G+C)\), where \( A, T, G \) and \( C \) are the occurrences of the four nucleotides. The codon usage of each PCG were calculated using MEGA 5\(^{66} \). The ratios of nonsynonymous and synonymous substitutions rates (Ka/Ks) were estimated based on the Maximum-Likelihood (ML) method\(^{27} \) using KaKs_Calculator 2.0\(^{68} \) with the YN model.

The whole mitogenome sequence was tested for potentially tandem repeats by Tandem Repeats Finder 4.0\(^{69} \). Prediction of potential secondary structure was performed by the online version of the mfold software, version 3.2\(^{10} \), applying default settings. When multiple secondary structures were possible, the most stable (lowest free energy (\( ΔG)\)) was used.

The gene map of the \( T. \) \( \text{kiyoni} \) and \( P. \) \( \text{pilula} \) mitogenomes were generated with the program CGView\(^{71} \). The two mitochondrial genomes have been deposited in the GenBank database under the accession numbers KU975161 for \( T. \) \( \text{kiyoni} \) and KU975162 for \( P. \) \( \text{pilula} \).

Predicted lengths of gene products and mitogenome sizes for up to 278 molluscs (see Supplementary Table 2). The statistical analysis was performed by using IBM SPSS Statistics 19 with Spearman rank correlations, as this test makes no assumption about the distribution of the data.

Phylogenetic analyses. Along with mitochondrial genome sequence of \( T. \) \( \text{kiyoni} \) and \( P. \) \( \text{pilula} \), all currently available mitochondrial genomes from Arcidae, including \( S. \) \( \text{broughtonii} \) (AB729113), \( S. \) \( \text{kagoshimensis} \) (KF750628), \( T. \) \( \text{granosa} \) (JK607173) and \( A. \) \( \text{vellicata} \) (KP954700), were used in phylogenetic analysis.

The phylogenetic relationships were built based on the nucleotide sequences of 12 PCGs. Crassostrea gigas (AF177226) and Crassostrea hongkongensis (EU266073) from the family Ostreidae was used as outgroup. The twelve-partitioned nucleotide sequences of protein coding genes were aligned with MAFFT based on their nucleotide sequences using default settings\(^{22} \). The final nucleotide sequences of each gene were then concatenated into single contigs (6719 bp) for phylogenetic analyses. The best-fit nucleotide substitution models for each data partitions were selected by jModelTest\(^{22} \). We employed ML in RAxML Black-Box webserver (http://phylobench.vital-it.ch/raxml-bb/index.php)\(^{24} \) with GTR + G substitution model to each partition. For the ML analysis, 1000 bootstraps were used to estimate the node reliability.
References

1. Boore, J. L. Animal mitochondrial genomes. Nucleic Acids. Res. 27, 1767–1780 (1999).
2. Wolstenholme, D. R. Animal mitochondrial DNA: structure and evolution. Int. Rev. Cytol. 141, 173–216 (1992).
3. Liu, Y. G., Kurakov, T., Sekino, M., Tanabe, T. & Watanabe, K. Complete mitochondrial DNA sequence of the ark shell Scapharca brounhotica: an ultra-large metazoan mitochondrial genome. Comp. Biochem. Physiol. D: Genomics Proteomics. 8, 72–81 (2013).
4. Sun, S. E., Kong, L., Yu, H. & Li, Q. The complete mitochondrial genome of Scapharca kagogiominisis (Bivalvia: Arcidae). Mitochondrial DNA. 26, 957–958 (2014).
5. Burger, G., Gray, M. W. & Lang, B. F. Mitochondrial genomes: anything goes. Trends. Genet. 19, 709–716 (2003).
6. Simison, W. B. & Boore, J. L. Molluscan evolutionary genomics. In Phylogeny and evolution of the mollusca. Edited by Ponder, W. & Lindberg, D. R. Berkeley: University of California Press, 447–461 (2008).
7. Shadel, G. S. & Clayton, D. A. Mitochondrial DNA maintenance in vertebrates. Annu. Rev. Biochem. 66, 409–435 (1997).
8. Gissi, C., Iannelli, F. & Pesole, G. Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. Heredity. 101, 301–320 (2008).
9. Stöger, I. & Schrödl, M. Mitogenomics does not resolve deep molluscan relationships (yet?). Mol. Phylogenet. Evol. 69, 376–392 (2013).
10. Uliano-Silva, M. et al. The complete mitochondrial genome of the golden mussel Linnunaperna fortunei and comparative mitogenomics of Mytilidae. Gene, 577, 202–208 (2016).
11. Smith, D. R. & Snyder, M. Complete mitochondrial DNA sequence of the eastern oyster Crassostrea virginica. Mol. Phylogenet. Evol. 5, 1535–1554 (2013).
12. Xu, K. F., Kanno, M., Yu, H., Li, Q. & Kijima, A. Complete mitochondrial DNA sequence and phylogenetic analysis of Zhikong scallop Chlamys farreri (Bivalvia: Pectinidae). Mol. Biol. Rep. 38, 3067–3074 (2011).
13. Passentini, M. & Scali, V. Gender-associated mitochondrial DNA heteroplasmy in the venerid clam Tapes philippinarum (Mollusca Bivalvia). Curr. Genet. 39, 117–124 (2001).
14. Sun, S. E., Kong, L., Yu, H. & Li, Q. The complete mitochondrial DNA of Tegillarca granosa and comparative mitogenomic analyses of three Arcidae species. Gene, 557, 61–70 (2015a).
15. Smith, D. R. & Snyder, M. Complete mitochondrial genome of Anadara vellicata (Bivalvia: Arcidae): A unique gene order and large atypical non-coding region. Comp. Biochem. Physiol. D: Genomics Proteomics. 16, 73–82 (2015b).
16. Breton, S., Beaupré, H. D., Stewart, D. T., Hoch, W. R. & Blier, P. U. The unusual system of doubly uniparental inheritance of mtDNA: isn't one enough? Trends. Genet. 23, 465–474 (2007).
17. Ghiselli, F. et al. Structure, transcription, and variability of metazoan mitochondrial genome: perspectives from an unusual mitochondrial inheritance system. Genome. Biol. Evol. 5, 1355–1354 (2013).
18. Zouros, E. Biparental inheritance through uniparental transmission: the doubly uniparental inheritance (DUI) of mitochondrial DNA. Evol. Biol. 40, 1–31 (2013).
19. Milbury, C. A. & Gaffney, P. M. Complete mitochondrial DNA sequence of the eastern oyster Crassostrea virginica. Mar. Biotechnol. 7, 697–712 (2005).
20. Ren, J., Liu, X., Jiang, F., Guo, X. & Liu, B. Unusual conservation of mitochondrial gene order in Crassostrea oysters: evidence for recent speciation in Asia. BMC Evol. Biol. 10, 394 (2010).
21. Xu, K. F., Kanno, M., Yu, H., Li, Q. & Kijima, A. Complete mitochondrial DNA sequence and phylogenetic analysis of Zhikong scallop Chlamys farreri (Bivalvia: Pectinidae). Mol. Biol. Rep. 38, 3067–3074 (2011).
22. Wu, X. Y., Xu, X. D., Yu, Z. N., Wei, Z. P. & Xia, J. J. Comparison of seven Crassostrea mitogenomes and phylogenetic analyses. Mol. Phylogenet. Evol. 57, 448–454 (2010).
23. Serb, J. M. & Lydeard, C. Complete mtDNA Sequence of the North American Freshwater Mussel, Lampsilis ornata (Unionidae): An Examination of the Evolution and Phylogenetic Utility of Mitochondrial Genome Organization in Bivalvia (Mollusca). Mol. Biol. Evol. 20, 1854–1866 (2003).
24. Oliver, P. G. & Holmes, A. M. The Arcoidea (Mollusca: Bivalvia): a review of the current phenetic-based systematics. Zool. J. Linnean Soc. 148, 237–251 (2006).
25. Newell, N. D. Order Arcida Stoliczka, 1871. In: Moore, R. C. (Ed.), Treatise on invertebrate paleontology, Part N, Mollusca 6, Bivalvia, vol. 1 (of 3). Geological Society of America and University of Kansas, pp. N228–N270 (1969).
26. Beesley, P. L., Ross, G. J. & Wells, A. Mollusca: The Southern Synthesis. CSIRO Publishing, Melbourne, 1234 pp (1998).
27. Hassanin, A., Léger, N. & Deutsch, J. Evidence for multiple reversals of asymmetric mutational constraints during the evolution of the mitochondrial genome of Metazoa, and consequences for phylogenetic inferences. Syst. Biol. 54, 277–298 (2005).
28. Song, S. N., Tang, P., Wei, S. J. & Chen, X. Comparative and phylogenetic analysis of the mitochondrial genomes in basal hymenopterans. Sci. Rep. 6, 20972 (2016).
29. Reyes, A., Gissi, C., Pesole, G. & Saccone, C. Asymmetrical directional mutation pressure in the mitochondrial genome of mammals. Mol. Biol. Evol. 15, 957–966 (1998).
30. Xia, X. Mutation and selection on the anticodon of tRNA genes in vertebrate mitochondrial genomes. Gene, 345, 13–20 (2005).
31. Ohtia, T. Synonymous and nonsynonymous substitutions in mammalian genes and the nearly neutral theory. J. Mol. Evol. 40, 56–63 (1995).
32. Fay, J. C. & Wu, C. I. Sequence divergence, functional constraint, and selection in protein evolution. Annu. Rev. Genomics. Hum. Genet. 4, 213–235 (2003).
33. Saccone, C., De Giorgi, C., Gissi, C., Pesole, G. & Reyes, A. Evolutionary genomics in Metazoa: the mitochondrial DNA as a model system. Gene, 238, 195–209 (1999).
34. Jacobson, M. W., da Fonseca, R. R., Bernatchez, L. & Hansen, M. M. Comparative analysis of complete mitochondrial genomes suggests that relaxed purifying selection is driving high nonsynonymous evolutionary rate of the NADH2 gene in whitefish (Coregonus sp.). Mol. Phylogenet. Evol. 95, 161–170 (2016).
35. Neidhal, M. A. & Flynn, J. J. Do the combined effects of the asymmetric process of replication and DNA damage from oxygen radicals produce a mutation-rate signature in the mitochondrial genome? Mol. Biol. Evol. 15, 219–223 (1998).
36. Ren, J. E., Shen, X. & Jiang, F. The Mitochondrial Genomes of Two Scallops, Argoplectus varioides and Chlamys farreri (Mollusca: Bivalvia): The Most Highly Rearranged Gene Order in the Family Pectinidae. J. Mol. Evol. 70, 57–68 (2010).
37. Dreyer, H. & Steiner, G. The complete sequence and gene organization of the mitochondrial genome of the girdled scaphopod Siphonodentalium lobatum (Mollusca). Mol. Phylogenet. Evol. 31, 605–617 (2004).
38. Dreyer, H. & Steiner, G. 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, 1–14 (2006).
39. Wang, H., Zhang, S., Li, Y. & Liu, B. Complete mtDNA of 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 (2010).
40. Wang, H., Zhang, S., Xiao, G. & Liu, B. Complete mtDNA of the Meretrix lamarkii (Bivalvia: Veneridae) and molecular identification of suspected M. lamarkii based on the whole mitochondrial genome. Mar. Genom. 4, 263–271 (2011).
41. Chimmenarak, S., Jeppesen, M. G., Suzuki, T., Nyborg, I. & Watanabe, K. Dual-mode recognition of noncanonical tRNAs by seryl-tRNA synthetase in mammalian mitochondria. EMBO J. 24, 3369–3379 (2005).
Complete mitochondrial genomes of Trisidos kiyoni

The authors declare no competing financial interests.

Varied mitochondrial genome size and highly rearranged gene order in Arcidae. Sci. Rep. 6, 33794; doi: 10.1038/srep33794 (2016).

Author Contributions
Q.L. and S.S. conceived and designed the experiments, S.S., L.K. and H.Y. collected the data and performed the experiments, Q.L. and S.S. analysed the data and wrote the paper, and all authors were involved in the critical review of the manuscript.

Additional Information

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Sun, S. et al. Complete mitochondrial genomes of Trisidos kiyoni and Potiaatra pilala: Varied mitochondrial genome size and highly rearranged gene order in Arcidae. Sci. Rep. 6, 33794; doi: 10.1038/srep33794 (2016).
This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/

© The Author(s) 2016