

Inverted Replication of Vertebrate Mitochondria

Miguel M. Fonseca,*†‡ David Posada,‡ and D. James Harris*†

*CiBiO/UP, Centro de Investigação em Biodiversidade e Recursos Genéticos, Vairão, Portugal; †Departamento de Zoologia e Antropologia, Faculdade de Ciências da Universidade do Porto, Porto, Portugal; and ‡Departamento de Bioquímica, Genética e Inmunología, Facultad de Biología, Universidad de Vigo, Vigo, Spain

After analyzing the base composition asymmetry of coding regions in vertebrate mitochondria, we identified 2 fishes, *Albula glossodonta* and *Bathygadus antrodes*, with inverted compositional patterns. Both species appear to have an unusual control region (CR), and in *B. antrodes*, it has switched from the light strand to the heavy strand. To our knowledge, this is the first report in vertebrates of inverted mitochondrial replication, caused by an inversion of the CR. These findings support the strand-asymmetric model of mtDNA replication and suggest that vertebrate mtDNA can tolerate globally reversed mutational pressures. In addition, we propose that nucleotide bias is not strand specific but that it depends on the location of the CR.

The organization and evolution of the mitochondrial DNA (mtDNA) are strongly related to its replication. Two models have been proposed to explain this process (Falkenberg et al. 2007). The “strand-asymmetric model” proposes that replication is unidirectional, starting from the origin of heavy strand replication located in the noncoding control region (CR) and displacing the parental heavy strand (H-strand), which becomes temporarily single-stranded DNA (ssDNA) (Clayton 1982). When the H-strand synthesis reaches two-thirds of the genome, it exposes the origin of light strand replication (O_L), initiating the synthesis of a new light strand (L-strand) in the opposite direction. On the other hand, the “strand-coupled model” proposes that mtDNA is synthesized from multiple, bidirectional origins (Reyes et al. 2005). This model describes ribonucleotide incorporation on the L-strand, acting as replication intermediates. The replacement of this RNA by DNA may initiate at dispersed sites, or around the O_L , or near the origin of the leading strand replication (Yasukawa et al. 2006). Importantly, although both models predict a delay between the synthesis of both strands, only the asymmetric model predicts a significant single-stranded exposure of the H-strand.

Although there is still no consensus about which of these models is correct, several attributes of mtDNA are only predicted by the strand-asymmetric model. Within these, the most important is the strand-specific bias (skew) in nucleotide composition (i.e., the H-strand is rich in G and T, whereas the L-strand is rich in A and C), explained by the exposition of the H-strand as ssDNA. The ssDNA is more prone to mutation/damage, especially to deaminations $C \rightarrow T$ and $A \rightarrow G$ (Frederico et al. 1990; Lindahl 1993). This different exposure to chemical damage between strands is confirmed by the strand-specific substitution patterns observed at 4-fold redundant sites, which should closely reflect the underlying patterns of mutation (Faith and Pollock 2003; Fonseca et al. 2006). In addition, the fact that redundant sites show the strongest bias suggests that selection may have a minor influence in the differences in nucleotide composition. On the other hand, it could be argued that this bias results mainly from errors that occur

during transcription and not during replication. In transcription, both strands become ssDNA, but deamination should occur more often in the L-strand (the “least-transcribed or “nontranscribed” strand). However, the observed mutational pattern is the opposite (Faith and Pollock 2003). Moreover, the asymmetrical model predicts the observed positive correlation between compositional bias and distance to the O_L (Reyes et al. 1998; Faith and Pollock 2003; Broughton and Reneau 2006). During asymmetrical replication, the parental H-strand becomes ssDNA while the nascent H-strand is synthesized, and genes closer to the O_L , in the direction of the L-strand synthesis, should remain exposed to mutation for less time.

Another important feature of the mtDNA is the CR, which regulates transcription and the initiation of replication. In most vertebrates and some invertebrates, the CR contains 3 conserved sequence blocks (CSB I, CSB II, and CSB III), which may be involved in the initiation of replication (Cao et al. 2004; Falkenberg et al. 2007). In mammalian mtDNA, the majority of the 5' ends of the nascent H-strand are located within or immediately downstream of CSB I (Clayton 1991), whereas CSB II increases the stability of the H-strand synthesis initiation (Xu and Clayton 1995). However, the absence of CSBs or their lack of conservation in some taxa suggests that other sequences may function as alternative regulatory elements. The location of the CSBs determines the leading strand during replication. Typically, CSBs are located in the L-strand, implying that the H-strand is the leading strand. Importantly, if CSBs switch strands, the L-strand becomes the leading strand and therefore the paternal L-strand will be temporarily single stranded and exposed to DNA damage during replication. Thus, the mutational scenario between strands will switch, and given enough time, the compositional bias and gradient may also revert, especially at synonymous positions.

In invertebrates, CR inversion has been proposed to be the cause of the reversion of the strand bias at 4-fold redundant sites in coding mtDNA (Scouras et al. 2004; Hassanin et al. 2005; Kilpert and Podsiadlowski 2006). In vertebrates, despite representing 80% of the available mtDNA sequences, CR inversions have never been detected. However, the increasing number of gene arrangements (San Mauro et al. 2006, Fonseca and Harris 2007), partial genome duplications (Fujita et al. 2007), and gene inversions (Amer and Kumazawa 2007) reported, suggest that vertebrate mtDNA should not be much constrained.

Key words: mtDNA, control region, inversion, strand bias, replication mechanism, vertebrates.

E-mail: mmfonseca@mail.icav.up.pt.

Mol. Biol. Evol. 25(5):805–808. 2008

doi:10.1093/molbev/msn050

Advance Access publication February 21, 2008

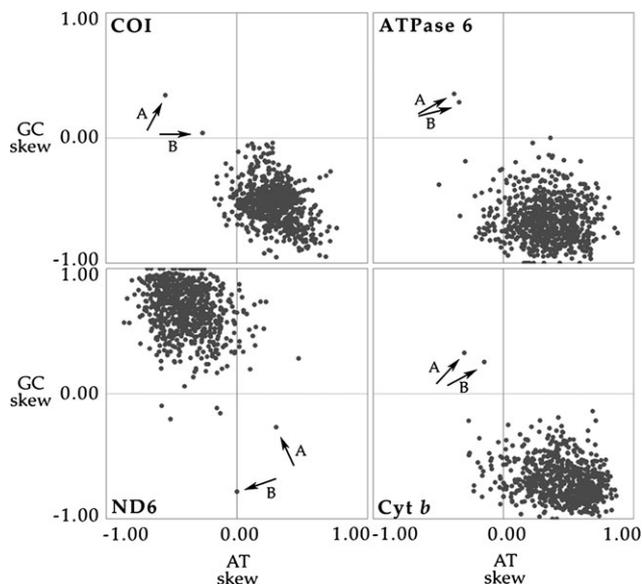


FIG. 1.—Plot of species-specific GC and AT skew at 4-fold redundant sites of 4 mitochondrial genes in 748 vertebrate taxa. We have chosen to present these 4 genes because they code for different proteins and they have the strongest inverted values for the species in state. Arrows highlight the species with reversed skews. A: *Albula glossodonta* and B: *Bathygadus antrodes*. National Center for Biotechnology Information GenBank accession numbers: AP002973 and AP008988 for species A and B, respectively. In species A, all genes have reversed GC and AT skews. In species B, COI, COII, ATPase6, ND3, Cyt b, and ND1 have reversed GC and AT skews; ATPase8, ND4L, ND4, ND5, ND2, and COIII have reversed AT skew; and ND6 has reversed GC skew.

Here, we report the first evidence for reversed mtDNA replication in vertebrates. We downloaded all 748 complete vertebrate mtDNA genomes available in GenBank (<http://www.ncbi.nlm.nih.gov>) in June 2007 and calculated the GC skew = $(G - C)/(G + C)$ and the AT skew = $(A - T)/(A + T)$ (Perma and Kocher 1995) at 4-fold redundant sites using BACA (Antão et al. 2007). The redundant codons examined were alanine (GCN), proline (CCN), serine (TCN), threonine (ACN), arginine (CGN), glycine (GGN), leucine (CTN), and valine (GTN). Two Teleostei fishes, *Albula glossodonta* and *Bathygadus antrodes*, showed significant and independent reversion of the GC and AT skews (figs. 1 and 2; if the complete genome constraint is relaxed and if all the complete coding segment for the mitochondrial genes are considered [multiplying sample size by 20], *A. glossodonta* and *B. antrodes* are still the 2 clear outliers [anonymous reviewer personal communication]). In both species, the CR presented atypical features. In *A. glossodonta*, the CR appears to be nonfunctional, as it is relatively short, very dissimilar from other CRs, and it has no CSBs, although it can form secondary structures. In addition to showing reversed skews and from the fact that the CR reverse complement sequence is more similar to a typical CR and may contain a putative CSB II (AACCCCCACTCCTTTCC), we did not find any additional features that could support its inversion. On the other hand, we found strong evidence for a CR inversion in *B. antrodes*. In this species, an H-strand noncoding fragment (flanked by 2 L-tRNAs, fig. 3), originally described

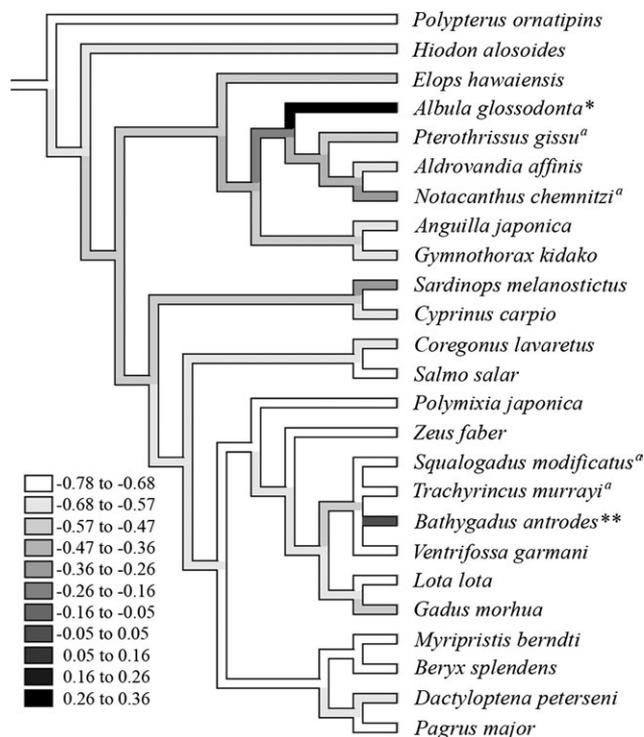


FIG. 2.—Consensus phylogeny of Teleostei fishes based on published phylogenetic trees. The evolution of the GC skew (using all 4-fold redundant sites) was reconstructed under the parsimony criterion using Mesquite (Maddison WP and Maddison DR 2007). Asterisk indicates species with unusual CR; double asterisk indicates species with inverted CR; and superscript a indicates species with CR not completely sequenced. The legend refers to GC skew values.

as an ND1-pseudogene (Satoh et al. 2006), is in fact the main CR: 1) it has significant sequence similarity with other CRs of closely related species (fig. 4), 2) it is capable of forming secondary structures, 3) it contains CSBs I and II in the same relative position (fig. 2; Saccone et al. 1991; Lee et al. 1995), and 4) the fragment located in the standard CR location is short and it appears to have no origin of replication (it does not have CSBs).

Although, ultimate proof can only be achieved with *in vitro* experiments, to the best of our knowledge, this constitutes the first known example in vertebrates of inversion of the mitochondrial replication mechanism, through an inversion of the CR.

These findings provide further support for the asymmetric model of replication and may be important to decipher, which conserved motifs of the CR are in fact crucial for an efficient replication/transcription. They also suggest that it is not necessarily correct to affirm that gene skew is strand specific. In these 2 examples, genes reverted their skew without changing the coding strand, and in *B. antrodes* only the CR inverted. Instead, we hypothesize that compositional bias depends much on which strand the CR (first origin of replication) is located. In addition, our results suggest that even vertebrate mtDNA may experience extreme structural changes and tolerate reversed mutational pressures, which may strongly influence the codon usage and amino acid composition (Min and Hickey 2007).

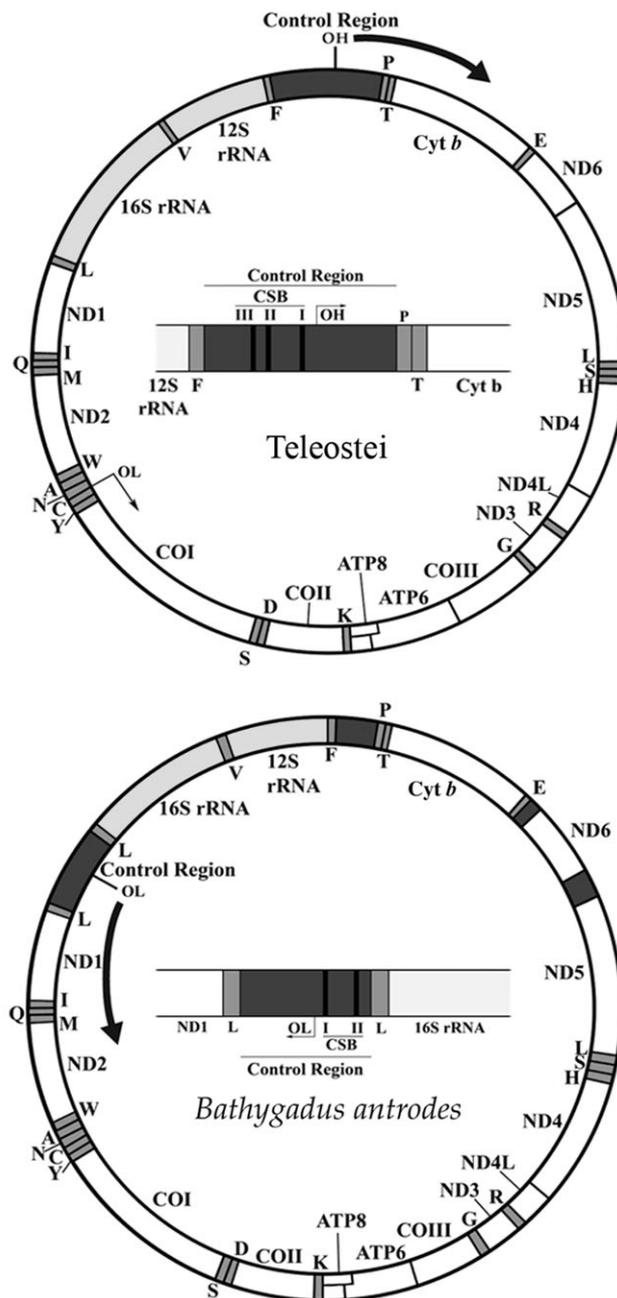


FIG. 3.—MtDNA organization. Typical Teleostei mtDNA is represented on the top and *Bathygadus antrodes* mtDNA is represented on the bottom. Genes transcribed from the H-strand are shown inside the circle and those transcribed from the L-strand are shown outside the circle. The location of the CSBs of the CR is also shown. Small arrows indicate the direction of replication, and big arrows highlight the replication origin and direction of the leading strand. Origin of lagging strand replication from *B. antrodes* was not clearly detected. Noncoding regions, dark gray; tRNAs, medium gray; rRNA, light gray; and protein-coding genes, white.

Indeed, it is important for accurate phylogenetic estimation to know the nucleotide bias across different taxa, as it has been shown that unrelated taxa with similar nucleotide composition tend to spuriously group together (Rosenberg and Kumar 2003; Hassanin et al. 2005).

Acknowledgments

We acknowledge R. Faria, R. Vasconcelos, and the anonymous reviewer for their helpful comments and suggestions. M.M.F. and D.J.H. are supported by FCT (Fundação para a Ciência e a Tecnologia) grants SFRH/BD/31158/2006 and POCI/65745/2006, respectively. D.P. is funded by the Spanish Ministry of Education and Science.

Literature Cited

- Amer SAM, Kumazawa Y. 2007. The mitochondrial genome of the lizard *Calotes versicolor* and a novel gene inversion in South Asian draconine agamids. *Mol Biol Evol.* 24:1330–1339.
- Antão T, Beja-Pereira A, Fonseca MM, Harris DJ. 2007. Baca: a mitochondrial genome retriever, organizer and visualizer. *Mol Ecol Notes.* 7:217–220.
- Broughton RE, Reneau PC. 2006. Spatial covariation of mutation and nonsynonymous substitution rates in vertebrate mitochondrial genomes. *Mol Biol Evol.* 23:1516–1524.
- Cao L, Kenchington E, Zouros E, Rodakis GC. 2004. Evidence that the large noncoding sequence is the main control region of maternally and paternally transmitted mitochondrial genomes of the marine mussel (*Mytilus* spp.). *Genetics.* 167:835–850.
- Clayton DA. 1982. Replication of animal mitochondrial DNA. *Cell.* 28:693–705.
- Clayton DA. 1991. Replication and transcription of vertebrate mitochondrial DNA. *Annu Rev Cell Biol.* 7:453–478.
- Faith JJ, Pollock DD. 2003. Likelihood analysis of asymmetrical mutation bias gradients in vertebrate mitochondrial genomes. *Genetics.* 165:735–745.
- Falkenberg M, Larsson NG, Gustafsson CM. 2007. DNA replication and transcription in mammalian mitochondria. *Annu Rev Biochem.* 76:679–699.
- Fonseca MM, Froufe E, Harris DJ. 2006. Mitochondrial gene rearrangements and partial genome duplications detected by multi-gene asymmetric compositional bias analysis. *J Mol Evol.* 63:654–661.
- Fonseca MM, Harris DJ. Forthcoming. 2007. Relationship between mitochondrial gene rearrangements and stability of origin of light strand replication. *Genet Mol Biol.*
- Frederico LA, Kunkel TA, Shaw BR. 1990. A sensitive genetic assay for the detection of cytosine deamination: determination of rate constants and the activation energy. *Biochemistry.* 29:2532–2537.
- Fujita MK, Boore JL, Moritz C. 2007. Multiple origins and rapid evolution of the duplicated mitochondrial genes in parthenogenetic geckos (*Heteronotia binoei*: squamata, Gekkonidae). *Mol Biol Evol.* 24:2775–2786.
- Hassanin A, Léger N, Deutsch J. 2005. 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.
- Lee WJ, Conroy J, Howell WH, Kocher TD. 1995. Structure and evolution of teleost mitochondrial control regions. *J Mol Evol.* 41:54–66.
- Lindahl T. 1993. Instability and decay of the primary structure of DNA. *Nature.* 362:709–715.
- Kilpert F, Podsiadlowski L. 2006. The complete mitochondrial genome of the common sea slater, *Ligia oceanica* (Crustacea,

```

G. morhua      TTTAAATCTGAATACCACTAAATATCGCCGCCGCCCTTTTAAATCTGAATACCACTAAATATCGCCGCCGCCCTTTTAAATATTGATTTTCGCT 150
B. antrodes    -----TTTATAACATATATTTCTACAAAGACATTAATATATGTCCTTAAAGTAGACTAGACGCTCGCC 63
                *****

G. morhua      TGAGCTTACCCCTCCTTTTCTCCCCCCCCCCGCTGCATATATGTCCTAGAGATCATATTTCTTTTAGTAAA--TTATTATTATACAT-ATGTAT 247
B. antrodes    AGGCCTTCCCCGCTTCTT-----CCCAATAGCGGGGAAAGTGA-TAAATACCTCAATTTAAATATTAAGGCTATAAGTGATATATTAGGTT 153
                *****

G. morhua      AATCACCATTA-ATTAAGTTAACCATACAAGGAGA-AATAATCATGAAAAGTCAAGCATTGAGGATAAAACAACAATATATTTATAGAACAATATGTT 345
B. antrodes    AATTTACGTTATGCTGACTAAGCCATACAGGTAATTAATACGTTCACTTAATTCGACCTT-AACATTAATTTATTTATGATATTTATTTATGGCAGAT 252
                *****

G. morhua      ATTTTAAACCAATTTATGGAATTTAGTGAAGAAATTTGTAACATAACCG--GACTTCTTGCCAAAGGCAACTGTCCAATGAAGGTGAGGGACACGTA 442
B. antrodes    TATTTATCTTCATAAGTAATCT--TGCAAGCATACGTTAAATCACCCTGAGAAATGCCCCA--ATGAAATACGTTCTTATGAAGGTCAAGGGCAGTTA 348
                *****

G. morhua      TTGAAGACCTGCATCCGTAACACGTTTCTGGCTATTCTGCTAGCTCAGGTCATGCTTGACAATCGCTCATAATTTGCACCTTTTGTACATCTCTT 542
B. antrodes    TTGCAACTCCCCATTTTATGACGTTTCTGGCC-ATTCAGCTAGTTTACGGCCATATTAATATATAGCTCATAGCTTACATTTTGTCCATCTCTA 447
                *****

G. morhua      AATGCTATACACATATATACTATAACTACTCCCATGCCGGGCTTCTTCTAATGGGCTACGGGTTCTTTTTTTTCTTCAAGTCATTTGACATTTT 642
B. antrodes    TATGGTGTATTTCATATA-ATGATAACTACTCAGCATGCCGAGCATTCTTCTAAGGGCAAGGGGTACTTTTTTTTCTTCCCTTCATTAAGCTATTC 546
                *****

G. morhua      AGCAGTGCAGAGCGTCGACCGCGGACAAGGTGGAGCTAGTCCGTTT-ATAGATATATAAAATTTATTTATTTAGTCCCCATAAAGATAATACATT 741
B. antrodes    A-CAGTGCACACGGTAAATATTAG-TAAGGGCAGCTATTCTTATTTTCAAGGATAAATTA-TAATGTTATAAACATTTATATATAAA-AAATACATT 642
                *****

G. morhua      AAGTTTTTCAAGAGCATAAGGCTAAAATTTTCTCGATGAGTCCCTAATATCTACATTTTACCCCTTTGTTTTTGTGCGTAAACCCCTCCGCCCA 841
B. antrodes    AATTTATCTCAGGGCATATATATAAATTTGATTTTAAAA---TAAAAATTAACAGATAAAACCTAAATAAATGT---AACCCCTTACGGCC 733
                *****

G. morhua      GTTCTCCTGAGATTACTAATATCTGTAACCCCGGAAACAGGAAAATCCCTAGAAGTGGTTTTTTTATCAAATATGACCAATAATGTATAAAT 941
B. antrodes    CTTTCTGCAATT-----TTGCAAGAAATTTTATACTAAAAATGACTGTTAACGGGATTTAA----- 795
                *****

G. morhua      TTGTGTTATTGCATTATTGCAAATTTAAAAATTT 977
B. antrodes    -----

```

FIG. 4.—Alignment of a standard CR (L-strand) from a Teleostei fish (*Gadus morhua*) with the inverted CR (H-strand sequence) from *Bathygadus antrodes*. The orientation is 5'–3'. CSBs are identified as darker boxes. Sequences were aligned using ClustalW (Thompson et al. 1994) with the default parameter values.

Isopoda) bears a novel gene order and unusual control region features. BMC Genomics. 7:241.

Maddison WP, Maddison DR. 2007. Mesquite: a modular system for evolutionary analysis. Version 2.0. Available from: <http://mesquiteproject.org>.

Min XJ, Hickey DA. 2007. DNA asymmetric strand bias affects the amino acid composition of mitochondrial proteins. DNA Res. 14:201–206.

Perna NT, Kocher TD. 1995. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. J Mol Evol. 41:353–358.

Reyes A, Gissi C, Pesole G, Saccone C. 1998. Asymmetrical directional mutation pressure in the mitochondrial genome of mammals. Mol Biol Evol. 15:957–966.

Reyes A, Yang MY, Bowmaker M, Holt IJ. 2005. Bidirectional replication initiates at sites throughout the mitochondrial genome of birds. J Biol Chem. 280:3242–3250.

Rosenberg MS, Kumar S. 2003. Heterogeneity of nucleotide frequencies among evolutionary lineages and phylogenetic inference. Mol Biol Evol. 20:610–621.

Saccone C, Pesole G, Sbisà E. 1991. The main regulatory region of mammalian mitochondrial DNA: structure-function model and evolutionary pattern. J Mol Evol. 33:83–91.

San Mauro D, Gower DJ, Zardoya R, Wilkinson M. 2006. A hotspot of gene order rearrangement by tandem duplication and random loss in the vertebrate mitochondrial genome. Mol Biol Evol. 23:227–234.

Satoh TP, Miya M, Endo H, Nishida M. 2006. Round and pointed-head grenadier fishes (Actinopterygii: Gadiformes) represent a single sister group: evidence from the complete mitochondrial genome sequences. Mol Phyl Evol. 40:129–138.

Scouras A, Beckenbach K, Arndt A, Smith MJ. 2004. Complete mitochondrial genome DNA sequence for two ophiuroids and a holothuroid: the utility of protein gene sequence and gene maps in the analyses of deep deuterostome phylogeny. Mol Phyl Evol. 31:50–65.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22:4673–4680.

Xu B, Clayton DA. 1995. A persistent RNA-DNA hybrid is formed during transcription at a phylogenetically conserved mitochondrial DNA sequence. Mol Cell Biol. 15:580–589.

Yasukawa T, Reyes A, Cluett TJ, Yang MY, Bowmaker M, Jacobs HT, Holt IJ. 2006. Replication of vertebrate mitochondrial DNA entails transient ribonucleotide incorporation throughout the lagging strand. EMBO J. 25:5358–5371.

Franz Lang, Associate Editor

Accepted February 15, 2008