J. Appl. Ichthyol. 24 (Suppl. 1) (2008), 46–49

ISSN 0175–8659

Received: August 31, 2007
Accepted: December 10, 2007
doi: 10.1111/j.1439-0426.2008.01089.x

Complete mitochondrial DNA sequence analysis of Ponto-Caspian sturgeon species

By S. Rastorguev, N. Mugue, A. Volkov and V. Barmintsev

Russian Federal Research Institute for Fisheries and Oceanography (VNIRO), Moscow, Russia

Summary

We present an analysis of sequence variation across complete mitochondrial (mt) genomes of eleven sturgeon species. We have sequenced five mt-genomes from four Ponto-Caspian species: *Acipenser gueldenstaedtii* (Russian sturgeon), *Acipenser persicus* (Persian sturgeon), *Acipenser naccarii* (Adriatic sturgeon), and *Acipenser baerii* (Siberian sturgeon). Two specimens of *A. gueldenstaedtii* were included representing the *gueldenstaedtii*-like haplotype and the *baerii*-like haplotype. Phylogenetic information content of each gene, clusters of genes sharing similar functions (tRNAs, protein coding sequences) and of the entire mitochondrial DNA genome was evaluated using the Shimodaira and Hasegawa test. Based on their suitability to reconstruct sturgeon phylogeny, mitochondrial genes were ranked between 'very good' (*ND5*, *cytb* – large agreement between phylogenies based on these genes and the phylogeny calculated on the entire mt genome), 'good' (*COI*, *ND4*, *ATP6*), 'medium' (*ATP8*, *COII*, *COIII*, *ND1*, *ND2*, *ND3*, *ND4L*) and 'weak' (*ND6* – tree topology is significantly deviated from the expected one). The maximum parsimony (MP) tree based on entire mt genomes showed that *Scaphirinchus* is a sister clade of the genera *Acipencer* and *Huso*. However, maximum likelihood (ML) algorithm with the most suitable nucleotide evolution model (GTR + I + G) places *Scaphyrinchus* within *Acipenser* as a sister branch to Ponto-Caspian species. Both algorithms (MP, ML) agreed in the following points: (i) Atlantic and Pacific species are separated in different clades, *Huso huso* is basal with the Atlantic species and (ii) all species of the *gueldenstadetii*-complex are closely related. Finally, eight substitutions between *A. baerii* and the *baerii*-like specimen of *A. gueldenstaedtii* were found overall protein coding sequences, and 131 substitutions were detected between the *gueldenstaedtii*-like and *baerii*-like haplotype over the entire mt genome. Taking the small sample size under consideration, these differences have to be handled with care and should be addressed in further studies.

Introduction

The family Acipenseridae consists of four genera (*Acipenser, Huso, Scaphirhynchus*, and *Pseudoscaphirhynchus*) taking together 25 extant species. Recently published molecular phylogenies are not congruent in several points with the classical taxonomy based on morphology (Artyukhin, 2006). For example both species of the genus *Huso* (*Huso huso* and *Huso dauricus*) were grouped within the species of the genus *Acipenser* (Birstein and DeSalle, 1998) and separated in two different clusters representing their zoogeographic distribution (Ludwig et al., 2001; Peng et al., 2007). Additionally, the genus *Pseudoscaphirhynchus* was also grouped within the genus *Acipenser* most closely connected to *Acipenser stellatus* (Birstein et al., 2002). Finally, the genus *Scaphynchinchus* was an internal clade within the genus *Acipencer* (Peng et al., 2007) or its position is unresolved (Krieger et al., 2008).

Most of the previous studies based on single mitochondrial (mt) genes [mostly *cytochrome b* (*cytb*) – Ludwig et al., 2000, 2001] or on a few mt genes (Krieger et al., 2008) or sometimes only on parts of them (Birstein and DeSale, 1998). To overcome the limitations of these limited data sets, we used entire mt genomes for phylogenetic reconstruction.

Besides phylogenetic questions, we would like to address the identification of characteristic substitutions for species and population identification. Depending on different evolutionary constrains, mt genes have a different degree of inter- and intraspecific variability. Sequence analysis of entire mt genomes is time consumed and expensive, therefore we would like to identify these regions which are characterized by a large level of inter- or intraspecific variability. Although during the second Status Workshop for Identification of Acipenseriformes in Trade in Berlin September 30–October 1, 2006 was decided to focus on *cytb* sequences for molecular species identification (Ludwig et al., 2002; Ludwig, 2008), additional informative regions could be fruitful especially for closely related species (Ludwig, 2006).

There were several recent studies on individual mitochondrial DNA (mtDNA) variability and its suitability for phylogenetic analysis. For example, based on maximum parsimony (MP), maximum likelihood (ML), and neighbor joining (NJ) analyses of several vertebrates, the following classification was proposed for mitochondrial genes: good genes are *ND4*, *ND5*, *ND2*, *cytb*, *COI*; medium genes are *COII*, *COIII*, *ND1*, *ND6*; and poor genes are *ATP6*, *ND3*, *ATP8*, *ND4L* (Zardoya and Meyer, 1996). Miya and Nishida (2000) classified the same genes especially for teleostei as follows: very good (*ND5*, *ND4*, *COIII*, *COI*); good (*COII*, *cytb*); medium (*ND3*, *ND2*); poor (*ND1*, *ATP6*); and very poor (*ND4L*, *ND6*, *ATP8*). Although there seem to be a large species-overlapping information content among all vertebrates, there is always a rest of lineage specificity in such data sets. Therefore, we test the sturgeon specific information content of all mt sequences.

Materials and methods

At present the following complete mtDNA sequences are deposited in GenBank (Table 1): *H. huso*, *A. stellatus*, *Acipenser transmontanus*, *Acipenser mikadoi*, *Scaphirhynchus cf. albus*, *Polyodon spathula*. We sequenced the complete mtDNA genome of five specimens from four sturgeon species. *Acipenser naccari*, *Acipenser persicus*, and *Acipenser baerii*, and *Acipenser gueldenstaedtii* (n = 2). We included two specimens

Table 1
Origin of samples, genbank accession numbers and collection numbers

Species	Sampling location	GenBank accession no.	Reference	RNCSGM no.
Acipenser baerii	Irtish River	Pending	This study	BAE0154
Acipenser gueldenstaedtii (typical mitotype)	Volga (Volgogradskiy ORZ)	Pending	This study	GUE0830
Acipenser gueldenstaedtii (*baerii* mitotype)	Caspian Sea (Zhemchuzhniy Isl.)	Pending	This study	GUE0365
Acipenser naccarii	Sierra Nevada Factoria r. Riofrio Spain.	Pending	This study	NAC0020
Acipencer persicus	Caspian Sea (near Turkmenian coast)	Pending	This study	PER0161
Acipenser stellatus		AJ585050	Arnason et al. (2004)	
Acipenser transmontanus		NC_004743	Inoue et al. (2003)	
Huso huso		NC_005252	Dunn et al., direct submission Genbank	
Acipenser dabryanus		AY510085	Peng et al. (2007)	
Scaphirhynchus cf. *albus*		NC_004420	Inoue et al. (2003)	
Polyodon spathula		AY510086	Peng et al. (2007)	

of *A. gueldenstaedtii* because this species is divided into two major haplogroups: one haplogroup possesses typical *gueldenstaedtii*-like haplotypes, and the other haplogroup has *baerii*-like haplotypes. We sequenced for each haplogroup one specimen.

Ethanol fixed tissue samples were obtained from the Russian National Collection of Standard Genetic Materials (RNCSGM) of sturgeon fishes (maintained by VNIRO). Primers were designed by multiple alignments of complete mtDNA sequences deposited at GenBank. Primers design and amplification conditions were done using the FastPCR program (Kalendar, 2007).

We designed 84 primers, which were used for amplification of 42 overlapping PCR-fragments. List of primers and PCR conditions are available on request. To increase PCR efficiency and to avoid non-specific product amplification, a two stages nested PCR was used. The first stage produced ~2400-bp product, and the second stage of nested PCR resulted in a ~1200-bp product. The smaller one was cleaned and used for further sequencing. Final PCR products (42 amplicons, ~1200 bp) covered all mtDNA at least two times. PCR was done in a DNA engine PTC tetrad 2 thermal cycler (MJ Research) and reactions carried out with 30 cycles of a 15 μl reaction volume containing H_2O, 1.5 μl of 10× PCR buffer (Sileks), 1.2 μl NTP's (2.5 mM each), 1.5 μl of each primer (5 μM), 0.06 μl of 0.5 U *Taq* polymerase (Sileks), and 0.6 μl DNA template. Cycle profile was as follows: denaturation at 94°C for 15 s, annealing at 50°C (but sometimes it was adjusted according to primer properties) for 30 s, and extension at 72°C for 90 s. PCR products were run on a 1.0% agarose gel and stained with ethidium bromide. Double-stranded PCR products were purified enzymatically by treatment with *Exo*I (Fermentas) and phosphatase SAP (USBiological) according to Werle et al. (1994), or by gel excising using a DNA Extraction Kit (Fermentas). Purified PCR product were subsequently used for direct cycle sequencing with BigDye v.1.1 Sequencing Kit (ABI). All sequencing reactions were performed according to the manufacturer's instructions on a 3100 DNA sequencer (ABI). Primary sequence analysis was performed using Sequencing Analysis software (ABI). Condon reconstruction was done with CodonCode Aligner v.1.5.2. (CodonCode Corporation).

Phylogenetic analyses were done on complete mitochondrial genomes and on protein-coding genes. Control regions were excluded from phylogenetic reconstructions depending on heteroplasmy and large numbers of indels. MP and ML reconstructions were performed using PAUP*4.0b10 (Swofford, 2001). The best-fit model of nucleotide substitution was assessed by ModelTest 3.7 (Posada and Crandall, 1998). The statistical confidence of the resulting best trees of each ML analysis was calculated by corrected non-parametric test of topologies, proposed by Shimodaira and Hasegawa (1999). Shimodaira and Hasegawa (SH) test implemented in PAUP*4.0b10 (Swofford, 2001) by using RELL bootstrap (one-tailed test). This was done by comparing gene topology with the expected topology. We used the best fitting evolutionary model, which were separately calculated for each gene or functional group of genes. This allows estimate the information content of each genes / group of genes for sturgeon phylogeny. Estimation of dN / dS ratio for protein coding sequences was performed by PALM program (Yang, 2000).

Results

GenBank accession numbers are presented in Table 1. Analysis of complete mtDNA sequences confirmed a conserved order of genes overall sturgeons. Gene order is identical to the order observed in other teleostei (Arnason et al., 2004).

Maximum parsimony topology based on complete mitochondrial genomes (Fig. 3) is concordant with previously published phylogenies based on *cytb* gene (Fontana et al., 2001; Ludwig et al., 2001; Peng et al., 2007), and on control region variation (Ludwig et al., 2000). But ML topology differed from previously published topologies by the grouping of *Scaphirhynchus*, which became a sister clade to the Ponto-Caspian species within the genus Acipenser (Fig. 4). The same topology is obtained by ML analysis of all protein-coding sequences. Within the closely related *gueldenstaedtii*-species complex (*A. gueldenstaedtii*, *A. naccarii*, *A. persicus*), *A. gueldenstaedtii* is most closely related to *A. persicus*, but bootstrap support is for this cluster low (55%). Deleting third codon position, phylogenetic relationships are not resolved for *A. gueldenstaedtii*, *A. naccarii*, and *A. persicus*.

Swofford et al. (2001) showed that in analysis of extensive DNA sequences ML algorithm is superior to MP in robustness of phylogenetic reconstruction. Therefore, further analyses of separated genes were focused on ML. The genus Scaphirhynchus was excluded from this analysis resulting from its unsafe grouping in MP and ML trees. Results of the SH-test are presented in Fig. 1. Two SH tests were conducted for each rRNA and protein coding gene, unified tRNA genes, and

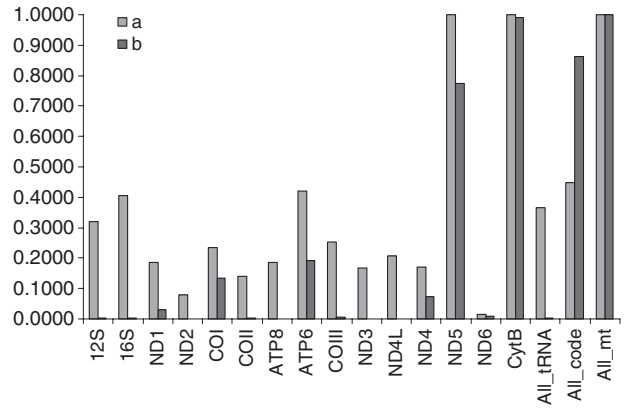

Fig. 1. Comparative analysis of individual maximum likelihood trees with expected one was done by Shimodaira–Hasegawa test (Shimodaira and Hasegawa, 1999) concordance of individual gene sequence data with topology of 'expected', or complete mtDNA dataset tree (a), and reciprocal test of concordance of the whole genome mt sequence with tree topology built on individual gene or functional groups of genes (b)

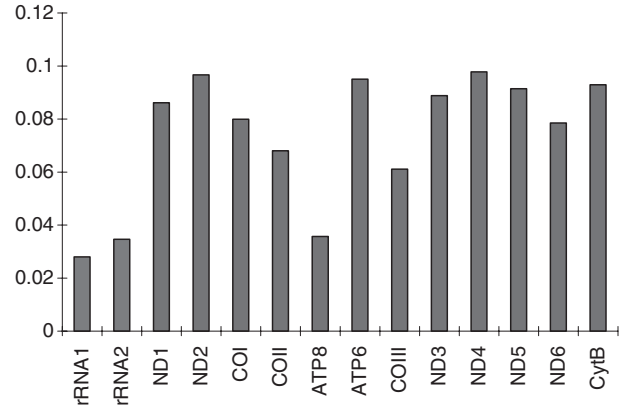

Fig. 2. Relative number of polymorphic sites for eleven sturgeon species

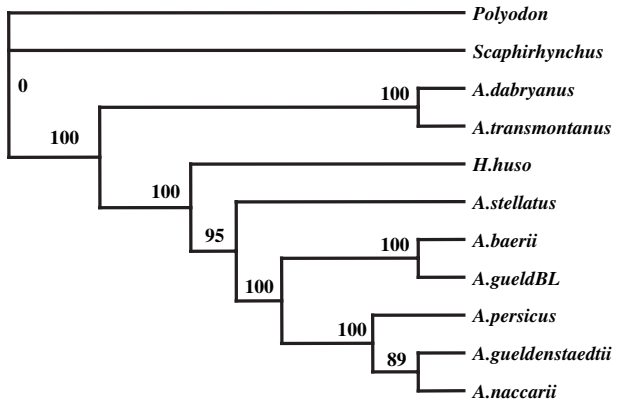

Fig. 3. Maximum parsimony tree based on complete mt genomes. Bootstrap support after 1000 replications is shown

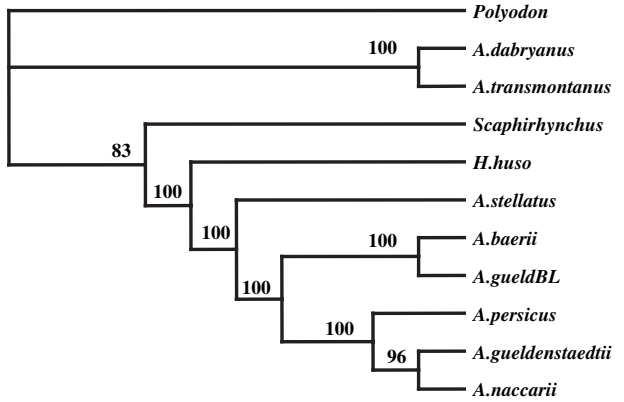

Fig. 4. Maximum likelihood tree based on complete *Acipenseridae* mt genomes with (GTR + I + G) nucleotide evolution model. Bootstrap support after 100 replications is shown

protein coding genes: (i) agreement between topology based on each gene and the 'expected' topology or based on entire mtDNA dataset (left bars on Fig. 1) as well as (ii) the reciprocal test (right bars on Fig. 1). Gene polymorphism was calculated as a proportion of variable sites to the total gene length (Fig. 2). All protein coding genes have a more or less similar degree of variability, and their variability's are about twice times larger than the values observed for ribosomal genes. Interestingly, remarkable variation was found especially in very short genes (for example *ATP8*).

Discussion

All Ponto-Caspian species were unified in one clade in both algorithms MP and ML. Using MP, Ponto-Caspian species are the sister clade of Pacific species (Fig. 3). However, in ML analysis *Scaphirhynchus* is grouped together with Ponto-Caspian species (Fig. 4). Grouping within the Ponto-Caspian clade is similar in both approaches and in agreement with the previously published molecular phylogenies (Birstein et al., 2002; Peng et al., 2007; Krieger et al., 2008).

The relationship of *baerii*-like haplotypes (Jenneckens et al., 2000) within the *A. gueldenstaedtii* clade is of special interest for both trade control and management of Russian sturgeons of the Northern Caspian Sea because these closely connected species (Ludwig et al., 2003; Birstein et al., 2005) are important caviar producers. Only eight substitutions between *A. baerii* and the *baerii*-like specimen of *A. gueldenstaedtii* were found overall protein coding sequences. However, taking entire genomes under consideration, 131 substitutions were detected between *gueldenstaedtii*-like and *baerii*-like haplotypes. Because only single representatives for both haplogroups were analyzed, we can draw no conclusions if any of these substitutions is species(haplogroup)-specific. Larger sample sets should be screened for these substitutions. Almost fixed diagnostic substitution separating *A. baerii* and *baerii*-like haplotypes of *A. gueldenstaedtii* are discussed in an additional article within this issue (Voynova et al., 2008).

Several phylogenies based on single genes have lower phylogenetic information in relation to the molecular phylogeny based on the entire genome. However, only the *ND6* phylogeny is significantly different. This is in agreement with the previously published observation based on primates (Yoder and Yang, 2000) that *ND6* has very different codon usage and substitution patterns. But besides the *ND6* noise in phylogenetic reconstructions, the impact of this gene is not significant for phylogenetic reconstruction. Parameters of nucleotide evolution model (GTR + I + G) are similar and tree topology remains constant if *ND6* is excluded or not. Taken together, we propose the following classification of sturgeon mitochondrial genes: (i) very good (*ND5*, *cytb*), good (*COI*,

ND4, *ATP6*), medium (*ATP8*, *COII*, *COIII*, ND1, *ND2*, *ND3*, *ND4L*) and poor (*ND6*). Nevertheless, the robustness of these analyses increases with the number of specimens incorporating intraspecific variation often observed within a species (Zardoya and Meyer, 1996).

Additionally, the position of the Scaphyrinchus clade requires further analysis. Our MP tree is concordant with previously published and widely accepted phylogeny that *Scaphirinchus* is a sister clade of the genera *Acipencer* and *Huso*. Nevertheless, ML places *Scaphyrinchus* within *Acipenser* having high statistical support. This grouping is in agreement with recently published data (Dillman et al., 2007). Further analyses are necessary to solve this topic.

References

Arnason, U.; Gullberg, A.; Janke, A.; Joss, J.; Elmerot, C., 2004: Mitogenomic analyses of deep gnathostome divergences: a fish is a fish. Gene **333,** 61–70.

Artyukhin, E. N., 2006: Morphological phylogeny of the order Acipenseriformes. J. Appl. Ichthyol. **22** (Suppl. 1)**,** 66–69.

Birstein, V. J.; DeSalle, R., 1998: Molecular phylogeny of Acipenserinae. Mol. Phylogenet. Evol. **9,** 141–155.

Birstein, V. J.; Doukakis, P.; DeSalle, R., 2002: Molecular phylogeny of Acipenseridae: Nonmonophyly of Scaphynchinae. Copeia **2002,** 287–301.

Birstein, V. J.; Ruban, G.; Ludwig, A.; Doukakis, P.; DeSalle, R., 2005: The enigmatic Caspian Sea Russian sturgeon: how many cryptic forms does it contain? System Biodivers. **3,** 203–218.

Dillman, C. B.; Wood, R. M.; Kuhajda, B. R.; Ray, J. M.; Salnikov, V. B.; Mayden, R. L., 2007: Molecular systematics of Scaphirhynchinae: an assessment of North American and Central Asian freshwater sturgeon species. J. Appl. Ichthyol. **23,** 290–296.

Fontana, F.; Tagliavini, J.; Congiu, C., 2001: Sturgeon genetics and cytogenetics: recent advancements and perspectives. Genetica **111,** 359–373.

Inoue, J. G.; Miya, M.; Tsukamoto, K.; Nishida, M., 2003: Basal actinopterygian relationships: a mitogenomic perspective on the phylogeny of the 'ancient fish'. Mol. Phylogenet. Evol. **26,** 110–120.

Jenneckens, I.; Meyer, J.-N.; Debus, L.; Pitra, C.; Ludwig, A., 2000: Evidence of mitochondrial DNA clones of Siberian sturgeon, *Acipenser baerii*, within Russian sturgeon, *Acipenser gueldenstaedtii*, caught in the River Volga. Ecol. Lett. **3,** 503–508.

Kalendar, R., 2007: FastPCR: a PCR primer design and repeat sequence searching software with additional tools for the manipulation and analysis of DNA and protein. Available at: http://www.biocenter.helsinki.fi/bi/programs/fastpcr.htm (accessed July 2007).

Krieger, J.; Hett, A. K.; Fuerst, P. A.; Artyukhin, E.; Ludwig, A., 2008: The molecular phylogeny of the order Acipenseriformes revisited. J. Appl. Ichthyol. **24** (Suppl. 1), 36–45.

Ludwig, A., 2006: A sturgeon view on conservation genetics. Eur. J. Wildl. Res. **52,** 3–8.

Ludwig, A., 2008: Identification of Acipenseriformes species in trade. J. Appl. Ichthyol. **24** (Suppl. 1), 2–19.

Ludwig, A.; May, B.; Debus, L.; Jenneckens, I., 2000: Heteroplasmy in the mtDNA control region of sturgeon (*Acipenser*, *Huso* and *Scaphirhynchus*). Genetics **156,** 1933–1947.

Ludwig, A.; Belfiore, N. M.; Pitra, C.; Svirsky, V.; Jenneckens, I., 2001: Genome duplication events and functional reduction of ploidy levels in sturgeon (*Acipenser*, *Huso* and *Scaphirhynchus*). Genetics **158,** 1203–1215.

Ludwig, A.; Debus, L.; Jenneckens, I., 2002: A molecular approach for trading control of black caviar. Int. Rev. Hydrobiol. **87,** 661–674.

Ludwig, A.; Congiu, L.; Pitra, C.; Fickel, J.; Gessner, J.; Fontana, F.; Patarnello, T.; Zane, L., 2003: Nonconcordant evolutionary history of maternal and paternal lineages in Adriatic sturgeon. Mol. Ecol. **12,** 3253–3264.

Miya, M.; Nishida, M., 2000: Use of mitogenomic information in teleostean molecular phylogenetics: a tree-based exploration under the maximum-parsimony optimality criterion. Mol. Phylogenet. Evol. **17,** 437–455.

Peng, Z.; Ludwig, A.; Wang, D.; Diogo, R.; Wei, Q.; He, Sh., 2007: Age and biogeography of major clades in sturgeons and paddlefishes (Pisces: Acipenseriformes). Mol. Phylogenet. Evol. **42,** 854–862.

Posada, D.; Crandall, K. A., 1998: MODELTEST: testing the model of DNA substitution. Bioinformatics **14,** 817–818.

Shimodaira, H.; Hasegawa, M., 1999: Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. **16,** 1114–1116.

Swofford, D. L., 2001: PAUP* version 4.0b10. Phylogenetic analysis using parsimony (and other methods). Sinauer Associates, Sunderland, MA.

Swofford, D. L.; Waddell, P. J.; Huelsenbeck, J. P.; Foster, P. G.; Lewis, P. O.; Rogers, J. S., 2001: Bias in phylogenetic estimation and its relevance to the choice between parsimony and likelihood methods. Syst. Biol. **50,** 525–539.

Voynova, N. V.; Mirzoyan, A. V.; Timoshkina, N. N.; Rynza, E. T., 2008: Occurrence of non-native specimens of Caspian origin within the Sea of Azov population of the Russian sturgeon (*Acipenser gueldenstaedtii*). J. Appl. Ichthyol. **24** (Suppl. 1), 50–51.

Werle, E.; Schneider, C.; Renner, M.; Völker, M.; Fiehn, W., 1994: Convenient single-step, one tube purification of PCR products for direct sequencing. Nuc. Acids Res. **22,** 4354–4355.

Yang, Z., 2000: Phylogenetic analysis by maximum likelihood (PAML), 3rd edn. University College London, London.

Yoder, A. D.; Yang, Z., 2000: Estimation of primate speciation dates using local molecular clocks. Mol. Biol. Evol. **17,** 1081–1090.

Zardoya, R.; Meyer, A., 1996: Phylogenetic performance of mitochondrial protein-coding genes in resolving relationships among vertebrates. Mol. Biol. Evol. **13,** 933–942.

Author's address: Sergey Rastorguev, Russian Federal Research Institute for Fisheries and Oceanography (VNIRO), 17, V.Krasnoselskaya, 107140 Moscow, Russia.
E-mail: rastorgueff@gmail.com

The authors have not declared any conflict of interests.