Genes Genom
DOI 10.1007/s13258-017-0548-1

Online ISSN 2092-9293
Print ISSN 1976-9571

RESEARCH ARTICLE

Complete mitochondrial genome sequence of *Cucullaea labiata* (Arcoida: Cucullaeidae) and phylogenetic implications

Yanwei Feng[1,2] · Qi Li[2] · Hong Yu[2] · Lingfeng Kong[2]

Received: 6 March 2017 / Accepted: 13 April 2017

Abstract The complete mitochondrial genome of *Cucullaea labiata* (Arcoida: Cucullaeidae) was firstly determined in this study in order to better understand the phylogenetic relationship between Cucullaeidae and Arcidae. The *C. labiata* mitochondrial genome was 25,845 bp in size and contained 12 protein-coding genes, 2 rRNA and 22 tRNA genes. The number and the location of the tRNA genes were different from three Arcidae species (*Scapharca broughtonii, Scapharca kagoshimensis* and *Tegillarca granosa*). Gene arrangement also differed dramatically. The length of the non-coding regions was 10,559 bp, in which the largest one (6057 bp) included eight point nine copies of a 659 bp repeat motif. The number of repeated sequences was different in different individuals, similar to the findings from the mitochondrial genome of *S. broughtonii* and *Placopecten magellanicus*. One intron was found in *cox1* gene both in CL_98 and in CL_99 individuals of *C. labiata*. The reason why mitochondrial introns are retained so scarcely in bivalve taxa needs further research. Phylogenetic analyses based on 12 concatenated amino acid sequences of protein-coding genes supported Cucullaeidae was the sister group of Arcidae.

Keywords Mitochondrial genome · *Cucullaea labiata* · Cucullaeidae · Intron · Phylogenetic relationships

✉ Qi Li
qili66@ouc.edu.cn

[1] Shandong Provincial Key Laboratory of Marine Ecology Restoration, Shandong Marine Resource and Environment Research Institute, Yantai 264006, China

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

Introduction

Metazoan mitochondrial (mt) DNA is the only extranuclear genome and forms a unit of genetic information. Most metazoan mitochondrial genomes are usually circular double-stranded molecules containing 13 protein-coding genes (*cox1-3, cytb, atp6, atp8, nad1-6* and *nad4l*), two ribosomal RNAs (*rrnL* and *rrnS*), 22 transfer RNA genes (*tRNA*s) and a single large non-coding region containing elements that control the initiation of replication and transcription (Wolstenholme 1992; Shadel and Clayton 1997). However, as more and more mt genome sequences are available, many mtDNAs deviate from this model such as some arcidae species *Scapharca broughtonii* (Liu et al. 2013), *Scapharca kagoshimensis* (Sun et al. 2014) and *Tegillarca granosa* (Sun et al. 2015).

Size variation, different gene arrangements and tandem repeated sequences are common features of metazoan mtDNA (Lunt et al. 1998; Boore and Brown 2000). The size of metazoan mtDNA ranged from 14 kb to about 50 kb. The mtDNA of *S. broughtonii* is, to date, the longest reported metazoan mtDNA sequence with a maximum length of about 50 kb (Liu et al. 2013). Gene arrangements are usually conserved within major lineages (Boore 1999). However, mollusks, especially bivalves, are highly rearranged, which can be seen from currently available mt genomes. Gene arrangement has been shown to be useful phylogenetic markers for evolutionary studies. The tandem repeats described in published reports range in size from small microsatellite-like repeats of several bp to long minisatellite repeats of hundreds of bp to well over 1000 bp (Snyder et al. 1987). Variation in the number of these repeats is responsible for the size of mtDNA.

Mitochondrial genome has several valuable characteristics including maternal inheritance, lack of recombination

Published online: 20 April 2017

and higher evolutionary rates. Because of these, mtDNAs have been extensively used for studying population structure and phylogenetic relationships (e.g. Yamanoue et al. 2007; Yuan et al. 2012). The number of complete mitochondrial genome sequences in public databases has been increasing rapidly in recent years. Nevertheless, only three complete mitochondrial genomes of Arcoid species have been determined.

Arcoid bivalves (Bivalvia: Pteriomorphia: Arcoida) are abundant and diverse in modern seawater. They comprise seven families mainly such as Arcidae, Noetiidae, Cucullaeidae, Glycymerididae and Limopsidae. Phenetic characters form a basis for the present classification of arcoids (Oliver and Holmes 2006), but it was challenged by recent phylogenetic studies based on molecular data (Feng et al. 2015). For example, the family Cucullaeidae did not appear as sister group to the Arcidae, but as subgroup within it. The family Cucullaeidae, which flourished in the late Mesozoic, is represented by a single, widely distributed, Indo-Pacific species, *Cucullaea labiata* (Boss 1982). The Cucullaeidae appear to be contemporary with the Arcidae and both have their origins in the Jurassic (Oliver and Holmes 2006). However, the result of Feng et al. (2015) indicates it may be younger than the Arcidae. The origin of the Cucullaeidae and its relationship with the Arcidae need further studies.

In the present study, we determined the complete mitochondrial genome sequence of Cucullaeidae from the species *C. labiata*, and compared it with the Arcidae species. This study will contribute to getting insights into the phylogenetic position of Cucullaeidae, as well as better understanding the relationship between Cucullaeidae and Arcidae.

Materials and methods

Sample collection and DNA extraction

The original *C. labiata* sample (CL_98) was collected from Beihai, Guangxi province, China. Additional one specimen (CL_99) was from Lingshui, Hainan province, China. The total genomic DNA was extracted from adductor muscle following a modification of the standard phenol–chloroform procedure described by Li et al. (2002) and visualized on 1.0% agarose gel.

PCR amplification and DNA sequencing

The complete mitochondrial genome of CL_98 was sequenced using long-PCR, followed by a mixed of shotgun sequencing and primer walking. To design long-PCR primers, partial *cox1, rrnL, cytb* and *cox3* gene sequences were amplified using the universal primers of cox1F/cox1R (Matsumoto 2003), 16Sar/16Sbr (Palumbi 1996), CytbF/CytbR and cox3F/cox3R (Boore and Brown 2000), respectively. PCR was performed in a total volume 50 µl including 2 U Taq DNA polymerase (TaKaRa, Dalian, China), approx. 100 ng template DNA, 1 µM forward and reverse primers, 200 µM of each dNTP, 1 × PCR buffer, and 2 mM $MgCl_2$. The PCR reaction was carried out in TaKaRa PCR Thermal Cycler Dice Model TP600 (Takara Bio Inc.) under the following conditions: an initial denaturation at 94 °C for 3 min, followed by 35 cycles of denaturation at 94 °C for 30 s, annealing at 50–53 °C for 1 min, and extension at 72 °C for 1 min, and final extension at 72 °C for 5 min. PCR products were purified and sequenced using an ABI PRISM 3730 automatic sequencer (Applied Biosystems, Foster City, CA, USA).

The sequences of the four short fragments were used to design sets of long range primers (Table 1) to amplify the whole mitochondrial genome. PCR reactions were carried out in 50 µl volume containing 33.5 µl of sterile distilled H_2O, 5 µl of 10 × LA PCR buffer II (Mg^{2+} plus, Takara), 8 µl of dNTP (10 mM each), 1 µl of each primer (10 µM), 0.5 µl of LA Taq polymerase (5 U/µl, Takara), and 1 µl of DNA template (50 ng). The long PCR cycling was set up with an initial denaturation at 94 °C for 2 min, followed by 35 cycles comprising denaturation at 94 °C for 20 s, annealing at 50–68 °C for 30 s and extension at 68 °C for from 7 to 15 min depending on the expected length of the PCR products. The process was completed with a final extension at 72 °C for 10 min. The long PCR products amplified by primer sets cytb-cox1-F/cytb-cox1-R, rrnL-cytb-F/rrnL-cytb-R and cox1-cox3-F/cox1-cox3-R were sequenced by primer walking, and the product amplified by cox3-rrnL-F/cox3-rrnL-R was sequenced by shotgun sequencing approach because of the repeated sequences. The cox3-rrnL and cox1 section of CL_99 was also amplified and sequenced.

Table 1 Primers used for amplification of long fragments in *Cucullaea labiata* mitochondrial genome

Primer name	Sequence (5′-3′)	Annealing T (°C)
cytb-cox1-F	TGTTTTTCCTTGAGGTCAAATGTCTT	53
cytb-cox1-R	CCTTTTTGTACCAAGTTTCTCATCGTT	53
rrnL-cytb-F	GCAGATGTGCGTTGAACCAGGTGTAT	68
rrnL-cytb-R	AATGGATTGGAACCTCCAGTCTCATG	68
cox1-cox3-F	TAGTGGGAACGATGAGAAACTTGGTA	64
cox1-cox3-R	CTGCTCCTTCTTAGTATGTTGACC TTGA	64
cox3-rrnL-F	TCGTGAGTCTAGGTTTCAAGGTCAAC	62
cox3-rrnL-R	ATCCCCAGAGTAACTTATTTCTCCATC	62

Sequence analysis and gene annotation

All sequence data were analyzed and arranged to create the full genome using Seqman in the DNAStar software package (DNASTAR Inc., Madison, WI; Burland 2000). Protein coding genes were defined by ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) using the invertebrate mitochondrial code. The rRNA genes were identified with DOGMA (Wyman 2004) and BLAST searches (http://www.ncbi.nlm.nih.gov/BLAST). The boundaries of each gene were determined with multiple alignments of other published bivalve mitochondrial sequences. The tRNA genes were predicted by DOGMA and tRNAscan-SE Search Server (Lowe and Eddy 1997) with a COVE score cutoff of 1.0 and the invertebrate mitochondrial genetic code for secondary structure prediction. The whole sequence was tested for potentially tandem repeats by Tandem Repeat Finder, Version 4.07b (Benson 1999). Intronic sequences were analyzed using RNAweasel (Lang et al. 2007). InterProScan 5 (http://www.ebi.ac.uk/Tools/pfa/iprscan5/, Zdobnov and Apweiler 2001), Phobius (http://phobius.sbc.su.se/, Käll et al. 2004), and TMpred (http://www.ch.embnet.org/software/TMPRED_form.html, Hofmann and Stoffel 1993) were used to identify putative signal peptides and transmembrane helices. @TOME 2 (http://atome.cbs.cnrs.fr/AT2B/meta.html, Pons and Labesse 2009) was used to find similarities with known proteins.

Phylogenetic analyses

The phylogenetic analyses are based on the complete mt sequences of *C. labiata* (this study), 4 arcoids from the family Arcidae, and other 24 bivalve species from GenBank (Table 2). *Solen grandis, Paphia euglypta* and *Lucinella divaricata* from the subclass Heteroconchia were used as outgroups. Owing to the fact that most bivalve species lack

Table 2 List of the species whose mitochondrial genome sequences were used in phylogenetic analyses

Species	Classification	Number of protein-coding genes	Accession no.
Cucullaea labiata	Bivalvia; Pteriomorphia; Arcoida; Arcoidea; Cucullaeidae	12	KP091889
Scapharca broughtonii	Bivalvia; Pteriomorphia; Arcoida; Arcoidea; Arcidae	12	AB729113
Anadara sativa	Bivalvia; Pteriomorphia; Arcoida; Arcoidea; Arcidae	12	KF667521
Scapharca kagoshimensis	Bivalvia; Pteriomorphia; Arcoida; Arcoidea; Arcidae	12	KF750628
Tegillarca granosa	Bivalvia; Pteriomorphia; Arcoida; Arcoidea; Arcidae	12	KJ607173
Argopecten irradians	Bivalvia; Pteriomorphia; Pectinoida; Pectinoidea; Pectinidae	12	EU023915
Chlamys farreri	Bivalvia; Pteriomorphia; Pectinoida; Pectinoidea; Pectinidae	12	EU715252
Mimachlamys nobilis	Bivalvia; Pteriomorphia; Pectinoida; Pectinoidea; Pectinidae	12	FJ415225
Mizuhopecten yessoensis	Bivalvia; Pteriomorphia; Pectinoida; Pectinoidea; Pectinidae	12	FJ595959
Placopecten magellanicus	Bivalvia; Pteriomorphia; Pectinoida; Pectinoidea; Pectinidae	12	DQ088274
Crassostrea angulata	Bivalvia; Pteriomorphia; Ostreoida; Ostreoidea; Ostreidae	12	EU672832
Crassostrea ariakensis	Bivalvia; Pteriomorphia; Ostreoida; Ostreoidea; Ostreidae	12	EU672835
Crassostrea gigas	Bivalvia; Pteriomorphia; Ostreoida; Ostreoidea; Ostreidae	12	EU672831
Crassostrea hongkongensis	Bivalvia; Pteriomorphia; Ostreoida; Ostreoidea; Ostreidae	12	FJ593172
Crassostrea iredalei	Bivalvia; Pteriomorphia; Ostreoida; Ostreoidea; Ostreidae	12	FJ841967
Crassostrea sikamea	Bivalvia; Pteriomorphia; Ostreoida; Ostreoidea; Ostreidae	12	EU672833
Crassostrea virginica	Bivalvia; Pteriomorphia; Ostreoida; Ostreoidea; Ostreidae	12	AY905542
Crassostrea nippona	Bivalvia; Pteriomorphia; Ostreoida; Ostreoidea; Ostreidae	12	HM015198
Ostrea denselamellosa	Bivalvia; Pteriomorphia; Ostreoida; Ostreoidea; Ostreidae	12	HM015199
Ostrea edulis	Bivalvia; Pteriomorphia; Ostreoida; Ostreoidea; Ostreidae	12	JF274008
Saccostrea mordax	Bivalvia; Pteriomorphia; Ostreoida; Ostreoidea; Ostreidae	12	FJ841968
Mytilus galloprovincialis	Bivalvia; Pteriomorphia; Mytiloida; Mytiloidea; Mytilidae	12	AY497292
Mytilus californianus	Bivalvia; Pteriomorphia; Mytiloida; Mytiloidea; Mytilidae	12	GQ527172
Mytilus edulis	Bivalvia; Pteriomorphia; Mytiloida; Mytiloidea; Mytilidae	12	AY484747
Mytilus trossulus	Bivalvia; Pteriomorphia; Mytiloida; Mytiloidea; Mytilidae	12	AY823625
Musculista senhousia	Bivalvia; Pteriomorphia; Mytiloida; Mytiloidea; Mytilidae	12	GU001953
Solen grandis	Bivalvia; Heteroconchia; Veneroida; Solenoidea; Solenidae	12	HQ703012
Paphia euglypta	Bivalvia; Heteroconchia; Veneroida; Veneroidea; Veneridae	12	GU269271
Lucinella divaricata	Bivalvia; Heteroconchia; Veneroida; Lucinoidea; Lucinidae	13	EF043342

the gene *atp8*, amino acid sequences of 12 concatenated protein-coding genes were used in phylogenetic analyses. The alignment of the amino acid sequences of 12 protein-coding genes was performed using the Clustal W module in BioEdit 7.0.9 (Hall 1999). Areas of dubious alignment were isolated using Gblocks 0.91b (Castresana 2000) and excluded from the analyses.

Phylogenetic trees were reconstructed using Maximum likelihood (ML) and Bayesian inference (BI) approaches. The best-fit model of amino acid substitutions was selected using ProtTest version 2.4 (Abascal et al. 2005). LG+I+G+F model was chosen as the best-fit model based on Akaike information criterion (AIC). As the LG model cannot be implemented in Bayesian analysis, the best scoring alternative, the WAG model was used. The ML analysis was conducted with PhyML 3.0 (online server: http://www.atgc-montpellier.fr/phyml/) and 1000 bootstraps were used to assess the support of nodes. BI was performed using MrBayes 3.1.2 (Ronquist and Huelsenbeck 2003). The analysis included four Markov Chain Monte Carlo chains and was run twice in parallel for 200,000 generations with trees sampled every 100 generations. Stationarity

Table 3 Organization of the mitochondrial genome of *Cucullaea labiata*

Gene	From	To	Size (bp)	Start	Stop	Intergenic nucleotide
cox1	1	1182	1182	GTG		651
cox1	1834	2247	414		TAA	−1
tRNA^{Trp}	2247	2315	69			1
tRNA^{Lys}	2317	2388	72			99
nad1	2488	3420	933	ATG	TAG	28
tRNA^{Cys}	3449	3515	67			10
tRNA$^{Leu(UUR)}$	3526	3592	67			55
nad4l	3648	3932	285	GTG	TAG	110
nad4	4043	5278	1236	ATA	TAG	36
tRNA^{Thr}	5315	5382	68			45
tRNA^{Ile}	5428	5499	72			50
tRNA$^{Ser(AGN)}$	5550	5609	60			−2
nad3	5608	6021	414	ATG	TAA	102
tRNA^{Asp}	6124	6192	69			13
tRNA^{Glu}	6206	6277	72			493
cox3	6771	7541	771	ATA	TAG	1310
nad5	8852	10,591	1740	ATA	TAA	6057
tRNA^{Met}	16,649	16,718	70			37
tRNA^{Gly}	16,756	16,827	72			20
tRNA^{Val}	16,848	16,918	71			0
rrnL	16,919	18,368	1450			0
tRNA^{Arg}	18,369	18,436	68			38
tRNA^{Tyr}	18,475	18,540	66			74
tRNA$^{Leu(CUN)}$	18,615	18,681	67			124
tRNA$^{Ser(UCN)}$	18,806	18,865	60			519
Cytb	19,385	20,608	1224	ATA	TAA	49
cox2	20,658	21,305	648	ATG	TAG	229
tRNA^{Pro}	21,535	21,600	66			83
tRNA^{Gln}	21,684	21,752	69			14
tRNA^{Asn}	21,767	21,833	67			134
nad6	21,968	22,465	498	ATG	TAA	42
nad2	22,508	23,527	1020	ATG	TAA	0
rrnS	23,528	24,808	1281			0
tRNA^{His}	24,809	24,882	74			1
tRNA^{Ala}	24,884	24,953	70			32
atp6	24,986	25,675	690	ATA	TAG	58
tRNA^{Phe}	25,734	25,800	67			45

was defined as mean standard deviation of split frequency less than 0.01. The first 25% sampled trees determined by Tracer v.1.5 (Drummond and Rambaut 2007) were discarded as burn-in. A majority-rule consensus tree was constructed using the remaining trees with support calculated by Bayesian posterior probabilities (BPP).

Results and discussion

Genome features

The complete mt genome of *C. labiata*, which contained 12 protein-coding genes, 2 rRNA and 22 tRNA genes, was

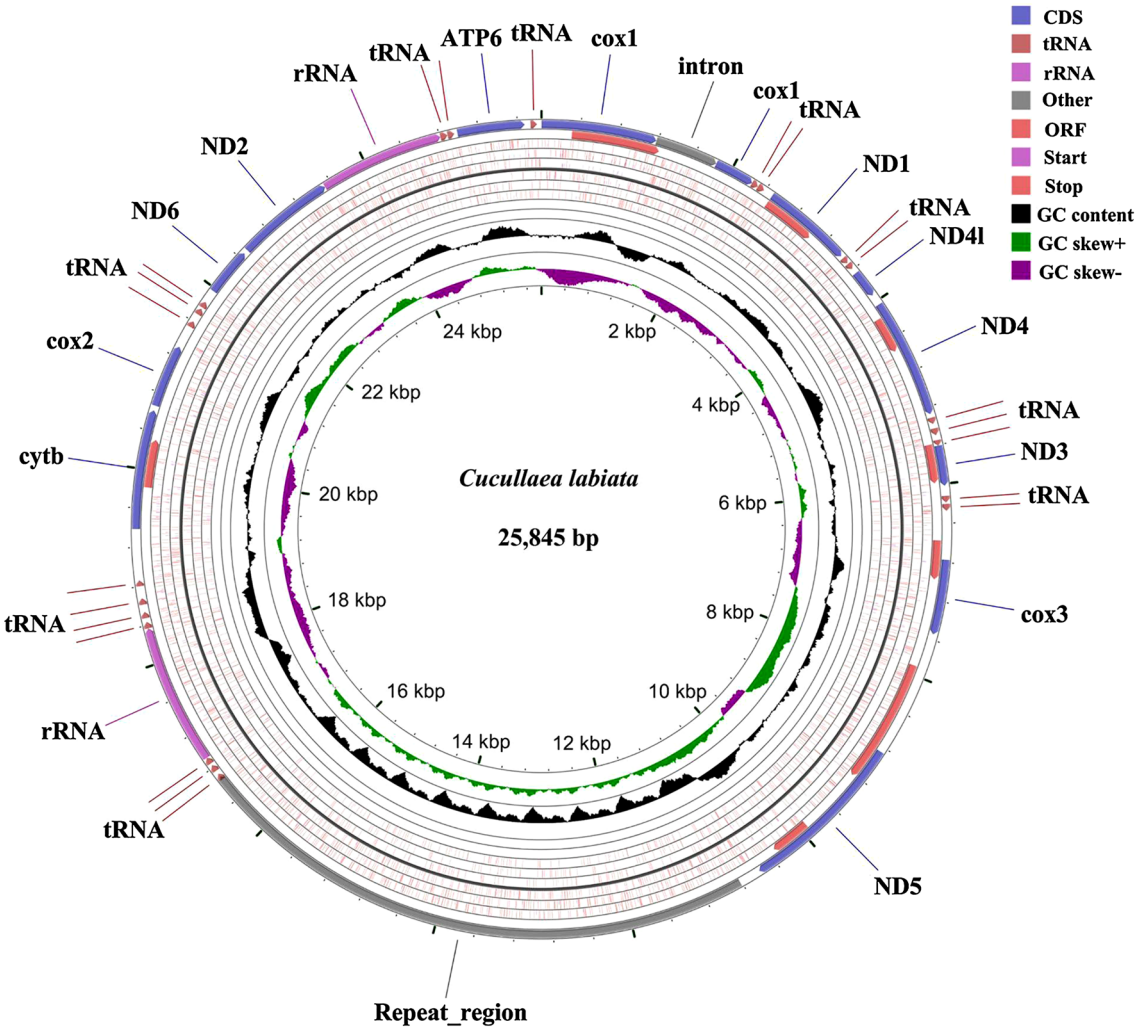

Fig. 1 Gene map of the mitochondrial genome of *Cucullaea labiata*

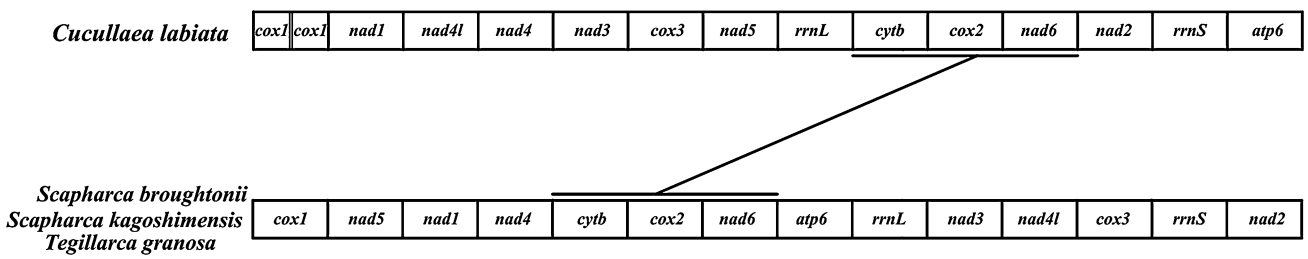

Fig. 2 Gene arrangement of *Cucullaea labiata, Scapharca broughtonii, Scapharca Kagoshimensis* and *Tegillarca granosa. tRNA* genes are excluded. The *bars* indicate identical gene block

25,845 bp long (Table 3; Fig. 1). The length is less than that of three Arcidae species, *S. broughtonii* (46,985 bp, Liu et al. 2013), *S. kagoshimensis* (46,713 bp, Sun et al. 2014) and *T. granosa* (31,589 bp, Sun et al. 2015). *Atp8* gene was not identified, as is the most distinctive features of marine bivalve mt genomes (Yu et al. 2008; Xu et al. 2011). The number and the location of the tRNA genes were different from three Arcidae species. Comparing gene arrangement of *C. labiata, S. broughtonii, S. kagoshimensis* and *T. granosa*, even excluding the tRNA genes, only one gene block (*cytb*/*cox2*/*nad6*) was shared (Fig. 2). It is common that Mt-gene order is dramatically variable in the major groups of bivalves, e.g., pectinidae species (Xu et al. 2011) and *Crassostrea* genus species (Wu et al. 2010). The length of the non-coding regions found in the mt genome of *C. labiata* was 10,559 bp. Most of these were observed within two zones. One (6057 bp) lied between *nad5* and $tRNA^{Met}$, and the other (1310 bp) lied between *cox3* and *nad5*.

The *C. labiata* genome had an overall A+T content of 63.9%, which is higher than that of *S. kagoshimensis* (62.75%) and *T. granosa* (60.18%), but lower than that of *S. broughtonii* (67.89%). The nucleotide compositions indicate the occurrence of more T and G than A and C (AT skew, −0.164; GC skew, 0.529).

Protein-coding genes and codon usage

Ten out of the 12 protein-coding genes of *C. labiata* started with the conventional invertebrate start codons, with *nad4, cox3, nad5, cytb, atp6* using ATA and *nad1, nad3, cox2, nad6, nad2* using ATG. The *cox1* and *nad4l* gene employed GTG as start codon. This alternate start codon GTG was also discovered in one Arcidae species *S. broughtonii* (Liu et al. 2013). All the genes had complete termination codons. *cox1, nad3, nad5, cytb, nad6*, and *nad2* were terminated by a TAA codon, while the remaining six genes were terminated by a TAG codon. Two *cox1* sequences did not have separate start and termination codons.

Excluding stop codons, the mt genome of *C. labiata* encoded 3673 amino acids, in which the most and the least frequent amino acids are leucine (Leu) and aspartic acid (Asp), respectively. UUU (Phenylalanine) is the most frequently used codon, similar to three Arcidae species. Codon usages of the protein-coding genes in *C. labiata* mt genome were detailed in Table 4.

Transfer RNAs and ribosomal RNAs

There were 22 tRNA genes in the mt genome of *C. labiata*, which ranged from 60 to 74 bp and could be folded into typical cloverleaf secondary structures. It is well known that the mt genomes of most bivalves contain more than one $tRNA^{Met}$ (Xu et al. 2011), but *C. labiata* had only one $tRNA^{Met}$, different from the three Arcidae species. Two copies of $tRNA^{Leu}$ and $tRNA^{Ser}$ existed in *C. labiata*. In the three Arcidae species, $tRNA^{Leu}$ and $tRNA^{Ser}$ also had more than one copy, except for *T. granosa*.

The locations of the *rrnL* and *rrnS* genes were identified by BLAST searches in this study. The *rrnL* was located between $tRNA^{Val}$ and $tRNA^{Arg}$, and the length was 1450 bp. The *rrnS* consisted of 1281 nucleotides and located

Table 4 Codon usages of the protein-coding genes in *Cucullaea labiata* mitochondrial genome

Codon	*N*	%	Codon	*N*	%	Codon	*N*	%	Codon	*N*	%
UUU(F)	327	8.903	UCU(S)	97	2.641	UAU(Y)	112	3.049	UGU(C)	99	2.695
UUC(F)	18	0.490	UCC(S)	6	0.163	UAC(Y)	4	0.109	UGC(C)	12	0.327
UUA(L)	204	5.554	UCA(S)	30	0.817	UAA([a])	6	0.163	UGA(W)	42	1.143
UUG(L)	177	4.819	UCG(S)	18	0.490	UAG([a])	6	0.163	UGG(W)	91	2.478
CUU(L)	69	1.879	CCU(P)	78	2.124	CAU(H)	73	1.987	CGU(R)	35	0.953
CUC(L)	6	0.163	CCC(P)	7	0.191	CAC(H)	6	0.163	CGC(R)	3	0.082
CUA(L)	13	0.354	CCA(P)	23	0.626	CAA(Q)	26	0.708	CGA(R)	12	0.327
CUG(L)	19	0.517	CCG(P)	9	0.245	CAG(Q)	37	1.007	CGG(R)	23	0.626
AUU(I)	174	4.737	ACU(T)	47	1.280	AAU(N)	65	1.770	AGU(S)	58	1.579
AUC(I)	6	0.163	ACC(T)	2	0.054	AAC(N)	8	0.218	AGC(S)	5	0.136
AUA(M)	98	2.668	ACA(T)	21	0.572	AAA(K)	28	0.762	AGA(S)	44	1.198
AUG(M)	88	2.396	ACG(T)	10	0.272	AAG(K)	45	1.225	AGG(S)	119	3.240
GUU(V)	254	6.915	GCU(A)	121	3.294	GAU(D)	61	1.661	GGU(G)	141	3.839
GUC(V)	6	0.163	GCC(A)	8	0.218	GAC(D)	4	0.109	GGC(G)	11	0.299
GUA(V)	76	2.069	GCA(A)	22	0.599	GAA(E)	29	0.790	GGA(G)	78	2.124
GUG(V)	119	3.240	GCG(A)	26	0.708	GAG(E)	59	1.606	GGG(G)	164	4.465

N is representative for the total number of particular codon in all protein-coding genes

[a]A total of 3673 codons were analyzed excluding the termination codons

between *nad2* and $tRNA^{His}$. The A+T contents of the *rrnL* and *rrnS* were 65 and 63.6%, respectively.

Non-coding regions

The mt genome of *C. labiata* contained 31 non-coding regions ranging in size from 1 to 6057 bp. The total length of all non-coding regions was 10,559 bp, accounting for 40.86% of the entire mtDNA. This percentage is much lower than that of *S. broughtonii* (67.4%), *S. kagoshimensis* (70.6%) and *T. granosa* (51.9%). Within these non-coding regions, the largest one (6057 bp) included a 659 bp repeat motif. Eight point nine copies were found and the copies were almost identical (10,817–11,475, 11,476–12,134, 12,135/;–12,793, 12,794–13,452, 13,453–14,111, 14,112–14,769, 14,770–15,428, 15,429–16,087, and 16,088–16,685). There are numerous reports describing tandem repeats in the non-coding regions of mt genomes of bivalves (Smith and Snyder 2007; Yuan et al. 2012; Meng et al. 2012). In mtDNA of the three Arcidae species, repeated sequences have also been described. Smith and Snyder (2007) and Endo et al. (2005) have reported repeated sequences in the mt genome of *Placopecten magellanicus* and *Lingula anatina* are associated with

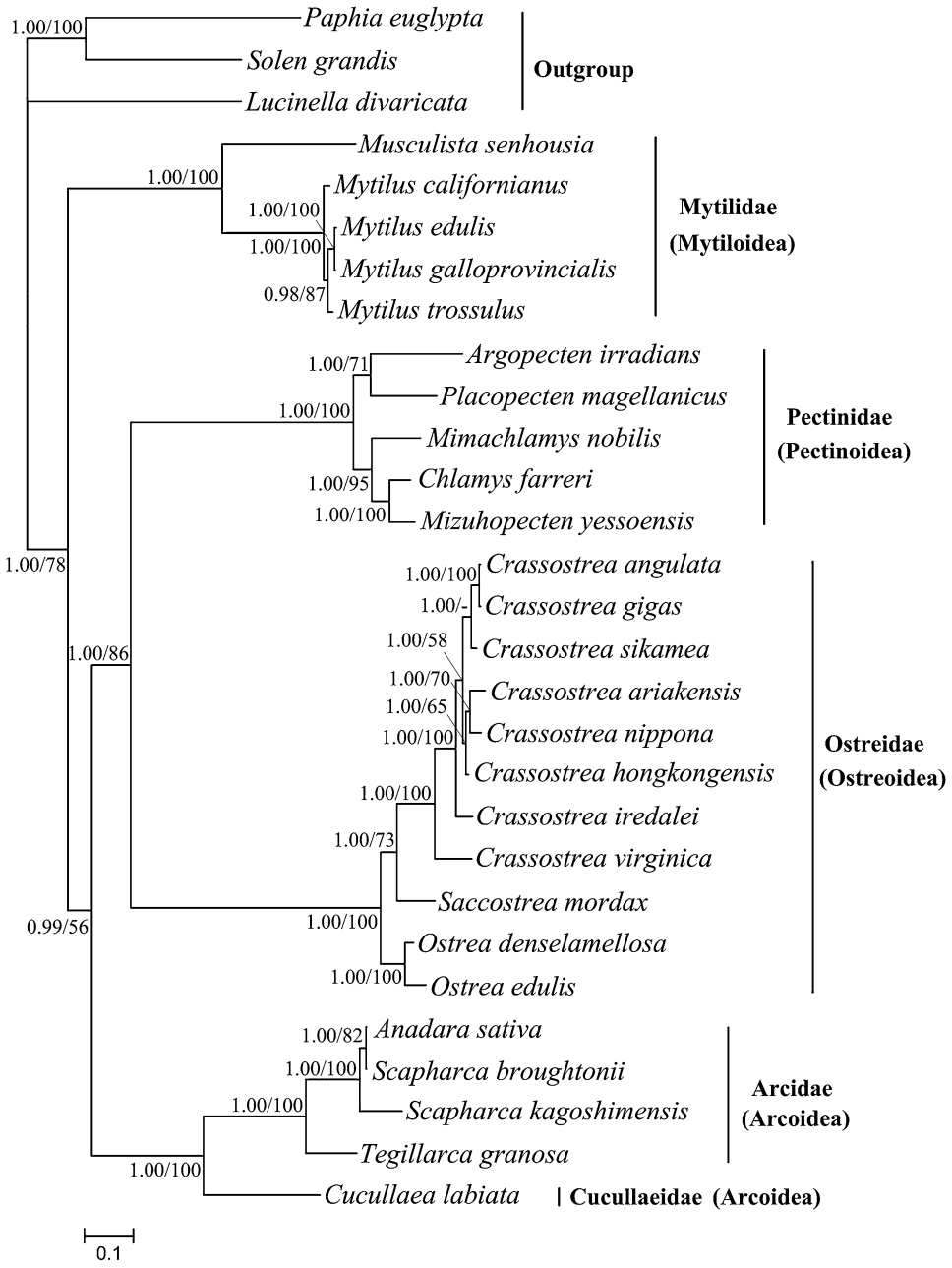

Fig. 3 Phylogenetic trees based on the concatenated amimo acid sequences of 12 protein-coding genes. The first number at each node is Bayesian posterior probability and the second number is ML bootstrap values. *Dashes* indicate support values below 50%

tRNA or pseudo-tRNA structures. This feature was also found in *C. labiata* ($tRNA^{Met}$). Furthermore, the copy numbers of repeated sequences are different in different individuals, e.g., *P. magellanicus* and *S. broughtonii*. In *C. labiata*, the individual CL_98 had eight point nine copies of repeated sequences, and CL_99 only contained one copy.

One intron in *cox1*

Although introns have been found in *cox1* gene of two demosponge species (*Tetilla sp*., Rot et al. 2006; *Plakortis angulospiculatus*, Wang and Lavrov 2008) and scleractinian corals (Fukami et al. 2007), no intron are reported in mtDNA of bivalves. In this study, we found one intron in *cox1* gene both in CL_98 and in CL_99 individual of *C. labiata*. The intron was 651 bp long, divided *cox1* gene into two sections (1182 and 414 bp). Among the intron sequences, a total of 8 open reading frames were found, ranging in size from 11 to 47 amino acids. No sequence similarity was detected between the ORFs and known proteins. In the 8 ORFs, only one had both signal peptide and transmembrane helices, 4 had transmembrane helices, and the remaining three had neither signal peptide nor transmembrane helices.

Introns are common in mtDNA of cnidarians and placozoans, but in demosponges, only two species were detected (Wang and Lavrov 2008). Why the introns are retained in demosponges so scarcely is still unknown. In the same way, why the mt genome of *C. labiata* had one intron, whereas other bivalves especially the Arcidae species did not have, still need further research.

Phylogenetic analyses

To further study the relationship between Cucullaeidae and Arcidae, and the phylogenetic positions within Bivalvia, ML and BI trees were reconstructed using amino acid sequence of 12 concatenated protein-coding genes (Fig. 3). The topological structure inferred by these two different methods were in almost complete agreement. The results showed that *C. labiata* formed a single group, being the sister group of Arcidae clade, which indicates Cucullaeidae and Arcidae have a common ancestor and supports the viewpoint of Oliver and Holmes (2006) that the Cucullaeidae appear to be contemporary with the Arcidae. At present, the number of mitochondrial genomes of Arcidae is still too limited and the relationship between Cucullaeidae and Arcidae needs to be further resolved. Within Arcidae, *Anadara sativa*, the synonym of *S. kagoshimensis*, attained from GenBank (KF667521) clustered together with *S. broughtonii*, suggests it may be contaminated or misidentified.

The families Cucullaeidae and Arcidae both belong to the superfamily Arcoidea. There are considerable discrepancies in regard to the phylogenetic positions of Arcoidea among bivalve taxa (Waller 1998; Steiner and Hammer 2000; Plazzi et al. 2011; Sharma et al. 2012). In our study, the phylogenetic tree showed that Pectinoidea and Ostreoidea had the closest relationship, as the sister group of Arcoidea, then Mytiloidea clustered with (Pectinoidea+Os treoidea)+Arcoidea, which is compatible with the opinion of Waller (1998).

Acknowledgements This study was supported by research grants from the National Natural Science Foundation of China (41276138), and Fundamental Research Funds for the Central Universities.

Compliance with ethical standards

Conflict of interest Yanwei Feng declares that he/she has no conflict of interest. Qi Li declares that he/she has no conflict of interest. Hong Yu declares that he/she has no conflict of interest. Lingfeng Kong declares that he/she has no conflict of interest.

Ethical approval The research was conducted in the absence of any ethical issue on aquatic animal research.

References

Abascal F, Zardoya R, Posada D (2005) ProtTest: selection of bestfit models of protein evolution. Bioinformatics 21:2104–2105

Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res 27:573–580

Boore JL (1999) Animal mitochondrial genomes. Nucleic Acids Res 27:1767–1780

Boore JL, Brown WM (2000) Mitochondrial genomes of *Galathealinum, Helobdella*, and *Platynereis*: sequence and gene arrangement comparisons indicate that Pogonophora is not a phylum and Annelida and Arthropoda are not sister taxa. Mol Biol Evol 17:87–106

Boss KJ (1982) Mollusca. In: Parker SP (ed) Synopsis and classification of living organisms, vol 1. McGraw-Hill, New York, p 1111

Burland TG (2000) DNASTAR's Lasergene sequence analysis software. Methods Mol Biol 132:71–91

Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17:540–552

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7:214

Endo K, Noguchi Y, Ueshima R, Jacobs HT (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 YW, Li Q, Kong LF (2015) Molecular phylogeny of Arcoidea with emphasis on Arcidae species (Bivalvia: Pteriomorphia) along the coast of China: challenges to current classification of arcoids. Mol Phylogenet Evol 85:189–196

Fukami H, Chen CA, Chiou CY, Knowlton N (2007) Novel group I introns encoding a putative homing endonuclease in the mitochondrial cox1 gene of scleractinian corals. J Mol Evol 64:591–600

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 95–98

Hofmann K, Stoffel W (1993) TMbase—a database of membrane spanning proteins segments. Biol Chem Hoppe-Seyler 374:166

Käll L, Krogh A, Sonnhammer ELL (2004) A combined transmembrane topology and signal peptide prediction method. J Mol Biol 338:1027–1036

Lang BF, Laforest MJ, Burger G (2007) Mitochondrial introns: a critical view. Trends Genet 23:119–125

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 YG, 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 8:72–81

Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25:955–964

Lunt DH, Whipple LE, Hyman BC (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 HL, Yan BL, 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 7:175–179

Oliver PG, Holmes AM (2006) The Arcoidea (Mollusca: Bivalvia): a review of the current phenetic-based systematics. Zool J Linn Soc 148:237–251

Palumbi SR (1996) Nucleic acids II: the polymerase chain reaction. In: Hillis D, Moritz C (eds) Molecular systematics. Sinauer, Sunderland, pp 205–247

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 JL, Labesse G (2009) @TOME-2: a new pipeline for comparative modeling of protein-ligand complexes. Nucleic Acids Res 37:W485–W491

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574

Rot C, Goldfarb I, Ilan M, Huchon D (2006) Putative cross-kingdom horizontal gene transfer in sponge (Porifera) mitochondria. BMC Evol Biol 6:71

Shadel GS, Clayton DA (1997) Mitochondrial DNA maintenance in vertebrates. Annu Rev Biochem 66:409–435

Sharma PP, González VL, Kawauchi GY, Andrade SCS, Guzmán A, Collins TM, Glover EA, Harper EM, Healy JM, Mikkelsen PM et al (2012) Phylogenetic analysis of four nuclear protein-encoding genes largely corroborates the traditional classification of Bivalvia (Mollusca). Mol Phylogenet Evol 65:64–74

Smith DR, 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

Snyder M, Fraser AR, LaRoche J, Gartner-Kepkay KE, Zouros E (1987) A typical mitochondrial DNA from the deep-sea scallop *Placopecten magellanicus*. Proc Natl Acad Sci USA 84:7595–7599

Steiner G, Hammer S (2000) Molecular phylogeny of the Bivalvia inferred from 18 S rDNA sequences with particular reference to the Pteriomorphia. Geol Soc Lond Spec Publ 177:11–29

Sun SE, Kong LF, Yu H, Li Q (2014) The complete mitochondrial genome of *Scapharca kagoshimensis* (Bivalvia: Arcidae). Mitochondrial DNA 1–2

Sun SE, Kong LF, Yu H, Li Q (2015) The complete mitochondrial DNA of *Tegillarca granosa* and comparative mitogenomic analyses of three Arcidae species. Gene 557:61–70

Waller TR (1998) Origin of the molluscan class Bivalvia and a phylogeny of major groups. Bivalves 1(4):5

Wang X, Lavrov DV (2008) Seventeen new complete mtDNA sequences reveal extensive mitochondrial genome evolution within the Demospongiae. PLoS ONE 3:e2723

Wolstenholme DR (1992) Animal mitochondrial DNA: structure and evolution. Int Rev Cytol 141:173–216

Wu XY, Xu XD, Yu ZN, Wei ZP, Xia JJ (2010) Comparison of seven *Crassostrea* mitogenomes and phylogenetic analyses. Mol Phylogenet Evol 57:448–454

Wyman SK, Jansen RK, Boore JL (2004) Automatic annotation of organellar genomes with DOGMA. Bioinformatics 20:3252–3255

Xu KF, 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

Yamanoue Y, Miya M, Matsuura K, Yagishita N, Mabuchi K, Sakai H (2007) Phylogenetic position of tetraodontiform fishes within the higher teleosts: Bayesian inferences based on 44 whole mitochondrial genome sequences. Mol Phylogenet Evol 45:89–101

Yu ZN, Wei ZP, Kong XY, 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 LF (2012) The complete mitochondrial genomes of six heterodont bivalves (Tellinoidea and Solenoidea): variable gene arrangements and phylogenetic implications. PLoS ONE 7:e32353

Zdobnov EM, Apweiler R (2001) InterProScan—an integration platform for the signature—recognition methods in InterPro. Bioinformatics 17:847–848