Molecular Phylogenetics and Evolution 120 (2018) 233–239

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Short Communication

Mitochondrial gene order evolution in Mollusca: Inference of the ancestral state from the mtDNA of *Chaetopleura apiculata* (Polyplacophora, Chaetopleuridae)

Davide Guerra*, Karim Bouvet, Sophie Breton

Université de Montréal, Département de Sciences Biologiques, 90 avenue Vincent-D'Indy, H2V 2S9 Montréal, Québec, Canada

ARTICLE INFO

Keywords:
Chaetopleura apiculata
Polyplacophora
Mollusca
Mitochondrial genome
Gene order evolution
Phylogeny

ABSTRACT

The mitochondrial genome architecture of polyplacophorans has been usually regarded as being very ancient in comparison to all mollusks. However, even if some complete chiton mtDNAs have been recently sequenced, thorough studies of their evolution are lacking. To further expand the set of complete chiton mtDNAs and perform such analysis, we sequenced the mitochondrial genome of the Eastern beaded chiton *Chaetopleura apiculata* (Chaetopleuridae) using next-generation sequencing. With mitochondrial sequences from all available chiton mtDNAs, we also built a phylogeny on which we reconstructed the evolution of gene arrangement in this class. The arrangement of *C. apiculata* proved to be the most primitive known so far for polyplacophorans. Comparing this gene order to those of other molluscan classes, we found that it most probably is the original gene order of the last common ancestor to all extant Mollusca. The ancient mitochondrial genome organization of *C. apiculata* is an important information that may help reconstructing the phylogeny of Mollusca and their relationship with other lophotrochozoans.

1. Introduction

Mollusca is the second largest animal phylum and comprises the eight classes Bivalvia, Caudofoveata (= Chaetodermomorpha), Cephalopoda, Gastropoda, Monoplacophora, Polyplacophora, Scaphopoda, and Solenogastres (= Neomeniomorpha) (Ponder and Lindberg, 2008). The phylogenetic relationships among these classes have been thoroughly investigated with various methods, resulting in different reconstructions (Kocot, 2013). The most supported model of molluscan evolution (see for example the recent studies of Kocot et al., 2011; Smith et al., 2011; Vinther et al., 2012; Osca et al., 2014 and references therein) is the so-called "Aculifera hypothesis", where two major sister lineages of mollusks are recognized: Aculifera, comprising Polyplacophora as sister to the aplacophorans Caudofoveata and Solenogastres; and Conchifera, comprising the remaining five classes of shell-bearing mollusks.

At the time of writing this article (February 2017), 804 complete mt genome sequences are available on GenBank for 336 molluscan species: 161 gastropods, 122 bivalves, 41 cephalopods, and 12 for four of the remaining classes (no mtDNA sequence is available yet for Solenogastres). A mt genome carries phylogenetic signals on many levels, the most widely used being the nucleotide sequences of its 37 genes, i.e. the 13 protein-coding genes and the 24 RNA genes required for their translation, the amino acid sequences of the proteins encoded by the protein-coding genes, and the arrangement of all these genes on the two strands of the mtDNA molecule. Attempts were recently made (Stöger and Schrödl, 2013; Osca et al., 2014; Stöger et al., 2016) to assess the potential of mitogenomics in reconstructing the evolution of Mollusca and test various phylogenetic hypotheses, but the results were sometimes indecisive or affected by issues such as long-branch attraction. Given the extremely ancient age of the phylum, radiated during the early Cambrian or maybe even before (Ponder and Lindberg, 2008), most of the phylogenetic signal contained in the mtDNAs might have been lost during evolution, leading to artefactual phylogenies (Bernt et al., 2013a; Stöger and Schrödl, 2013; Stöger et al., 2016).

Complete mt genomes are extremely underrepresented for Polyplacophora, with only six available mtDNAs up to date (Boore and Brown, 1994; Irisarri et al., 2014; Veale et al., 2016) all coming from

Abbreviations: 12S and *16S*, genes encoding respectively the small and large mitochondrial ribosomal RNA subunits; A, adenine; *A, C, D, E, F, G, H, I, K, L1, L2, M, N, P, Q, R, S1, S2, T, V, W, Y*, genes encoding the 22 mitochondrial tRNAs; *atp6-8*, genes for ATP-synthase F_0 subunits 6 and 8; C, cytosine; *cob*, gene for cytochrome *b*; *cox1-3*, genes for cytochrome *c* oxidase subunits 1-3; G, guanine; mt, mitochondrial; *nd1-6* and *4L*, genes for NADH dehydrogenase subunits 1-6 and 4L; rRNA, ribosomal RNA; T, thymine; tRNA, transfer RNA; Y, cytosine or thymine

* Corresponding author.
E-mail addresses: davide.guerra@umontreal.ca (D. Guerra), karim.bouvet-hasab.alla@umontreal.ca (K. Bouvet), s.breton@umontreal.ca (S. Breton).

https://doi.org/10.1016/j.ympev.2017.12.013
Received 10 May 2017; Received in revised form 18 November 2017; Accepted 11 December 2017
Available online 16 December 2017
1055-7903/

species of the order Chitonida (see Irisarri et al., 2014). The organization of protein-coding genes and rRNA genes in the black chiton *Katharina tunicata* has been traditionally considered very ancestral for mollusks, since it is found also in Cephalopoda and Gastropoda (Stöger and Schrödl, 2013). The gene arrangement descriptions of the other five polyplacophoran genomes available, especially for the two belonging to the genus *Sypharochiton*, have been rather discordant, creating some confusion about the gene order variability in this group (Veale et al., 2016; Stöger et al., 2016). The need for more complete mtDNA sequences is thus of pivotal importance to expand our knowledge on polyplacophoran mitogenomics. We obtained the complete mt genome of the Eastern beaded chiton *Chaetopleura apiculata* (Say in Conrad, 1834) (1) as part of a broader project on molluscan mitogenomics ongoing in our lab, and (2) to widen the coverage of the order Chitonida to solve the ambiguities revolving the descriptions of polyplacophoran mt gene order variability. The gene order of *Chaetopleura* proved to be unexpectedly primitive among Polyplacophora, and this prompted us to compare it with those of other mollusks. We inferred that *C. apiculata* gene arrangement may be the plesiomorphic condition for all Mollusca.

2. Materials and methods

2.1. mtDNA sequencing and annotation

A specimen of *C. apiculata* from Gulf of Mexico was provided by Gulf Specimen Marine Laboratories, Inc. (Panacea, FL, USA) in June 2016. Total DNA was extracted from the foot using the MasterPure Complete DNA and RNA purification kit (Epicentre, Madison, WI, USA). A DNA library was constructed using a TruSeq DNA PCR-free kit (Illumina) and paired-end sequenced (2 × 100 bp) using a HiSeq platform (Illumina) by Génome Québec Innovation Centre (Montréal, QC, Canada). Reads were trimmed with Trimmomatic 0.32 (Bolger et al., 2014) and assembled with Pear 0.9.6 (Zhang et al., 2013) on the Galaxy online platform (Afgan et al., 2016), checking the quality of reads at every step with FastQC 0.11.5 (Andrews, 2016). MITObim 1.8 (Hahn et al., 2013) was used to assemble the complete genome sequence, starting from a *C. apiculata* partial *cox1* sequence (GenBank accession number KP254304) as an initial seed. The sequence obtained was annotated with MITOS (Bernt et al., 2013b). The predicted gene boundaries were manually adjusted by comparing them with those of the other six polyplacophoran mtDNAs (*Cryptochiton stelleri*, KJ569363; *Cyanoplax caverna*, KJ569361; *Katharina tunicata*, U09810; *Nuttallina californica*, KJ569362; *Sypharochiton pelliserpentis*, KJ534307; *Sypharochiton sinclairi*, KJ534306) (Boore and Brown, 1994; Irisarri et al., 2014; Veale et al., 2016) using Geneious 4.8.5 (Kearse et al., 2012). By doing this, we also assessed the overall good quality of the MITObim assembly by checking the integrity of the coding sequences. Coverage and variability data for the complete mt genome were obtained with Bowtie (Langmead et al., 2009), SAMTools 'pileup' (Li et al., 2009), and Naive Variant Caller (Blankenberg et al., 2014) on the Galaxy online platform.

2.2. Phylogenetic analyses

DNA- and protein-based phylogenies of Polyplacophora were reconstructed using the sequences of all 13 mitochondrial protein-coding genes from *C. apiculata*, the other six available polyplacophoran species mentioned above, *Haliotis rubra* (Gastropoda; AY588938) (Maynard et al., 2005), and *Solemya velum* (Bivalvia; JQ728447) (Plazzi et al., 2013). The gene sequences were translated with the invertebrate mitochondrial genetic code to obtain the respective proteins, which were aligned with M-Coffee (Wallace et al., 2006). These protein alignments were then used to retro-align the respective codons using TranslatorX (Abascal et al., 2010). All alignments were trimmed with Gblocks 0.91b online (Castresana, 2000), specifying the respective sequence type ("Protein" or "Codon") and using the option "Do not allow many contiguous nonconserved positions" for a more stringent selection. Evolutionary models of nucleotide alignments were retrieved with jModelTest 2.1.4 (Darriba et al., 2012), using the Bayesian Information Criterion as criterion of choice. Gene and protein alignments were concatenated into two distinct alignments with Geneious 4.8.5 and used as input for MrBayes 3.2.2 (Ronquist et al., 2012) to reconstruct the two phylogenies. Both analyses consisted of two runs of four chains, each ran for 1,000,000 generations, sampling every 100 generations with a 1% relative burn-in. In the nucleotide-based analysis, the evolutionary models found by jModelTest were implemented into the input, specifying the '4by4' nucleotide substitution model, whereas in the protein-based analysis the rate matrix was specified as 'mixed'. At the end of the runs, trees were summarized using the 'sumt' command. Convergence of the runs was assessed by checking the standard deviation of average split frequencies trend over the generations: runs were considered good when they stabilized this value at < 0.01. Trees were graphically edited with FigTree 1.4.0 (Rambaut, 2012).

2.3. Gene order analyses

We mapped all known gene orders of Polyplacophora on the obtained phylogeny, and then we used CREx (Bernt et al., 2007) to infer which one could represent the most ancestral architecture for this class. The inferred plesiomorphic organization for polyplacophoran mtDNAs was then mapped onto a phylogeny of Mollusca summarizing the Aculifera hypothesis, together with the gene orders considered as most plesiomorphic of the other classes of molluscs. For Bivalvia, Cephalopoda, and Gastropoda we respectively used the gene orders of *Solemya velum* (Plazzi et al., 2013), *Vampyroteuthis infernalis* (Yokobori et al., 2007), and *Haliotis rubra* (Maynard et al., 2005), which are already hypothesized to be very ancient for their classes (Bernt et al., 2013a; Stöger and Schrödl, 2013). Two gene orders derived from complete mt genomes are available for each of the remaining classes Caudofoveata (*Chaetoderma nitidulum*, Dreyer H. and Steiner G., unpublished data; *Scutopus ventrolineatus*, Osca et al., 2014), Monoplacophora (*Laevipilina antarctica* and *Vema ewingi*; Stöger et al., 2016), and Scaphopoda (*Graptacme eborea*, Boore et al., 2004; *Siphonodentalium lobatum*, Dreyer and Steiner, 2004), and therefore all of them were used in our analysis. To avoid choosing any particular relationship among the five classes composing Conchifera (see Kocot, 2013) in our summary tree, we imposed a politomy at the base of their clade. Finally, the different gene orders were searched for the conserved gene blocks identified by Bernt et al. (2013a).

3. Results

3.1. Chaetopleura apiculata mt genome

14,902,649 paired-end assembled reads (42.3% of the initial 35,251,666 raw pairs) were selected as a starting pool to assemble the whole *C. apiculata* mt genome. The complete sequence is 15,108 bp long, and had median, average, and maximum coverages respectively of 479×, 458.9×, and 600×. The mitogenome contains the standard set of 37 genes (Fig. 1a; Appendix A Supp. Table 1), and has been deposited in GenBank under the accession number KY824658.

A highly variable portion with very low coverage inside the large unassigned region between *E* and *cox3* (Fig. 1a–b) precedes a segment composed of TA tandem repeats: given the uncertainties in assembling this unassigned region, we chose to insert in the final sequence the minimum number of repeats possible, i.e. five. This unassigned region has a length of 125 bp, similar to that of the related species *S. pelliserpentis* (169 bp), *S. sinclairi* (145 bp), *C. stelleri* (134 bp), and *K. tunicata* (141 bp). In all five species, this region preceding *cox3* contains a variable number of the dinucleotide TA, and a motif AAGTGYTAGCAA (followed by a long string of As only in *C. apiculata*, *C. stelleri*, and *K. tunicata*). In the Lepidochitonidae species *C. caverna* and *N. californica*, this region appears to have been translocated between *E* and *12S*