Molecular phylogenetics of pricklebacks and other percoid fishes from the Sea of Japan

Yuri P. Kartavtsev[1,2,*], Svetlana N. Sharina[1], Tadasuke Goto[3], Olesya A. Rutenko[2], Vladimir V. Zemnukhov[1], Alexander A. Semenchenko[2], Dmitri L. Pitruk[1], Naoto Hanzawa[3]

[1]**A.V. Zhirmunsky Institute of Marine Biology, FEB RAS, Vladivostok 690041, Russia**
[2]**Far Eastern State University, Vladivostok 690095, Russia**
[3]**Department of Biology, Faculty of Science, Yamagata University, Yamagata 990-8560, Japan**

ABSTRACT: Mitochondrial DNA of the *Co-1* gene region was sequenced for 8 fish species (in total 17 sequences of at least 532 bp) from the far east of Russia and compared to 3 other perciform sequences, altogether comprising 20 perch-like fish sequences and 2 outgroup sequences (Cypriniformes). The analysis of the protein-coding *Co-1* gene revealed a statistically substantiated bias in the (T+C):(A+G) content, confirming the basic findings on the pyrimidines-prunes ratio. The average scores of *p*-distances for different scales of the evolutionary history of the *Co-1* gene revealed a pattern of increased nucleotide diversity at 4 different levels: (1) intraspecies, (2) intragenus, (3) intrafamily and (4) intraorder. The scores of the average *p*-distances of the 4 categories of comparison were: (1) $0.11 \pm 0.04\%$, (2) $1.87 \pm 0.68\%$, (3) $12.67 \pm 0.28\%$, and (4) $16.52 \pm 0.10\%$, respectively (mean ± SE). These data support the concept that speciation in the order Perciformes in most cases follows a geographic mode through accumulation of numerous small genetic changes over a long time. The phylogenetic trees for 18 and 21 sequences of perch-like fishes together with 2 other fishes belonging to the ray-finned fishes (Actinopterygii) were developed using the *Co-1* gene and 4 different analytical approaches: neighbour-joining (NJ), maximum likelihood (ML), Bayesian (BA) and maximum parsimony (MP). The analysis revealed a monophyletic origin for the representatives of the Stichaeidae, which was the principal percoid family investigated (86, 96 and 100% support in our NJ, ML and BA analyses, respectively). Species identification on a per individual basis or *Co-1*-based DNA barcoding was high.

KEY WORDS: *Co-1* · cytochrome c oxidase 1 · Mitochondrial DNA · Identification · Percoid

INTRODUCTION

Mitochondrial DNA (mtDNA) is now widely used for phylogenetic purposes for a wide range of taxa. Many regions of mtDNA that code for protein genes or participate in regulation as the control region are used as genetic markers for investigation of intra- and interspecies diversity. mtDNA is able to accumulate many base substitutions over a long period of time, providing a comparative tool for taxonomic, evolutionary and phylogenetic research (Nei 1987, Wallace 1992, Avise 2000). Molecular phylogenetic approaches use sequence diversities of a single gene, multiple genes or the complete mitochondrial genome (mitogenome).

Very popular for such research are sequences of the cytochrome oxidase 1 (*Co-1*) and cytochrome *b* (*Cyt-b*) genes, which are best suited for analyses from the species up to the family level (Johns & Avise 1998, Hebert et al. 2004, Kartavtsev & Lee 2006). However, single-gene approaches provide inadequate phylogenetic data when applied to higher taxa (above the order level), a result of insufficient information capacity and homoplasy effects (Hillis et al. 1996). Another complication in the successful application of mtDNA sequence data, as well as any other markers, is a vast variation in the degree of genetic diversity within many taxa (Johns & Avise 1998, Hebert et al. 2004, Kartavtsev & Lee 2006). Thus, a simple distance-like

ranking even among vertebrate animals is quite a complicated task. Nevertheless, the bulk of evidence demonstrates that sequence data, e.g. for *Co-1* or *Cyt-b*, are a useful tool for species discrimination (as DNA barcoding tags) and very helpful for phylogenetic reconstructions, including numerous fish taxa (Avise 2000, Hebert et al. 2004, Kim et al. 2004, Ward et al. 2005, Kartavtsev & Lee 2006). Pooling data such as *p*-distances (Nei & Kumar 2000), or their derivates, across different taxon levels gives repetitive support for an increase of genetic diversity starting at the intraspecies level and up to higher taxonomic ranks (Johns & Avise 1998, Hebert et al. 2004, Kartavtsev 2005, Ward et al. 2005, Kartavtsev & Lee 2006, Kartavtsev et al. 2007a,b). This positive and proportional distance score increase along with a taxon rank increase may be used as a helpful attribute in searching for the most common speciation mode, a geographic mode (Ayala et al. 1974, Avise 2000). However, because of the complexity of species criteria, distance data alone may give only a part of the necessary information for discrimination among speciation modes (Kartavtsev et al. 2002, Kartavtsev 2005, Kartavtsev & Lee 2006).

The need for molecular genetic research in percoid fish species becomes apparent when considering taxonomic complications and population structure analysis. There are numerous taxonomic problems associated with many groups and certain perch-like species of Perciformes (Nelson 2006), which is one of the largest fish orders. Most families in many suborders are not currently definable in terms of shared derived characters and thus may not be monophyletic. This is especially true for those species that have a low commercial value, like pricklebacks (Stichaeidae). For instance, the family Stichaeidae has no consensus position in taxonomic revisions (e.g. Makushok 1986a,b, Nelson 2006). However, in accordance with the Integrated Taxonomic Information System (IT IS: www.itis.gov/), it does hold such a position. Currently (17 December 2009), 6 genera are placed in the family Stichaeidae by the NCBI Taxonomic Browser (http://www.ncbi.nlm. nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Tree& id=30892&lvl=3&lin=f&keep=1&srchmode=1&unlock). Meanwhile, 41 genera were placed within the family in an earlier version of FishBase (Froese & Pauly 2009), and 77 species of 38 genera are currently listed in the family Stichaeidae in a more recent version of FishBase (Froese & Pauly 2009). Thus, it is obvious that much molecular taxonomy and phylogenetic research on Stichaeidae and other perch-like fish species remains to be conducted.

Many molecular marker investigations in perch-like fishes with respect to phylogeny have been reported recently (Finnerty & Block 1995, Lydeard & Roe 1997, Song et al. 1998, Reed et al. 2002, Kim et al. 2004).

Some fishes in the family Stichaeidae have been investigated by phylogeneticists (Ward et al. 2005, Ward & Holmes 2007, Radchenko et al. 2009). To our knowledge, however, only a few representatives were compared simultaneously by using sequence divergence at *Co-1*. Promising results for percoids were obtained at other loci, such as 12S, 16S rRNA and *Cyt-b* of mtDNA that definitely showed monophyly of some investigated tribes and families in the order (Stepien et al. 1997, Reed et al. 2002, Orrell & Carpenter 2004, Kim et al. 2004). Serious disagreements with current taxonomy were also found at several DNA markers (Craig & Hastings 2007, Radchenko et al. 2009).

During a BLAST search of NCBI GenBank (www. ncbi.nlm.nih.gov) based on the fish *Pholidapus dybovskii,* a reference sequence investigation for close proximity among *Co-1* sequences revealed 1 hit with only a few mtDNA sequences, most of which are short partial sequences. Nevertheless, our search for *Co-1* provided several quite long sequences for the gene (~650 bp), which allowed us to make a molecular taxonomic and phylogenetic evaluation for perch-like fishes on this basis. However, taking into consideration the large diversity in the order Perciformes, which currently comprises 156 recent families, with 77 species in the family Stichaeidae alone (Froese & Pauly 2009), the possibility of obtaining an absolutely reliable molecular phylogenetic topology for the whole order Perciformes, from *Co-1*, *Cyt-b* or any other single gene, seems remote.

Here, we focus on the family Stichaeidae and discuss 3 to 4 representatives of other percoid fish families by molecular phylogenetic and taxonomic considerations using *Co-1* sequence data. Two main questions were considered: (1) whether the family Stichaeidae is monophyletic, and (2) whether the nucleotide diversity at *Co-1* supports the currently accepted intra-family and intra-genus divisions.

MATERIALS AND METHODS

In total, we sampled 17 white muscle tissue specimens of 8 percoid fish species sampled in 2005–2006: (1) *Stichaeus grigorjewi* (Herzenstein, 1890), (2) *S. nozawae* (Jordan & Snyder, 1902), (3) *Ammodytes hexapterus* (Pallas, 1814), (4) *Arctoscopus japonicus* (Steindachner, 1881), (5) *Chirolophis japonicus* (Herzenstein, 1890), (6) *Pholidapus dybowskii* (Steindachner, 1881), (7) *Opisthocentrus ocellatus* (Tilesius, 1811), (8) *O. tenuis* Bean & Bean, 1897, and 2 specimens of a single cyprinid species, *Tribolodon brandtii* (Dybowski, 1872) as an outgroup (Table 1). Muscle tissues were taken from all individuals, which were collected by gill nets in April to September in Vostok Bay (132° 44′

Table 1. GenBank accession numbers for *Co-1* gene sequences and location of collections of the examined species. Alcohol-preserved specimens and voucher individuals were deposited in collections at the A.V. Zhirmunsky Institute of Marine Biology

No.	Accession number	Species and specimen	Date (dd/mm/yy)
Peter the Great Bay, Sea of Japan, Russia			
1	EU200479	*Stichaeus grigorjewi* 8A	29/04/06
2	EU200485	*Stichaeus nozawae* 13A	29/04/06
3	EU200482	*Ammodytes hexapterus* 10A	29/04/06
4	EU200482	*Arctoscopus japonicus* 12A	29/04/06
5	EU200494	*Chirolophis japonicus* 39	12/09/06
6	EU200501	*Chirolophis japonicus* 50	13/09/06
7	EU200496	*Pholidapus dybowskii* 41	12/09/06
8	EU200499	*Pholidapus dybowskii* 48	13/09/06
9	EU200470	*Pholidapus dybowskii* 1P	15/10/05
10	EU200472	*Pholidapus dybowskii* 3P	15/10/05
11	EU200474	*Pholidapus dybowskii* 4P	15/10/05
12	EU200476	*Pholidapus dybowskii* 5P	18/10/05
13	EU200500	*Pholidapus dybowskii* 48P	26/09/05
14	EU200502	*Pholidapus dybowskii* 51P	26/09/05
15	EU200477	*Opisthocentrus ocellatus* 7P	04/08/06
16	EU200481	*Opisthocentrus tenuis* 9P	16/07/05
17	EU200486	*Opisthocentrus tenuis* 14P	24/07/05
18	EU200478	*Tribolodon brandtii* 8	12/09/06
19	EU200480	*Tribolodon brandtii* 9	12/09/06
Victoria, Australia			
20	EF609348	*Emmelichthys nitidus*	1996
Clarence River, Brushgrove, New South Wales, Australia			
21	DQ107936	*Macquaria novemaculeata*	1998
22	DQ107937	*Macquaria novemaculeata*	1998

The PCR reaction mixture (25 µl) consisted of 2.5 µl Ex-Taq buffer (TaKaRa), 1 µl of F1-primer (20 pmol µl^{-1}), 1 µl of R1-primer (20 pmol µl^{-1}), 2 µl of dNTP mixture, 1 µl of template, 0.5 µl of *Taq* polymerase (5 units µl^{-1}) and 17.4 µl of distilled water. The PCR reactions were carried out in 30 to 40 cycles at 94°C for 5 min, 94°C for 30 s, 60°C for 30 s, 72°C for 60 s cycle^{-1} and 72°C for 7 min cycle^{-1}. The PCR products were electrophoresed in 1.0% agarose gel, stained with ethidium bromide and visualised under UV light. When good PCR products were obtained, subsequent specimens were then subjected to cycle-sequencing with commercial sequencing kits. Cycle-sequencing was made in most cases with DYEnamic ET Terminator Cycle Sequencing Kits (Amersham). If during sequencing a decreased signal was obtained for a considerable portion of the gene, the PCR and cycle sequencing were repeated with the same or a new primer set.

Labelled mtDNA fragments were analysed on a model 3110 or 3130 Sequencer (Applied Biosystems). Nineteen *Co-1* sequences of 9 percoid fish species and outgroup species were set to 532 bp after alignment. They originally varied in length from 598 to 671 bp, and were deposited in this mode in GenBank (Table 1) and also in the BOLD database (www.boldsystems.org).

35″ E, 42° 52′ 33″ N), and in Posiet Bay (130° 55′ 0″ E, 42° 30′ 0″ N), Peter the Great Bay, Sea of Japan, Russia (Table 1). The nomenclature of scientific names follows Froese & Pauly (2004). All species were identified by characters described by Lindberg & Krasyukova (1969, 1975), including the number of gill rakers, the number of fin rays, the number of scales in the lateral line, the shape of a lateral line and the colour and shape of the body. The locations of 3 specimens of Percichthyidae from literature sources along with their accession numbers are also given in Table 1 (specimens 20 to 22).

Total genomic DNA was isolated from the muscle tissue of the sampled fish using a phenol-chloroform technique (Lee 2000) or a modification, which mainly included phenol removal from the mixture and an RNAase addition at the final step. To get the *Co-1* sequence data of the 19 specimens, we first used PCR to obtain products of approximately 600 bp in length with the set of 4 primers following Ward et al. (2005): FishF1-50T CAA CCA ACC ACA AAG ACA TTG GCA C30; FishR1-50T AGA CTT CTG GGT GGC CAA AGA ATC A30; FishF2-50T CGA CTA ATC ATA AAG ATA TCG GCA C30; and FishR2-50A CTT CAG GGT GAC CGA AGA ATC AGA A30.

For phylogenetic purposes, 16 percoid *Co-1* gene sequences (sequences 1 to 6 and 8 to 17) and 2 outgroup sequences from cyprinids (sequences 18 to 19) from our own records were subjected to an analysis together with 3 other perciform sequences taken for comparative purposes and retrieved from GenBank (sequences 20 and 22, Table 1). Sequence 7 was excluded because of an unexpected great difference of this specimen of *Pholidapus dybowskii* from other specimens of this species. All analyses, starting with new DNA extraction, will be repeated for this specimen. The sequences were aligned with Clustal W software (Thompson et al. 1994), as an integrated product with MEGA-3.1 (Kumar et al. 1993); gap opening penalties and gap extension penalties were set to 15.0 versus 5.0 for both pairwise and multiple alignment sessions for the first step of alignment (default options were used for other settings). Following this initial step, large gaps were manually deleted, and final alignments were made with decreased levels of penalties (5.0 versus 0.5). All gaps were then deleted manually again.

During alignment, several sequences were removed from further analysis because of their low proximity to the bulk of the data, and the final sets of sequences for phylogenetic purposes were reduced to 18 and 21 (most sequences removed were those from GenBank that did not fit the length limit of 532 bp that we achieved for our own sequences as mentioned above). Both coding (all positions) and non-coding nucleotides were used in the substitution models. Four approaches were used for *Co-1* tree building: neighbour-joining (NJ), maximum likelihood (ML), Bayesian (BA) and maximum parsimony (MP) (Nei & Kumar 2000, Hall 2001, Felsenstein 2004). The HKY+G substitution model was chosen as the best for phylogenetic analysis; the test that included a data set of 19 sequences was based on Modeltest 3.7 software (Posada & Crandall 1998). The HKY+G model was estimated as the best by both the ML test and by the Akaike information criterion (AIC). The HKY+G model ($-\ln$ L = 2475.69, K = 5, AIC = 4961. 25; A:C:G:T = 0.2658, 0.2673, 0.1679, 0.2990, G = 0.1751, Ti/Tv [transition to transversion] ratio = 3.3079) was used for illustrative purposes in BA and ML tree building. Therefore, the BA (n = 10^6 generations) and ML (k = 1000 bootstrap replicates) approaches were based on the described model parameters. The NJ tree with bootstrapping (k = 1000 replicates) was based on the maximum composite likelihood (MCL) distance with the G parameter as defined above. Starting trees were obtained via a stepwise addition, and branch lengths at start-up were obtained using the Rogers-Swofford approximation method as implemented in PAUP* 4.0 10b (Swofford 2002). A tree bisection-reconnection technique was used for branch swapping under a heuristic search (with 10 repetitions). The main calculations were performed by PAUP* 4.0 10b (Swofford 2002), MrBayes 3.1.1 (Huelsenbeck & Ronquist 2001, Ronquist & Huelsenbeck 2003), and MEGA-3.1 (Kumar et al. 1993) packages. The phylogenetic trees were visualised and edited where necessary with TreeView software (Page 1996), and a general statistical analysis was performed using STATISTICA 5.5 (StatSoft).

RESULTS

The proportion of pyrimidines (T, C) to purines (A, G) in the *Co-1* gene deviated from a 50:50 ratio. There was no sharp difference in the nucleotide content between the representatives of the Perciformes and the 2 sequences of Cypriniformes. The frequency distributions for the 4 sequence categories (which included the 2 above mentioned taxa, Perciformes and Cypriniformes, and our data versus GenBank data) of 4 nucleotide types were significantly different, with slightly overlapping confidence intervals. To exemplify this statistical trend more precisely, a 2-factor MANOVA (Multivariate Analysis of Variance interaction mode, random effects) was done based on data for all nucleotide positions for 4 types of nucleotides (T, C, A and G; Factor 1), and for the 3 compared groups: (1) including our 17 sequences; (2) including the 3 GenBank sequences of the Perciformes; and (3) including 2 cyprinid sequences (Factor 2). The main effect of Factor 1 was significant: mean square (MS) effect = 337.54, MS error = 0.41; F = 829.10, df = 3; 76, p < 0.000001, whereas the main effect of Factor 2 was not significant (p < 0.998); however the interaction of Factors 1 and 2 was significant (p < 0.00001).

Co-1 sequences were investigated to understand the phylogenetic relationships among percoid fishes, with cyprinids as an outgroup. The *p*-distances and their derivates were obtained from 22 *Co-1* gene sequences of 532 bp for 20 percoids: 17 from our collection and 3 GenBank sequences as well as 2 sequences of cyprinids. However, tree building and molecular phylogenetic comparisons were performed mainly on 18 sequences (see data below). Trees using 21 sequences were also constructed and are included in the Discussion.

Simple *p*-distance values are applicable for a comparative analysis of genetic diversity within species and among taxa, although they represent only an approximate view of sequence divergence. Here, the sequence diversities for the *Co-1* genes from the investigated taxa were tested by several approaches, which showed basically similar patterns. Even a simple visual inspection of the *p*-distances showed that the lowest diversity occurred between conspecific sequences and the highest diversity occurred between families. Although the variation row for pairwise *p*-distances was unimodal, the distribution deviated from normality and had a heavy tail (skewness = –1.1799, SE = 0.1763). Taking this information into account, we first compared the mean *p*-distances for different groups of perch-like fishes: (1) intraspecies, among individuals of the same species; (2) intragenus, among species of the same genus; (3) intrafamily, among genera of the same family; and (4) intraorder, among families of the same order (Fig. 1). To perform this analysis, we took all pairwise scores from a triangle *p*-distance matrix for the 20 sequences belonging to percoid fishes and sorted them into the corresponding 4 groups. Afterwards, several averages were obtained for each category (within species, genus, family, order). The variation row for mean *p*-distances agreed well with the normal distribution (skewness = 0.0716, SE = 0.6163; kurtosis = –2.1263, SE = 1.1909). A 1-factor ANOVA (random effect mode) on the 4 above-defined groups was used to identify variability in *p*-distances among
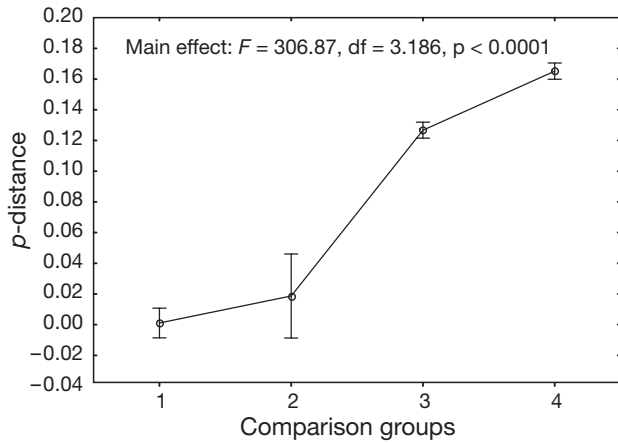
Fig. 1. Results of a 1-factor ANOVA and mean *p*-distance values at 4 levels of differentiation in fish species (Perciformes) for *Co-1* gene sequence data. All pairwise *p*-distances were obtained from a triangle matrix. Comparison groups: (1) intraspecies, among individuals of the same species; (2) intragenus, among species of the same genus; (3) intrafamily, among genera of the same family; (4) intraorder, among families of the order Perciformes. Variation is statistically significant. Error bars represent 95% confidence intervals for the mean

groups. Combined with the calculations of the average *p*-distance scores for these 4 groups, the ANOVA revealed a significant difference among averages of the 4 groups (F = 306.87, df = 3; 186, p < 0.0001). Because of the importance of this conclusion, we performed a nonparametric Kruskal-Wallis ANOVA for *p*-distance differences in the 4 compared groups. This test by ranks gave: H = 115.20, df = 3, N = 190, p < 0.0001 and a median test gave: X^2 = 89.81, df = 3, p < 0.0001. Thus, 2 different approaches revealed that the average *p*-distances among intraspecies, intragenus, intrafamily, and intraorder categories among Perciformes were different.

For phylogenetic purposes, NJ, ML, BA, and MP trees were constructed. The HKY+G substitution model was used for ML and BA, while MCL was used for building the NJ tree. When 18 sequences were considered, all trees showed a very similar topology, and all 4 methods gave trees that differed only at the levels of support for internal nodes and in joins between the Stichaeidae and the rest of the Perciformes branches, with an external position for cyprinids (Fig. 2). The NJ, ML and BA trees are shown here (Fig. 2). The BA-tree program parameters were as follows: Markov Chain Monte-Carlo Parameters (MCMCP) ngen = 1 000 000, printfreq = 1000, samplefreq = 1000, nchains = 4, SUMP burnin = 100; default program options were used for the remaining parameters. The consensus tree was calculated on the basis of the 11 672 best trees, and 1381 of them fitted the 50% majority rule. The MP tree differed from the NJ, ML and BA trees mostly

in somewhat lower resolution. Nevertheless, all trees showed 4 major properties: (1) a separate cluster for outgroup species, (2) a cluster and 3 branches, each comprising a single family of the order Perciformes, (3) the presence of clear-cut clusters representing the family Stichaeidae and its different genera, and (4) the existence of a set of the nearest clusters for all single individuals that are classified as specimens of the same species (Fig. 2). Repetition frequencies (posterior probabilities for the BA tree and bootstrap support levels for the others) for the 4 kinds of clusters were respectively 100% (1.0), 81–100% (0.81–1.0), 99–100% (0.99–1.0), and 100% (1.0) in the BA tree, while in the NJ and ML trees they were 100%, 71–84%, 60–99%, and 97–100%.
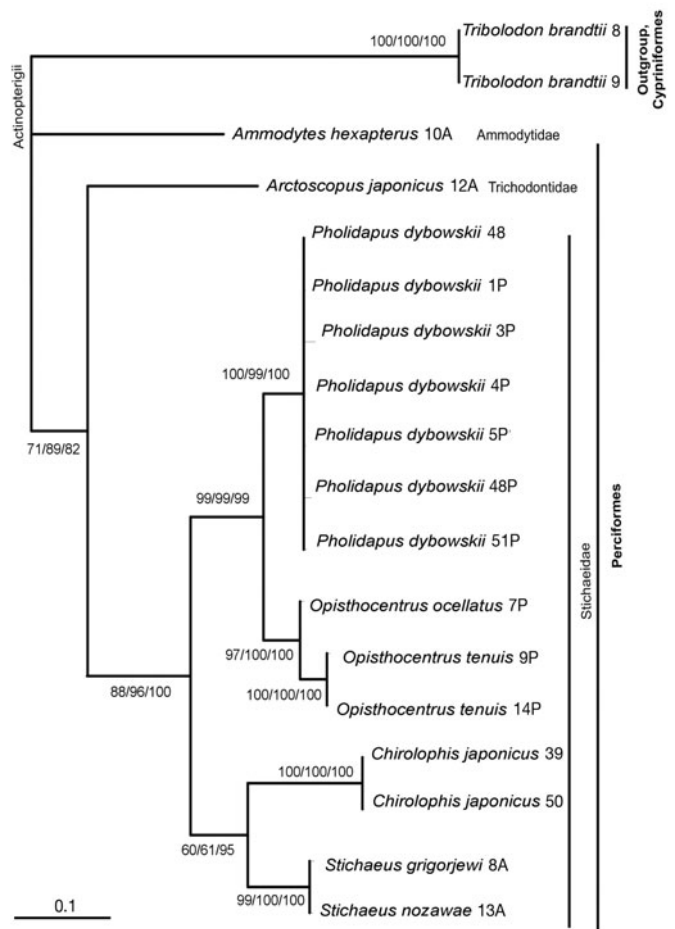

Fig. 2. Consensus (50%) trees showing phylogenetic interrelationships on the basis of *Co-1* sequence data for 7 analysed perch-like fish species (Perciformes) and 2 outgroup sequences (Cypriniformes). The rooted Bayesian tree was built for sample purposes. Bootstrap support values and posterior probabilities are shown in the nodes of the tree; support scores are given in the order neighbour-joining/maximum likelihood/Bayesian. Scale indicates the relative branch lengths

DISCUSSION

Bias in the ratio of (T+C):(A+G) has been well described in the literature for other protein-coding genes, e.g. *Cyt-b* (e.g. Kim et al. 2004), but is usually stated without statistical substantiation. We have validated this for *Co-1* in percoid fishes on a precise statistical basis (see 'Results'). The interpretation of nucleotide bias is that it reflects the hydrophobic properties of the encoded proteins (Naylor et al. 1995). However, taxonomic differences, when found, are more relevant to the evolution of lineages and might reflect their separate divergence (Kartavtsev et al. 2007a).

Scores of *p*-distances pooled for different scales of evolutionary history gave a clear pattern, presenting an increase in nucleotide diversity at (1) intraspecies, (2) intragenus, (3) intrafamily and (4) intraorder levels (Fig. 1). Scores of average *p*-distances for compared fish groups were: (1) $0.11 \pm 0.04\%$, (2) $1.87 \pm 0.68\%$, (3) $12.67 \pm 0.28\%$ and (4) $16.52 \pm 0.10\%$, respectively (mean ± SE). It is generally accepted that taxa of higher ranks have existed longer than those of lower ranks (Simpson 1961, Nei 1987). Thus, taking into consideration significant differences between intraspecies and interspecies levels within genera, we may attribute these data to a certain mode of speciation. More specifically, together with an increase in *p*-distances at 2 stages (subspecies and species), these data lead to the conclusion that speciation in the compared representatives of the order Perciformes followed a geographic mode (sensu Mayr 1982), with the accumulation of numerous per site changes (mutations) over a long period of time since separation of gene pools of local populations and/or subspecies. Finally, this mode forecasts the species formation and later independent evolution of species themselves and consequent divergence of representatives in higher lineages or phyletic evolution (Dobzhansky 1955, Bush 1975, Templeton 1981, Mayr 1982, King 1993, Harrison 1998). Recently, the geographic mode of speciation has been substantiated for many other animals and for a wider list of rank categories including subspecies, semispecies, sibling species, morphologically distinct species and genera (Johns & Avise 1998, Avise 2000, Kartavtsev 2005, Ward et al. 2005, Kartavtsev & Lee 2006, Kartavtsev et al. 2007a,b, 2008, Ward & Holmes 2007). In these, more than 24 000 species are listed for different comparison groups or taxa. Taking into account the complexity of a species as an entity, we are aware of the fact that such a conclusion derived solely from the genetic distance scale is not quite comprehensive. Actually, both a wider set of descriptors (Kartavtsev et al. 2002, Kartavtsev 2005, Kartavtsev & Lee 2006) and a focused investigation of this specific point in the Stichaeidae

and other Perciformes are required to draw a more precise conclusion on the speciation mode.

For the sake of brevity, we will consider in detail the correspondence between the trees built in this research and in literature data mostly for the Stichaeidae. The investigated representatives of the family Stichaeidae joined as a monophyletic branch with 86, 96, and 100% of support in the NJ, ML, and BA trees, respectively (Fig. 2) and 74% bootstrap support in the MP tree (not shown). Unlike the Stichaeidae with 74 to 100% support, the resolution of the other family branches was lower in all approaches (Fig. 2, see Trichodontidae), and the Ammoditidae branch was not even resolved. The same conclusion remains basically true when a tree was built on 21 sequences (not shown). Comparing those smaller and larger trees, we conclude that the reliability of monophyly even for the Stichaeidae is diminished and the resolution of the larger tree decreased. One of the most probable explanations for this is that the information capacity in our 532 bp *Co-1* sequences is not enough to cover the increase in the diversity of taxa.

A close genetic proximity of the family Trichodontidae to the family Stichaeidae and monophyly of the latter was supported by 3 kinds of trees (NJ, ML, BA). Whether families are monophyletic is an important question for molecular phylogenetics, especially for big families, such as the Sparidae, Carangidae and Stichaeidae, the basic information for which is currently zoological (Froese & Pauly 2009). However, some attention to this point has been paid by molecular phylogeneticists (Stepien et al. 1997, Reed et al. 2002, Orrell & Carpenter 2004, Craig & Hastings 2007, Radchenko et al. 2009). The monophyly for at least 3 tribe clades has been suggested within the Carangidae based on the results for *Cyt-b* sequence data (Reed et al. 2002) and for 16S rRNA and *Cyt-b* among several subfamilies of the family Sparidae (Orrell & Carpenter 2004). However, evidence based on mtDNA and nDNA sequences (4 genes) brings into question the monophyly of the large family Serranidae, as well as the genera *Cephalopholis, Epinephelus* and *Mycteroperca* as defined by Craig & Hastings (2007). Our own data show that the investigated representatives of the family Stichaeidae formed a monophyletic clade with high support (Fig. 2). This fact disagrees with the polyphyly found for Stichaeidae on 3 mtDNA genes (*Cyt-b, Co-1* and 16S rRNA with a combined sequence length of 2043 to 2046 bp; Radchenko et al. 2009) among a wider set of taxa (Zoarcoidei). To our knowledge, the monophyletic origin of most families in the order Perciformes has not yet been investigated thoroughly; there is a need for further examination of this question (Johnson & Paterson 1993, Smith & Craig 2007). Both the present and previous results corresponded fairly

well to the recent morphological classification and conclusions about the status of the family Stichaeidae in particular (Makushok 1986a, Nelson 2006). Our own conclusion on monophyly is mostly valid for the Stichaeidae (Fig. 2), which was represented by 4 genera and several species. Regarding the discrepancies found, the real tree topology may remain obscure, despite statistical support, if the analysed taxa are not representative of natural diversity (Hillis 1998) and/or if markers used have insufficient information capacity as discussed in the 'Introduction'. This is certainly the case with the level of the order due to several unresolved branches in our *Co-1*-based trees (Fig. 2). However, even for the family level, the above-mentioned problems of not representing enough species diversity and insufficient molecular signal may be a real obstacle for a correct phylogenetic reconstruction when a family is speciose and the analysed sequence length is relatively short (1000 to 2000 bp or less), as it is with the Stichaeidae. The limited molecular genetic information we have in general for fish (Saitoh et al. 2000, Reed et al. 2002, Miya et al. 2003, Orrell & Carpenter 2004, Smith & Craig 2007) is also a significant obstacle.

The validity of all considered genera in the Stichaeidae (*Pholidapus*, *Opisthocentrus*, *Chrisophis* and *Stichaeus*) is well supported by our data (Fig. 2). However, more abundant species representation is highly desirable. These data also agreed with the recent systematic revision that moved *Pholidapus dybowskii* out of the genus *Opisthocentrus* (Shiogaki 1984); nevertheless, the molecular phylogenetic proximity of these 2 genera is very high, and they comprise a monophyletic clade in all trees (Fig. 2). The current taxonomic separation of *Stichaeus grigorjewi* (Herzenstein, 1890) and *S. nozawae* (Jordan & Snyder 1902) seems artificial in view of their close genetic proximity, comparable to intraspecies diversity (Fig. 2). Certainly, both questions require repeated examination using other markers, additional specimens and species of the 2 genera discussed.

One of the most important properties of the constructed trees (see points 3 and 4 at the end of the 'Results' section) might be the sharp difference among different genera and among congeneric species in the major family considered (Fig. 2), except for the 2 species in the genus *Stichaeus* (as stressed in the above paragraph). One of the most reliable and interesting points of our research is the difference between intraspecies and interspecies (congeneric) levels of genetic diversity. This is basically an indicator of the barcoding value of a gene sequence for species discrimination and may be used as a tool for comparative evolutionary genetics. These kinds of data are very reliable as they are supported both by the tree topology seen as the nearest intraspecies clusters for all

cases (Fig. 2) and by very small average intraspecies *p*-distance scores (Fig. 1), both indicating that some individuals of the same species have nearly identical sequences. In contrast, interspecies congeneric distances are quite distinct within major genera in our data set (Fig. 2). Thus, our findings support a barcoding value of *Co-1* gene marker for most current taxa, as also shown in many previous observations (Hebert et al. 2004, Ward et al. 2005, Kartavtsev & Lee 2006, Kartavtsev et al. 2007a,b, 2008, Rock et al. 2007, Ward & Holmes 2007, Hubert et al. 2008).

Despite the representation of distant taxa in the present study, we did not face a long-branch attraction problem as has been described in some other diverse groups (Wilcox et al. 2004). The outgroup taxon, Cypriniformes, belonging to ray-finned fishes of the class Actinopterygii, formed phylogenetic clusters related to the Perciformes (Fig. 2). However, this high taxonomic level will likely be resolved in the future by studies involving more powerful approaches, such as a comparative analysis of complete mtDNA sequences (e.g. Miya et al. 2003, Nohara et al. 2005) and nuclear genes.

LITERATURE CITED

Avise JC (2001) Phylogeography. The history and formation of species. Harvard University Press, Cambridge, MA

Ayala FJ, Tracey ML, Hedgecock D, Richmond RC (1974) Genetic differentiation during the speciation process in *Drosophila*. Evolution 28:576–592

Bush GL (1975) Modes of animal speciation. Annu Rev Ecol Syst 6:339–364

Craig MT, Hastings PA (2007) A molecular phylogeny of the groupers of the subfamily Epinephelinae (Serranidae) with a revised classification of the Epinephelini. Ichthyol Res 54:1–17

Dobzhansky T (1955) Evolution, genetics and man. John Wiley & Sons, New York

Felsenstein J (2004) Inferring phylogenies. Sinauer Associates, Sunderland, MA

Finnerty JR, Block BA (1995) Evolution of cytochrome-*b* in the Scombroidei (Teleostei)—molecular insights into Billfish (Istiophoridae and Xiphiidae) relationships. Fish Bull 93(1): 78–96

Froese R, Pauly D (eds) (2004) FishBase. www.fishbase.org

Froese R, Pauly D (eds) (2009) World Wide Web electronic publication. www.fishbase.org, version (10/2009)

Hall B (2001) Phylogenetic trees made easy. A how-to manual for molecular biologists. Sinauer Associates, Sunderland, MA

Harrison RG (1998) Linking evolutionary pattern and process: the relevance of species concepts for the study of speciation. In: Howard DJ, Berlocher SH (eds) Endless forms: species and speciation. Oxford University Press, Oxford, p 19–31

Hebert PD, Stoeckle MY, Zemlak TS, Francis CM (2004) Identification of birds through DNA barcodes. PLoS Biol 2:e312 doi: 10.1371/journal.pbio.0020312

Hillis DM (1998) Taxonomic sampling, phylogenetic accuracy, and investigator bias. Syst Biol 47:3–8

Hillis DM, Moritz C, Mable B (eds) (1996) Molecular systematics. Sinauer Associates, Sunderland, MA

Hubert N, Hanner R, Holm E, Mandrak NE and others (2008) Identifying Canadian freshwater fishes through DNA barcodes. PLoS ONE 3:e2490. doi:10.1371/journal.pone. 0002490

Huelsenbeck JP, Ronquist F (2001) Mr. BAYES: Bayesian inference of phylogeny. Bioinformatics 17:754–755

Johns GC, Avise JC (1998) A comparative summary of genetic distances in the vertebrates from the mitochondrial cytochrome b gene. Mol Biol Evol 15:1481–1490

Johnson GD, Paterson C (1993) Percomorph phylogeny: a survey of acanthomorphs and new proposal. Bull Mar Sci 52:554–626

Kartavtsev YP (2005) Molecular evolution and population genetics. Far Eastern State University Press, Vladivostok

Kartavtsev YP, Lee JS (2006) Analysis of nucleotide diversity at genes *Cyt-b* and *Co-1* on population, species, and genera levels. Applicability of DNA and allozyme data in the genetics of speciation. Genetika 42:437–461

Kartavtsev YP, Sviridov VV, Sasaki T, Hanzawa N (2002) Genetic divergence of far eastern dace belonging to the genus *Tribolodon* (Pisces, Cyprinidae) and closely related taxa: some insights in taxonomy and speciation. Genetika 38:1518–1531 (In Russian)

Kartavtsev YP, Park TJ, Vinnikov KA, Ivankov VN, Sharina SN, Lee JS (2007a) Cytochrome *b* (*Cyt-b*) gene sequences analysis in six flatfish species (Pisces, Pleuronectidae) with phylogenetic and taxonomic insights. Mar Biol 152: 757–773

Kartavtsev YP, Jung SO, Lee YM, Byeon HK, Lee JS (2007b) Complete mitochondrial genome in the bullhead torrent catfish, *Liobagrus obesus* (Siluriformes, Amblycipididae): genome description and phylogenetic considerations inferred from the Cyt *b* and 16S rRNA genes. Gene 396: 13–27

Kartavtsev YP, Park TJ, Lee JS, Vinnikov KA, Ivankov VN, Sharina SN, Ponomarev AS (2008) Phylogenetic inferences introduced on cytochrome *b* (*Cyt-b*) gene sequences data for six flatfish species (Teleostei, Pleuronectidae) and species synonymy between representatives of genera *Pseudopleuronectes* and *Hippoglossoides* from far eastern seas. Russ J Genet 44:451–458

Kim IC, Kweon HS, Kim YJ, Kim CB and others (2004) The complete mitochondrial genome of the javeline goby *Acanthogobius hasta* (Perciformes, Gobiidae) and phylogenetic considerations. Gene 336:147–153

King M (1993) Species evolution: the role of chromosome change. Cambridge University Press, Cambridge

Kumar S, Tamura K, Nei M (1993) Manual for MEGA: molecular evolutionary genetics analysis software. Pennsylvania State University, University Park, PA

Lee JS (2000) The internally self-fertilized hermaphroditic teleost *Rivulus marmoratus* (Cyprinodontiformes, Rivulidae)

β-actin gene: amplification and sequence analysis with conserved primers. Mar Biotechnol 2:161–166

Lindberg GU, Krasyukova ZV (1969) Fishes of the Sea of Japan and the adjacent areas of the Sea of Okhotsk and Yellow Sea, Part 3. Nauka, Leningrad

Lindberg GU, Krasyukova ZV (1975) Fishes of the Sea of Japan and the adjacent areas of the Sea of Okhotsk and Yellow Sea, Part 4. Nauka, Leningrad

Lydeard C, Roe KJ (1997) The phylogenetic utility of the mitochondrial cytochrome *b* gene for inferring relationships among Actinopterygian fishes. In: Kocher TD, Stepien CA (eds) Molecular systematics of fishes. Academic Press, San Diego, CA, p 285–303

Makushok VM (1986a) Stichaeidae. In: Whitehead PJP, Bauchot ML, Hureau JC, Nielsen J, Tortonese E (eds) Fishes of the North-eastern Atlantic and the Mediterranean, Vol. 3. UNESCO, Paris, p 1122–1123

Makushok VM (1986b) Lumpenidae (subfamily of Stichaeidae). In: Whitehead PJP, Bauchot ML, Hureau JC, Nielsen J, Tortonese E (eds) Fishes of the North-eastern Atlantic and the Mediterranean, Vol. 3. UNESCO, Paris, p 1126–1129

Mayr E (1982) Process of speciation in animals. In: Barigozzi C (ed) Mechanisms of speciation. Alan R. Liss, New York, p 1–20

Miya M, Kawaguchi A, Nishida M (2001) Mitogenomic exploration of higher teleostean phylogenies: a case study for moderate-scale evolutionary genomics with 38 newly determined complete mitochondrial DNA sequences. Mol Biol Evol 18:1993–2009

Miya M, Takeshima H, Endo H, Ishiguro NB and others (2003) Major patterns of higher teleostean phylogenies: a new perspective based on 100 complete mitochondrial DNA sequences. Mol Phylogenet Evol 26:121–138

Naylor GJ, Collins TM, Brown WM (1995) Hydrophobicity and phylogeny. Nature 373:565–566

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Nei M, Kumar S (2000) Molecular evolution and phylogenetics. Oxford University Press, New York

Nelson JS (2006) Fishes of the world, 4th edn. John Wiley and Sons, New York

Nohara M, Nishida M, Miya M, Nishikawa T (2005) Evolution of the mitochondrial genome in Cephalochordata as inferred from complete nucleotide sequences from two *Epigonichthys* species. J Mol Evol 60:526–537

Orrell TM, Carpenter KE (2004) A phylogeny of the fish family Sparidae (porgies) inferred from mitochondrial sequence data. Mol Phylogenet Evol 32:425–434

Page RDM (1996) TREEVIEW: an application to display phylogenetic trees on personal computers. Comput Appl Biosci 12:357–358

Posada D, Crandall KA (1998) MODELTEST: testing the model DNA substitution. Bioinformatics 14:817–818

Radchenko OA, Chereshnev IA, Petrovskaya AV, Balanov AA (2009) Molecular systematics and phylogeny of the suborder Zoarcoidei. Vestn DVO 3:40–47 (In Russian)

Reed DL, Carpenter KE, deGravelle MJ (2002) Molecular systematics of the Jacks (Perciformes : Carangidae) based on mitochondrial cytochrome *b* sequences using parsimony, likelihood, and Bayesian approaches. Mol Phylogenet Evol 23(3):513–524

Rock J, Costa FO, Walker DI, North AW, Hutchinson WF, Carvalho GR (2007) DNA barcodes of fish of the Scotia Sea, Antarctica, indicate priority groups for taxonomic and systematics focus. Antarct Sci 20:253–262

Ronquist F, Huelsenbeck JP (2003) Mr.BAYES 3: Bayesian

phylogenetic inference under mixed models. Bioinformatics 19:1572–1574

Saitoh K, Hayashizaki K, Yokoyama Y, Asahida T, Toyohara H, Yamashita Y (2000) Complete nucleotide sequence of Japanese flounder (*Paralichthys olivaceus*) mitochondrial genome: structural properties and cues for resolving teleostean relationships. J Hered 91:271–278

Shiogaki M (1984) A review of the genera *Pholidapus* and *Opisthocentrus* (Stichaeidae). Jpn J Ichthyol 31:213–224

Simpson GG (1961) Principles of animal taxonomy. The species and lower categories. Columbia University Press, New York

Smith WL, Craig MT (2007) Casting the percomorph net widely: the importance of broad taxonomic sampling in the search for the placement of serranid and percid fishes. Copeia 2007:35–55

Song CB, Near TJ, Page LM (1998) Phylogenetic relations among percid fishes as inferred from mitochondrial cytochrome *b* DNA sequence data. Mol Phylogenet Evol 10:343–353

Stepien CA, Dillon AK, Brooks MJ, Chase K, Hubers A (1997) The evolution of blennoid fishes based on an analysis of mitochondrial 12S rDNA. In: Cocher TD, Stepien CA (eds)

Molecular systematics of fishes. Academic Press, San Diego, CA, p 245–270

Swofford DL (2002) PAUP*: phylogenetic analysis using parsimony, version 4.10. Sinauer Associates, Sunderland, MA

Templeton AR (1981) Mechanisms of speciation—population genetic approach. Annu Rev Ecol Syst 12:23–48

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

Wallace DC (1992) Diseases of the mitochondrial DNA. Annu Rev Biochem 61:1175–1212

Ward RD, Holmes BH (2007) An analysis of nucleotide and amino acid variability in the barcode region of cytochrome c oxidase I (cox1) in fishes. Mol Ecol Notes 7:899–907

Ward RD, Zemlak TS, Innes BH, Last PA, Hebert PDN (2005) DNA barcoding Australia's fish species. Philos Trans R Soc Lond B Biol Sci 360:1847–1857

Wilcox TP, Garcia de Leon FJ, Hendrickson DA, Hillis DM (2004) Convergence among cave catfishes: long-branch attraction and a Bayesian relative rates test. Mol Phylogenet Evol 31:1101–1113