Gene 442 (2009) 8–17

Contents lists available at ScienceDirect

Gene

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

The complete mitochondrial genome of *Cephalothrix simula* (Iwata) (Nemertea: Palaeonemertea)

Hai-Xia Chen [a], Per Sundberg [b], Jon L. Norenburg [c], Shi-Chun Sun [a,*]

[a] Mariculture Research Laboratory, Ocean University of China, 5 Yushan Road, Qingdao 266003, China
[b] Department of Zoology, Göteborg University, PO Box 463, SE-405 30 Göteborg, Sweden
[c] Department of Invertebrate Zoology, National Museum of Natural History, Smithsonian Institution, Washington, DC 20560-0163, USA

ARTICLE INFO

Article history:
Received 3 January 2009
Received in revised form 10 April 2009
Accepted 13 April 2009
Available online 3 May 2009

Received by P. Deininger

Keywords:
mtDNA
Nemertean
Gene order
Phylogeny

ABSTRACT

The first complete mitochondrial genome sequence for a nemertean, *Cephalothrix simula*, was determined by conventional and long PCR and sequencing with primer walking methods. This circular genome is 16,296 bp in size and encodes 37 genes (13 protein-coding genes, 2 ribosomal RNAs, and 22 transfer RNAs) typically found in metazoans. All genes are encoded on H-strand except two tRNAs (*trnT* and *trnP*). It differs from those reported for other metazoans, but some gene junctions are shared with those of other protostomes. Structure of the mitochondrial genome of *C. simula* is mostly concordant with the partial mitochondrial genome known for *Cephalothrix rufifrons*, but notable differences include three large indel events and transposition of 2 tRNAs. Nucleotide composition of the mitochondrial genome of *C. simula* is highly A+T biased. The compositional skew is strongly reflected in the codon-usage patterns and the amino acid compositions of the mitochondrial proteins. An AT-rich noncoding region with potential to form stem–loop structures may be involved in the initiation of replication or transcription. Gene adjacencies and phylogenetic analysis based on the 12 concatenated amino acid sequences (except *atp8*) of mitochondrial protein-coding genes show that the nemertean is close to the coelomate lophotrochozoans, rather than the acoelomate platyhelminths.

1. Introduction

Metazoan mitochondrial genomes, ranging in size from 14 to 18 kb, are typically circular and usually encode 37 genes including 13 protein-coding genes (*cox1–3*, *nad1–6*, *nad4L*, *atp6*, *atp8* and *cob*), two ribosomal RNA genes (*rrnL* and *rrnS*) and 22 tRNA genes (Wolstenholme, 1992; Boore, 1999). In addition, there usually is a noncoding (AT-rich, control or D-loop) region, which may contain elements that control the replication and transcription of the genome (Wolstenholme, 1992; Shadel and Clayton, 1997). With a few exceptions, the gene content of animal mitochondrial genomes is generally conserved (Boore, 1999), but gene order is more variable (Boore and Brown, 1998; Moret et al., 2001).

Complete mitochondrial genomes have been characterized from a variety of metazoan phyla and used for analyzing phylogenetic relationships. To date, there are over 1240 complete mitochondrial genomes of metazoans in GenBank, mainly for Arthropoda, Mollusca, Platyhelminthes, Nematoda and Chordata. Nemerteans (phylum Nemertea) are unsegmented worms possessing an eversible proboscis that is the synapomorphy for the taxon (Turbeville, 2002). The phylum, currently contains 1275 described species (Kajihara et al., 2008), and for a long time was considered acoelomate and closely related to platyhelminths on the basis of mainly morphological and developmental similarities (e.g., Bürger, 1897–1907; Hyman, 1951; Nielsen, 1995). The argument of a coelomate position for nemerteans was supported by ultrastructural similarity between the blood vessels of nemerteans and the coelom of spiralian coelomates (Turbeville and Ruppert, 1985; Turbeville, 1986; Jespersen and Lützen, 1988). There is an increasing number of studies using molecular information to provide increasingly refined estimates of within-phylum phylogenetic relationships of nemerteans (e.g., Envall and Sundberg, 1998; Sundberg and Saur, 1998; Thollesson and Norenburg, 2003; Sundberg et al., 2003; Strand and Sundberg, 2005; Sundberg and Strand, 2007). Few studies have addressed the position of phylum Nemertea among metazoans. Whereas none have supported a platyhelminth + nemertea clade; recent molecular analyses consistently place Nemertea within the coelomate Lophotro-

Abbreviations: *atp6* and *atp8*, genes for ATP synthase subunits 6 and 8; *cob*, gene for cytochrome *b*; *cox1–3*, genes for subunits I–III of cytochrome *c* oxidase; *nad1–6* and *nad4L*, NADH dehydrogenase subunits 1–6 and 4L; *rrnL* and *rrnS*, genes for the large and small subunits of ribosomal RNA; *trnX*, genes encoding for transfer RNA molecules with corresponding amino acids denoted by the one-letter code and codon indicated in parentheses (xxx) when necessary; DHU, dihydrouridine loop; mtDNA, mitochondrial DNA; NC, noncoding region; PCR, polymerase chain reaction; TΨC, pseudouridine loop; kb, kilobase; bp, base pair; BI, Bayesian inference; ML, Maximum likelihood. Nucleotide symbol combination R = A/G, Y = C/T, W = A/T, B = G/C/T.

* Corresponding author. Tel.: +86 532 82032216.
E-mail address: sunsc@ouc.edu.cn (S.-C. Sun).

0378-1119/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2009.04.015

chozoa, but at various positions—as sister to mollusks, brachiopods, entoprocts, etc. (e.g., Giribet et al., 2004; Turbeville and Smith, 2007; Struck et al., 2007; Bourlat et al., 2008; Dunn et al., 2008; Struck and Fisse, 2008). These findings have not been tested with nucleotide data and gene-order information from whole mitochondrial genome sequences. Turbeville and Smith (2007) published a continuous 10.1 kb fragment sequence for the palaeonemertean *Cephalothrix rufifrons*, and here we report for the first time a complete mitochondrial genome for a nemertean, *Cephalothrix simula* (Iwata, 1952). We present the genome structure, gene arrangement, nucleotide composition, and codon usage and use this information to assess the phylogenetic position of Nemertea within the lophotrochozoans.

2. Materials and methods

2.1. Samples, mtDNA extraction, PCR amplification, cloning and sequencing

C. simula was collected from the littoral zone in Qingdao, China, in December 2006. Genomic DNA was extracted from a single individual by a standard phenol-chloroform protocol (Sambrook and Russell, 2001). The complete mitochondrial genome sequence was obtained using a PCR-based strategy involving a combination of conventional PCR and long PCR to amplify overlapping mt genome fragments.

Initially, the small fraction *rrnL–nad1*, and the partial *cox1* and *cox3* genes were amplified using universal primers (see Table 1). The PCR products were cloned and sequenced. The sequence data in conjunction with the genomes published for some metazoans were used to design taxon-specific PCR primers (*nad1–nad4L*, *cox3–nad3*, *cox1–cox2*, Table 1). Then the determined sequences were used to design three additional PCR primer pairs (Table 1) bridging the gaps between *nad4L–cox3*, *nad3–cox1* and *cox2–rrnL*. Fragments of 3 kb or less were amplified by conventional PCR and fragments greater than 3 kb were amplified by long PCR. The complete genome was amplified in nine overlapping fragments that were pieced together, annotated and analyzed.

Table 1
PCR primers used to amplify the complete mitochondrial genome of *Cephalothrix simula*.

Primer name	Sequence (5′→3′)	References
Universal PCR primers		
rrnL–nad1		
16SarL	CGCCTGTTTATCAAAAACAT	Palumbi, 1996
nad1R	CCTGATACTAATTCAGATTCTCCTTC	Boore, 2006
cox1		
LCO-1490	GGTCAACAAATCATAAAGATATTGG	Folmer et al., 1994
HCO-2198	TAAACTTCAGGGTGACCAAAAAATCA	Folmer et al., 1994
cox3		
cox3F	TGCGWTGAGGWATAATTTTATTTATT	Turbeville and Smith, 2007
cox3R	ACCAAGCAGCTGCTTCAAAACCAAA	Turbeville and Smith, 2007
Specific PCR primers		
nad1–nad4L		
Psnad1_F	CACTACGCTCTGTTGGGTTCA	Present study
Psnad4L_R	GCCGTCCTACTCTTGCTTCA	Present study
nad4L–cox3		
Psnad4L_Flong	TCGGTTATTGCGGTGGTTCT	Present study
Pscox3_Rlong	TGGAGGAGGTCAACTACAAC	Present study
cox3–nad3		
Pscox3_F	GTAGACGGAGGTCGTTTAGGAGGAA	Present study
Psnad3_R	GRCCAAAGCCACAYTCAAAAGGAGA	Present study
nad3–cox1		
Psnad3_F	CAGGACCACCTCGTTTACCBTTTTC	Present study
Pscox1_R	TAATACCGCACCAGGTTGTC	Present study
cox1–cox2		
Pscox1_F	GGGGTGTAGGAACTGGATGAA	Present study
Pscox2_R	AGTTCGGARCAYTGHCCRTA	Present study
cox2–rrnL		
Pscox2_Flong	CATCGCACTGTAGTACCTTA	Present study
psrrnL_Rlong	ACGCTGTTATCCCTATGGTA	Present study

Conventional PCR reactions were carried out with 16.3 μl distilled H_2O, 2.5 μl 10× reaction buffer, 2 μl $MgCl_2$ (25 mM), 2 μl dNTPs (2.5 mM), 0.5 μl each primer (10 μM), 0.20 μl rTaq DNA polymerase (Takara) and 1 μl DNA template. Thermal cycling was performed in a thermal cycler (Bio-Rad MyCycler™), programmed for an initial denaturing step of 94 °C for 5 min, followed by 35 cycles at 94 °C for 40 s, 48–55 °C (depending on the targets) for 50 s, and 72 °C for 60 s, and final extension at 72 °C for 7–10 min.

For amplifications of two large fragments (*nad4L–cox3*, *cox2–rrnL*, approximately 4 kb), a long PCR was performed. The reactions were set up containing 18.3 μl distilled H_2O, 2.5 μl 10× LA PCR buffer II (Mg^{2+} plus, Takara), 2 μl dNTPs (2.5 mM), 0.5 μl of each primer (10 μM), 0.2 μl of LA-*Taq* polymerase (Takara), and 1 μl DNA template. Thermal cycling was programmed for an initial denaturing step of 95 °C for 3 min, and then 35 cycles with 98 °C for 10 s, 68 °C for 6 min, and final extension at 68 °C for 10 min.

All PCR products were inspected under UV transillumination and purified with the PCR Gel extraction kit (Takara). The fragments were cloned into PMD18-T vector (Takara) and sequenced in both directions by primer walking on an ABI 3730 Sequencer (Applied Biosystems, Foster, CA, USA). Two or more single clones were sequenced for each fragment.

2.2. Sequence alignments

All sequences were checked and aligned by visual inspection using the program Bioedit v. 7.0.1 (Hall, 1999). Analysis of transfer RNAs was conducted with tRNAscan-SE (Lowe and Eddy, 1997) using invertebrate mitochondrial genetic code. The complete rRNA genes were identified by comparison with *rrnL* and *rrnS* genes from a range of metazoans through BLAST (http://www.ncbi.nlm.nih.gov/BLAST) searches and boundaries were determined by the terminal ends of adjacent genes. Protein-coding genes were identified through BLAST and then aligned with sequences of *C. rufifrons* and genes available in GenBank. Positions of start and stop codons of the coding genes were identified using the invertebrate mitochondrial genetic code and amino-acid reading frames were checked. Boundaries of noncoding regions were recognized by the range of the coding genes. Final nucleotide composition and codon usage were analyzed with the program MEGA ver.3 (Kumar et al., 2004), and the mitochondrial genome was visualized using the program CG View (Stothard and Wishart, 2005).

The complete sequence of *C. simula* mtDNA is deposited in GenBank (FJ594739).

2.3. Phylogenetic analysis

The phylogenetic analyses are based on the complete mt sequence of *C. simula* (this study) and complete or nearly complete mitochondrial genomes of 32 species from GenBank (Table 2). Two arthropods, *Limulus polyphemus* and *Drosophila yakuba*, were used as outgroup. Gene *atp8* was excluded from the analysis as most species of Platyhelminthes, Acanthocephala, Rotifera and some species from the phylum Mollusca lack the gene. The nucleotides of 12 protein-coding genes were translated into amino acids using the invertebrate mitochondrial genetic code, and the translated amino acid sequences were aligned using the default parameters of CLUSTAL X version 1.83 (Thompson et al., 1997). Alignment ambiguities and gaps were excluded from phylogenetic analysis using GBLOCKS version 0.91b (Castresana, 2000) with default parameters. Finally, subsets of different genes for the same set of species were concatenated.

Phylogeny was estimated using Bayesian inference (BI; Huelsenbeck and Ronquist, 2001) and maximum likelihood (ML; Felsenstein, 1981). The best-fit models of amino acid substitutions were selected using ProtTest version 1.2.7 (Abascal et al., 2005). The Akaike Information Criterion was used to determine mtREV+G+I as the