

Speciation of two gobioid species, *Pterogobius elapoides* and *Pterogobius zonoleucus* revealed by multi-locus nuclear and mitochondrial DNA analyses☆



Akihito^a, Fumihito Akishinonomiya^{b,c}, Yuji Ikeda^d, Masahiro Aizawa^d, So Nakagawa^{e,i}, Yumi Umehara^e, Takahiro Yonezawa^{f,g}, Shuhei Mano^g, Masami Hasegawa^{f,g}, Tetsuji Nakabo^h, Takashi Gojobori^{e,j,*}

^a The Imperial Residence, 1-1 Chiyoda, Chiyoda-ku, Tokyo 100-0001, Japan

^b The University Museum, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

^c Tokyo University of Agriculture, 1737 Funako, Atsugi-shi, Kanagawa 243-0034, Japan

^d Imperial Household Agency, 1-1 Chiyoda, Chiyoda-ku, Tokyo 100-8111, Japan

^e Center for Information Biology, National Institute of Genetics, 1111 Yata, Mishima, Shizuoka 411-8540, Japan

^f School of Life Sciences, Fudan University, SongHu Rd. 2005, Shanghai 200438, China

^g The Institute of Statistical Mathematics, 10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan

^h The Kyoto University Museum, Kyoto University, Yoshida Honmachi, Sakyo-ku, Kyoto 606-8501, Japan

ⁱ Department of Molecular Life Science, Tokai University School of Medicine, 143 Shimokasuya, Isehara, Kanagawa 259-1193, Japan

^j Computational Bioscience Research Center, Biological and Environmental Science and Engineering, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia

ARTICLE INFO

Available online 28 October 2015

ABSTRACT

To understand how geographical differentiation of gobioid fish species led to speciation, two populations of the Pacific Ocean and the Sea of Japan for each of the two gobioid species, *Pterogobius elapoides* and *Pterogobius zonoleucus*, were studied in both morphological and molecular features. Analyzing mitochondrial genes, Akihito et al. (2008) suggested that *P. zonoleucus* does not form a monophyletic clade relative to *P. elapoides*, indicating that “Sea of Japan *P. zonoleucus*” and *P. elapoides* form a clade excluding “Pacific *P. zonoleucus*” as an outgroup. Because morphological classification clearly distinguish these two species and a gene tree may differ from a population tree, we examined three nuclear genes, S7RP, RAG1, and TBR1, in this work, in order to determine whether nuclear and mitochondrial trees are concordant, thus shedding light on the evolutionary history of this group of fishes. Importantly, nuclear trees were based on exactly the same individuals that were used for the previously published mtDNA trees. The tree based on RAG1 exon sequences suggested a closer relationship of *P. elapoides* with “Sea of Japan *P. zonoleucus*”, which was in agreement with the mitochondrial tree. In contrast, S7RP and TBR1 introns recovered a monophyletic *P. zonoleucus*. If the mitochondrial tree represents the population tree in which *P. elapoides* evolved from “Sea of Japan *P. zonoleucus*”, the population size of *P. elapoides* is expected to be smaller than that of “Sea of Japan *P. zonoleucus*”. This is because a smaller population of the new species is usually differentiated from a larger population of the ancestral species when the speciation occurred. However, we found no evidence of such a small population size during the evolution of *P. elapoides*. Therefore, we conclude that the monophyletic *P. zonoleucus* as suggested by S7RP and TBR1 most likely represents the population tree, which is consistent with the morphological classification. In this case, it is possible that the incongruent mitochondrial and RAG1 trees are either due to incomplete lineage sorting of ancestral polymorphisms or to introgression by hybridization. Because of a smaller effective population size of mitochondria compared with nuclear genes, the introgression might be a more likely scenario in explaining the incongruent mitochondrial tree than the incomplete lineage sorting. Because of smaller effective population size of “Sea of Japan *P. zonoleucus*” than that of *P. elapoides*, the direction of the introgression was likely to be from the latter to the former. This evolutionary work of the two gobioid species highlights the need of analyzing multiple gene trees for both nuclear and mitochondrial genes as well as scrutinization of morphological characteristics to obtain a population tree representing the organismal evolutionary history.

© 2015 Elsevier B.V. All rights reserved.

☆ The DDBJ/EMBL/GenBank accession numbers are from LC020659 to LC020784.

* Corresponding author at: Computational Bioscience Research Center, Biological and Environmental Science and Engineering, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia

E-mail address: tgojobor@nig.ac.jp (T. Gojobori).

1. Introduction

Forming a large suborder in the order of Perciformes, the Gobioidae consists of nine families and approximately 2211 species (Nelson, 2006). Such a vast number of species have apparently undergone their morphological specialization to diverse habitats including marine, estuarine and freshwater, making the classification of the gobioid fishes quite complicated. Moreover, gobioid fishes are considered suitable materials for studying an evolutionary process, in particular a speciation process, because various diversifications seem to have taken place from the common ancestors over the world, although the evolutionary origin was not well-known yet.

Akihito et al. (2000) determined and examined nucleotide sequences of mitochondrial cytochrome *b* (CytB) genes from 28 species of gobioid fishes including 8 families (Rhyacichthyidae, Odontobutidae, Eleotridae, Xenisthmidae, Gobiidae, Kraemeriidae, Microdesmidae and Schindleriidae). The authors revealed that the phylogenetic relationships followed differentiation of gobioid fishes which can be separated into many groups of morphological and ecological diversity. In particular, phylogenetic trees obtained show that most species examined have diverged from each other almost simultaneously or during an extremely short period of time, following that phylogenetic classifications of gobioid fishes are a kind of difficult tasks. It is also worth noting that this is one of the earliest studies using DNA sequences to examine the relationships of gobioid fishes (Pezold, 2011).

Akihito et al. (2008) focused on two populations of the Pacific Ocean and the Sea of Japan for each of the two gobioid species, *Pterogobius elapoides* and *Pterogobius zonoleucus*, from the viewpoints of both morphological and molecular features. Both *Pterogobius elapoides* and *Pterogobius zonoleucus* are common free-swimming gobies that can be found in rocky and weedy shores along the temperate coast of Japan. Those two species are good examples for studying a speciation process because geographical variations can be informative at the morphological level.

Collecting samples of both species from 23 different locations around the coast of Japan, Akihito et al. (2008) compared the mitochondrial nucleotide sequences of the two gene regions, CytB and NADH dehydrogenase subunit 2 (ND2). According to the phylogenetic trees they constructed, it indicated that all 125 samples of the two species can be clearly divided into the four clades: “Pacific *P. elapoides*”, “Sea of Japan *P. elapoides*”, “Pacific *P. zonoleucus*”, and “Sea of Japan *P. zonoleucus*”. Interestingly, although “Pacific *P. elapoides*” and “Sea of Japan *P. elapoides*” composed a single clade, this single clade formed a monophyletic group together with “Sea of Japan *P. zonoleucus*” with a bootstrap value of more than 95%. In other words, Akihito et al. have made an observation that at the mitochondrial DNA (mtDNA) level, the two populations of Pacific Ocean and the Sea of Japan of *P. elapoides* are considered to have originated from a population of the Sea of Japan of *P. zonoleucus*.

The above observation led the authors to consider the possibility that the Pacific Ocean and Sea of Japan populations of *P. zonoleucus* represent two separate species, although they were morphologically indistinguishable for most traits. However, the results obtained by Akihito et al. (2008) should be taken with caution, because it is the outcome obtained only from the mtDNA sequence data. MtDNA sequences are known to be informative makers for closely related species mainly because of relatively rapid rates of substitution (Brown et al., 1979) and the very low rate of recombination (Piganeau et al., 2004). However, it is also known that gene and species phylogenetic trees are not always congruent due to ancestral polymorphisms and/or introgression in particular when the time of divergence between different species is short (Nei, 1987). Indeed, in recent years, there are many cases where the phylogenetic relationships are different between nuclear DNAs (nucDNAs) and mtDNAs (reviewed in Rubinoff and Holland, 2005; Dupuis et al., 2012). Therefore, this has urged us to examine the nucleotide sequences of multiple unlinked nuclear loci for those species.

In the present study, we examined the phylogenetic relationships among the populations of Pacific Ocean and the Sea of Japan for both *P. elapoides* and *P. zonoleucus*, sequencing three nuclear gene regions, an exon of RAG1 and introns of S7RP and TBR1. These gene regions were chosen because they do not appear to be linked as well as influenced by any particular strong force of positive selection. Here in the present study of the nucDNA analysis, we have to note that we took advantage of using exactly the same individual samples that we used for mtDNAs (Akihito et al., 2008). This reinforces the validity of our comparative studies of phylogenetic relationships of the two species between mitochondrial and nuclear gene regions.

2. Materials and methods

2.1. Samples

All 25 fish samples used in this study including 14 individuals of *P. elapoides*, nine individuals of *P. zonoleucus*, and each of *P. zacalles* and *P. virgo* individuals were prepared as previously described (see Akihito et al., 2008). Abbreviations and the collection locations of all samples examined are summarized in Table 1.

2.2. DNA extraction, PCR amplification, sequencing, and cloning

We extracted DNA from each sample following the protocol described in Akihito et al. (2008). We designed PCR primers specific to the second intron of S7 ribosomal protein (S7RP), the third exon of recombination activating gene 1 (RAG1), and the fifth intron of T-box brain 1 (TBR1). Note that these were found to be unique (and not members of multi-gene families) in the stickleback (*Gasterosteus aculeatus*) genome, thus assuring the use of orthologous sequence. Each primer sequence was shown in Supplementary Table S1. The PCR products were purified using MinElute® Gel Extraction Kit (Qiagen), MinElute® PCR Purification Kit (Qiagen), and/or ExoSAP-IT (GE Healthcare). The purified PCR products were sequenced by an ABI 3730 DNA Analyzer (Applied Biosystems). When two or more polymorphisms were found in a nucleotide sequence of the obtained PCR product (of which species are shown in Table 1), we conducted cloning experiments to identify haplotype structures for all species examined as follows: an adenine was added to the 3' end of a PCR product using 10× A attachment mix (Toyobo), and it was then cloned into a pMD20-T vector using DNA Ligation Kit, Mighty Mix (TaKaRa Bio). The plasmid DNA containing the cloned the PCR product was transformed into *Escherichia coli* strain DH5α. Transformants were incubated for 1 h at 37 °C, and then plated onto LB plates with ampicillin, X-gal, and IPTG. The plasmid DNAs of white recombinant bacteria were purified using Montage Plasmid Miniprep Kit (Millipore), which were used as templates and directly cycle-sequenced using M13 forward and M13 reverse primers and an ABI Prism Dye Terminator Reaction Kit. The products were sequenced by an ABI 3730 DNA Analyzer. All sequence data obtained will be deposited in the DDBJ/EMBL/GenBank International nucleotide sequence database under accession numbers (Table 1).

2.3. Phylogenetic analysis

We made multiple alignments of the obtained nucleotide sequences for each gene, using a computer program L-INS-i in MAFFT (Katoh et al., 2005). Using each alignment sequence, we inferred the phylogenetic relationship utilizing neighbor-joining (NJ, Saitou and Nei, 1987) and maximum likelihood (ML, Felsenstein, 1981) methods as follows.

For NJ trees, the MEGA program version 5.05 (Tamura et al., 2011) was used. The evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura et al., 2004) with the rate variation among sites modeled with a gamma distribution. For the distances of RAG1 exon sequences, the codon positions included

Table 1

Accession numbers of nucDNA sequences used in this study.

Sample ^a and abbreviation	Locality	RAG1 ^b	S7RP ^b	TBR1 ^b
<i>Pterogobius elapoides</i>				
PeINagasaki	Sea of Japan, Sasebo, Nagasaki Pref.	LC020659/LC020660	LC020744/LC020745	LC020705/LC020706
PeShimaneHamada	Sea of Japan, Hamada, Shimane Pref.	LC020661	LC020746/LC020747	LC020707/LC020708
PeOki	Sea of Japan, Oki, Shimane Pref.	LC020662/LC020663	LC020748/LC020749	LC020709/LC020710
PeIshikawa	Sea of Japan, Tsukumo Bay, Ishikawa Pref.	LC020664/LC020665	LC020750/LC020751	LC020711/LC020712
PeSado	Sea of Japan, Sado, Niigata Pref.	LC020666/LC020667	LC020752	LC020713
PeAkita	Sea of Japan, Oga, Akita Pref.	LC020668/LC020669	LC020753/LC020754	LC020714
PeAomori	Sea of Japan, Aomori Bay, Aomori Pref.	LC020670/LC020671	LC020755/LC020756	LC020715/LC020716
PeIwate	Pacific Ocean ^c , Okirai Bay, Iwate Pref.	LC020672/LC020673	LC020757/LC020758	LC020717/LC020718
PeMiyagi	Pacific Ocean ^c , Ishinomaki, Miyagi Pref.	LC020674/LC020675	LC020759/LC020760	LC020719/LC020720
PeChiba	Pacific Ocean, Okinoshima, Chiba Pref.	LC020676/LC020677	LC020761/LC020762	LC020721
PeShizuoka	Pacific Ocean, Shimoda, Shizuoka Pref.	LC020678/LC020679	LC020763/LC020764	LC020722
PeMie	Pacific Ocean, Gokasho Bay, Mie Pref.	LC020680/LC020681	LC020765/LC020766	LC020723
PeOita	Pacific Ocean, Saiki Bay, Oita Pref.	LC020682/LC020683	LC020767	LC020724
PeHiroshima	Seto Inland Sea, In'noshima, Hiroshima Pref.	LC020684/LC020685	LC020768/LC020769	LC020725
<i>Pterogobius zonoleucus</i>				
PzoKagoshima	Sea of Japan, Nagashima, Kagoshima Pref.	LC020686/LC020687	LC020770	LC020726
PzoYamaguchi	Sea of Japan, Shimomoseki, Yamaguchi Pref.	LC020688/LC020689	LC020771	LC020727/LC020728
PzoOki	Sea of Japan, Oki, Shimane Pref.	LC020690	LC020772/LC020773	LC020729/LC020730
PzoIshikawa	Sea of Japan, Tsukumo Bay, Ishikawa Pref.	LC020691/LC020692	LC020774	LC020731/LC020732
PzoAkita	Sea of Japan, Oga, Akita Pref.	LC020693/LC020694	LC020775	LC020733
PzoChiba	Pacific Ocean, Okinoshima, Chiba Pref.	LC020695/LC020696	LC020776	LC020734/LC020735
PzoMie	Pacific Ocean, Gokasho Bay, Mie Pref.	LC020697/LC020698	LC020777/LC020778	LC020736/LC020737
PzoOita	Pacific Ocean, Saiki Bay, Oita Pref.	LC020699/LC020700	LC020779	LC020738
PzoKagawa	Seto Inland Sea, Takamatsu, Kagawa Pref.	LC020701/LC020702	LC020780/LC020781	LC020739/LC020740
<i>Pterogobius virgo</i>				
Pvilshikawa	Sea of Japan, Nanao Bay, Ishikawa Pref.	LC020703	LC020782/LC020783	LC020741/LC020742
<i>Pterogobius zacalles</i>				
PzaAomori	Sea of Japan, Aomori Bay, Aomori Pref.	LC020704	LC020784	LC020743

^a The sample number 1 for each individual described in Akihito et al. (2008) was used in this study.^b Single or double accession numbers in a column indicate the nucDNA sequence obtained as a homozygote or a heterozygote in the sample, respectively.^c The populations of the northern Pacific Ocean of Honshu belong to the Sea of Japan clade.

were 1st, 2nd, and 3rd positions. The robustness of the phylogenetic tree was evaluated by bootstrap analysis by using 1000 pseudoreplicate datasets.

For ML trees, the RAxML program version 7.0.3 (Stamatakis et al., 2008) was used with the GTR + I + Γ model. Difference among codon positions in RAG1 was taken into account with the separate model (Pupko et al., 2002). Bootstrap values were computed by the rapid bootstrap method (Stamatakis et al., 2008) with 1000 replications.

2.4. Population genetic analyses

The parameter $\theta = 4N_e u$, where N_e is effective population size and u is mutation rate per sequence per generation, is called the population mutation rate or the scaled mutation rate in population genetics. The parameter $\beta = 2N_e r$ is the scaled growth rate, where r is exponential growth rate per generation. Using Genetree Version 9.0. (R.C.Griffiths, <http://www.stats.ox.ac.uk/~griff/software.html>; Griffiths and Tavaré, 1994), we estimated these two parameters for each population of *P. elapoides*, “Sea of Japan *P. zonoleucus*” and “Pacific *P. zonoleucus*” to determine the relative effective population sizes of these populations assuming equal mutation rate.

The multispecies coalescent analysis of the nuclear genes was also carried out to estimate not only the phylogenetic relationships but also the ancestral population sizes and the divergence times. The analysis was conducted using the program BPP ver. 3 (Rannala and Yang, 2003; Yang and Rannala, 2010). This method accommodates the species phylogeny as well as incomplete lineage sorting due to ancestral polymorphism. More details were given in Supplementary Information.

3. Results and discussion

3.1. Incongruence among gene trees

Analyzing CytB and ND2 genes of mitochondria, Akihito et al. (2008) showed that *P. elapoides* form a monophyletic group together with “Sea

of Japan *P. zonoleucus*”, and that “Pacific *P. zonoleucus*” is a sister-group to the *P. elapoides* “Sea of Japan *P. zonoleucus*” clade. If this represents the population level relationship among *P. elapoides*, “Sea of Japan *P. zonoleucus*”, and “Pacific *P. zonoleucus*”, it is in sharp contrast with morphological evolution, in which morphological difference between “Sea of Japan *P. zonoleucus*” and “Pacific *P. zonoleucus*” is much smaller than that between the two populations of *P. elapoides*.

In this work, we analyzed three nuclear loci; an exon of RAG1 and introns of S7RP and TBR1. The gene tree of RAG1 exon by the ML method shown in Fig. 1 gave a topology in which a clade formed by “Sea of Japan *P. zonoleucus*” is a sister-group to *P. elapoides* consistently with the mitochondrial gene tree, although “Sea of Japan *P. zonoleucus*” is not monophyletic as in the mitochondrial tree. The NJ trees shown in Supplementary Fig. S1 give essentially the same relationships with the ML trees. On the other hand, the gene trees of S7RP and TBR1 introns both gave a monophyly of *P. zonoleucus*. (Figs. 2 and 3). Although the phylogenetic relationships among *P. zonoleucus* were ambiguous from the ML analysis of TBR1 (Fig. 3), a deletion common in all *P. zonoleucus* TBR1 unambiguously supports the monophyly of *P. zonoleucus* (Fig. 4), and supports the sister-group relationship of the whole *P. zonoleucus* with *P. elapoides*. This relationship was also supported by S7RP (Fig. 2).

Thus, the gene trees of the three nuclear loci we analyzed in this work are incongruent with the gene tree of mtDNA (Akihito et al., 2008) each in some respects, and some explanations are needed for the inconsistencies as discussed below.

3.2. Gene trees vs. population trees

In the reconstruction of phylogenetic trees, it is important to distinguish between a population (or species) tree and a gene tree (Pamilo and Nei, 1988). These two can often contradict each other, and a gene tree can be different among different genes (e.g., Maddison, 1997; Funk and Omland, 2003; Degnan and Rosenberg, 2009). In the species complex of *P. zonoleucus* and *P. elapoides*, the mitochondrial tree and

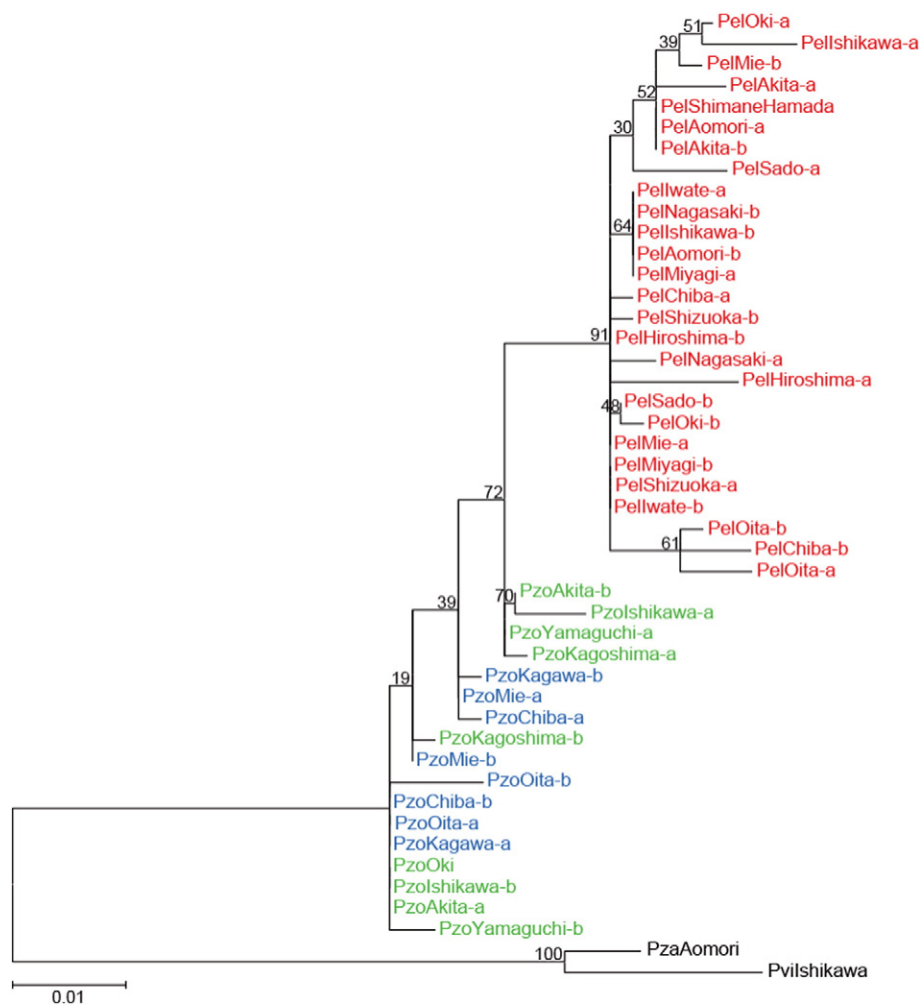


Fig. 1. ML tree of RAG-1 estimated by RAxML using the GTR + I + Γ model with partitioning among three codon positions. *P. zacalles* (Pza) and *P. virgo* (Pvi) are used as outgroup for rooting the tree. Bootstrap values from rapid bootstrap 1000 replications are shown. The scale bar corresponds to 0.01 substitutions per nucleotide site. Pel: *P. elapoides*, with red-color, Pzo: *P. zonoleucus*, Pzo of Sea of Japan with green-color, Pzo of Pacific Ocean (including Seto Inland Sea) with blue-color. "a" and "b" at the end of a label show heterozygotes in an individual, and a label without "a" and "b" represents a homozygote.

the RAG1 tree suggest the sister-group relationship between "Sea of Japan *P. zonoleucus*" and *P. elapoides* (Akihito et al., 2008 and Fig. 1, although "Sea of Japan *P. zonoleucus*" is not monophyletic in the RAG1 tree), while the S7RP and TBR1 trees support the monophyly of *P. zonoleucus* (Figs. 2 and 3). How can we compromise this incongruence among the gene trees?

Until recently, molecular phylogenetic analyses were typically based on sequences of a gene or only a few genes, and assumed the gene tree to be the estimate of the tree of population or species relationships representing the organismal history. However, as sequence data for multiple genes became increasingly available, analyses of these large data sets have made it clear that different genes can give markedly different trees (Maddison, 1997; Degnan and Rosenberg, 2009). Multiple gene data contain more information about the evolutionary history of populations than single gene data. Because of the stochastic way in which lineages sort during speciation, gene trees can be incongruent among different genes and from population trees (Pamilo and Nei, 1988), and the pattern of the incongruence among gene trees can provide us with the information on the effective population size of the common ancestor and on the timings of population separations.

In this regard, the example of the trichotomy problem among humans, chimpanzees and gorillas might shed some light on the *Pterogobius* problem. A gene tree with an unequivocal resolution of this trichotomy problem was first given by an mtDNA analysis, showing

chimpanzees are closer to humans than to gorillas (Horai et al., 1992). Although it was almost established that chimpanzees are closer to humans than to gorillas in the population (species) tree, 23% of gene trees turned out to be incongruent with the population tree of the human/chimpanzee/clade when whole genome data were analyzed (Ebersberger et al., 2007), and these incongruent gene trees are interpreted to be due to incomplete lineage sorting of polymorphisms in the common ancestral population of humans, chimpanzees and gorillas (Degnan and Rosenberg, 2009). This amount of incongruent gene trees is close to the amount expected if the time period separating the gorilla branching and the chimpanzee branching from the lineage leading to human in the population level is $2N_e$ generations, where N_e is the effective population size of the common ancestor of these three species, and if neutrality holds (Pamilo and Nei, 1988).

Another source of discord among gene trees is introgression by hybridization and this process is known in many groups of animals and plants (e.g., Gyllenstein and Wilson, 1987; Dowling and DeMarais, 1993; Mallet, 2007). It has been widely known in botany that hybridization and gene flow (introgression) between species are important forces in plant evolution (Mallet, 2007). Today, biologists increasingly find new examples of introgression due to hybridization among animal species as well (Gyllenstein and Wilson, 1987; Dowling and DeMarais, 1993), and this process cannot be dismissed in interpreting the incongruent gene trees. Funk and Omland (2003) emphasized incomplete

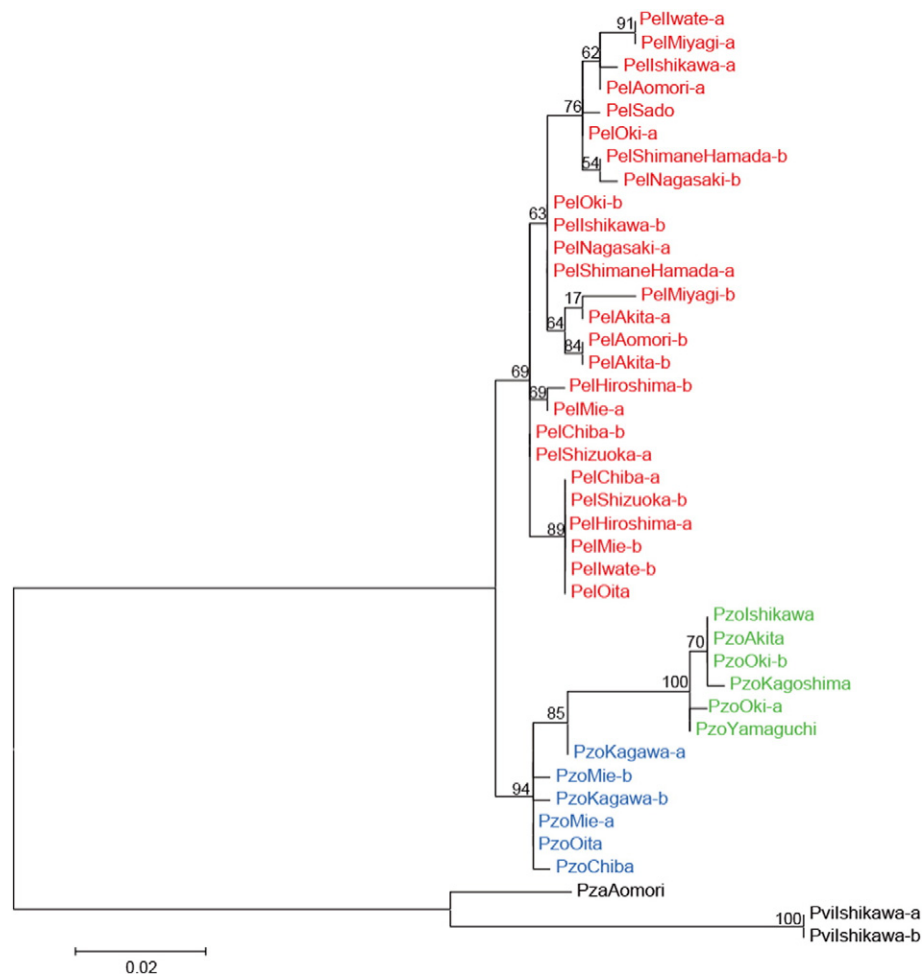


Fig. 2. ML tree of S7RP estimated by RAxML using the GTR + I + Γ model. The details of the figure are as described in the caption of Fig. 1.

lineage sorting and introgression as major causes of paraphyly and polyphyly that reflect fundamental aspects of organismal history with important biological implications.

3.3. How to resolve the incongruence among gene trees of *Pterogobius*?

We suggest two possible scenarios of the population-level evolution among *P. zonoleucus* and *P. elapoides* as follows; (1) Hypothesis-1: *P. elapoides* evolved from “Sea of Japan *P. zonoleucus*”, and (2) Hypothesis-2: “Sea of Japan *P. zonoleucus*” and “Pacific *P. zonoleucus*” form a monophyletic group, which is sister to *P. elapoides*.

Hypothesis-1 was suggested at first by Akihito et al. (2008) from the mitochondrial gene tree, and is supported by the RAG1 tree as well. However, in contrast to the monophyly of “Sea of Japan *P. zonoleucus*” in the mitochondrial tree, “Sea of Japan *P. zonoleucus*” is not monophyletic in the RAG1 tree, and, if Hypothesis-1 holds, gene flow of RAG1 must have occurred from Pacific to Sea of Japan populations as illustrated in Fig. 5A. The S7RP and TBR1 trees incongruent from this hypothesis can be interpreted as due to incomplete lineage sorting of the ancestral polymorphism. Ancestral polymorphism of neutral alleles is maintained for $2N_e$ generations in average, and if the two successive separations among populations occurred in not much longer than $2N_e$ generations, incomplete lineage sorting can occur in some noticeable frequency and the incongruent S7RP and TBR1 trees can be interpreted as such cases as shown in Fig. 5B.

Conversely, Hypothesis-2 is supported by the S7RP and TBR1 trees, and the mitochondrial and RAG1 trees incongruent from this hypothesis

can potentially be interpreted as due to incomplete lineage sorting (Figs. 6 and 7). The extent of the retention of ancestral polymorphism, however, is directly related to the effective population size, N_e , and N_e of mitochondria is 1/4 of those of nuclear autosomal genes because the mitochondrial genome is effectively haploid and is maternally inherited (Birky et al., 1989). Therefore, the incongruence of the mitochondrial tree from the population tree of Hypothesis-2 seems unlikely to be due to incomplete lineage sorting of ancestral polymorphism although we could not exclude such a possibility. If Hypothesis-2 holds, this incongruence may more likely be attributed to ancient introgression by hybridization (e.g., Gyllenstein and Wilson, 1987; Dowling and DeMarais, 1993; Mallet, 2007) (Fig. 7). The direction of introgression of mtDNA and RAG1 might be either from “Sea of Japan *P. zonoleucus*” to *P. elapoides* (Fig. 7A) or the opposite direction (Fig. 7B). Although the incomplete lineage sorting seems unlikely in explaining the incongruent mtDNA tree because of small effective population size of mitochondria, the incongruent RAG1 tree may be due either to incomplete lineage sorting (Fig. 6) or to ancient introgression from “Sea of Japan *P. zonoleucus*” to *P. elapoides* or the opposite (Fig. 7A and B).

The rapid pace of mtDNA nucleotide substitutions, coupled with the special mode of maternal non-recombining mtDNA inheritance, offers advantages of using this marker for phylogenetic analyses particularly at lower taxonomic level (for example, congeneric level as discussed here) that will not be matched easily by any nuclear markers (Avise et al., 1987). Because of the reduced N_e of mitochondrial genes compared to nuclear genes, ancestral polymorphism of mitochondrial genes is expected to be retained only a shorter time than that of nuclear genes. Thus, incomplete lineage sorting is less of a concern for

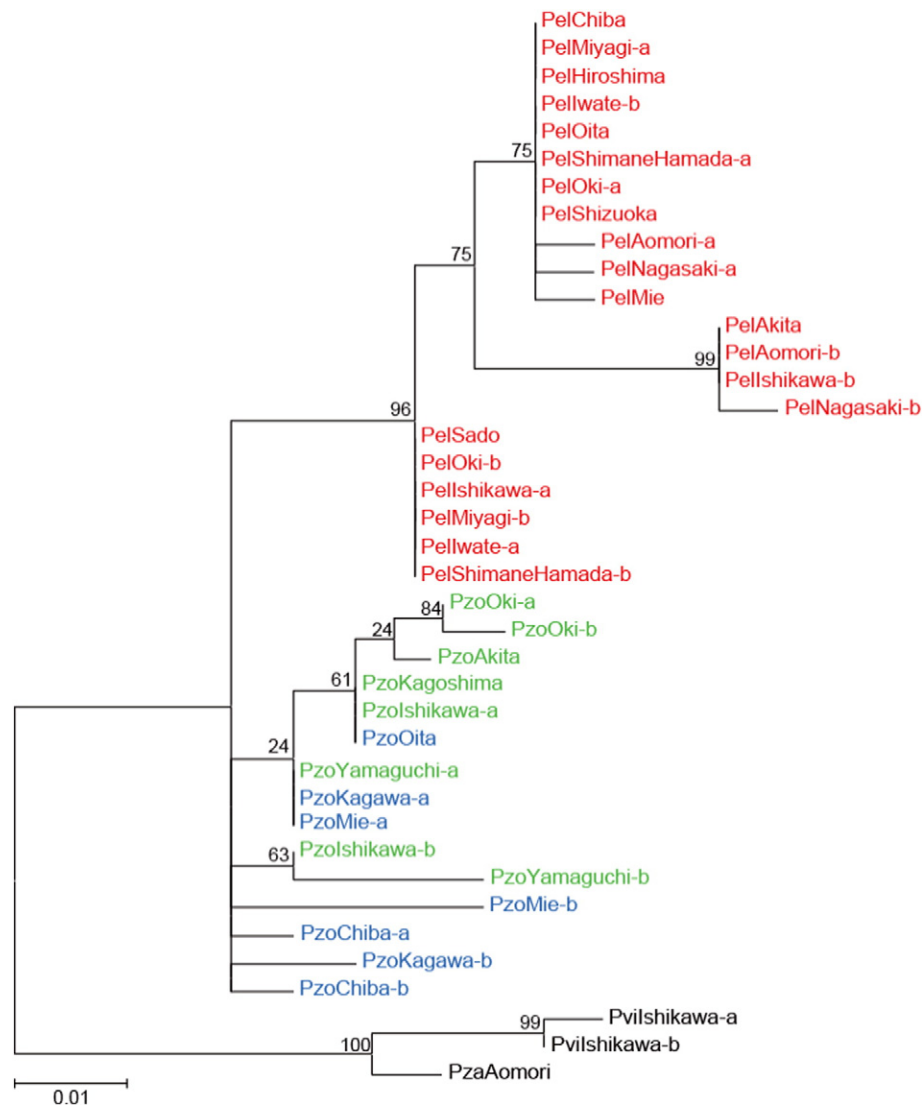


Fig. 3. ML tree of TBR1 estimated by RAxML using the GTR + I + Γ model. The details of the figure are as described in the caption of Fig. 1.

mitochondrial than for nuclear gene trees, other things being equal, providing an additional advantage to using mitochondrial gene trees as estimates of population trees (Avice et al., 1987; Palumbi et al., 2001; Hudson and Turelli, 2003). On the other hand, however, the smaller N_e of mitochondrial genes may facilitate fixation of the introgressed haplotype lineage, such that even low levels of introgression may be sufficient to establish a mitochondrial haplotype in a foreign population (Takahata and Slatkin, 1984). The small N_e of mitochondria is thus a double-edged sword for phylogenetic studies.

3.4. Which hypothesis is more likely to be correct

If Hypothesis-1 is true, it is expected that the population size of *P. elapoides* is smaller than that of “Sea of Japan *P. zonoleucus*” at least immediately after the former separated from the latter as illustrated in Fig. 5A and B. Estimation of $\theta = 4N_eu$, where N_e is the current effective population size and u is the mutation rate per sequence per generation, was carried out for each population of *P. elapoides*, “Sea of Japan *P. zonoleucus*”, and “Pacific *P. zonoleucus*” with Genetree software Version 9.0. With the assumption of equal mutation rate for a particular sequence, we could compare the relative population size among the three populations by using θ as shown in Supplementary. Table S2. The estimated θ , and accordingly the population size N_e , of *P. elapoides* is larger

than that of “Sea of Japan *P. zonoleucus*” consistently for the three nuclear genes, RAG1, S7RP and TBR1 (Supplementary. Table S2), in contradiction with the expectation of Hypothesis-1. As for mtDNA, although N_e of *P. elapoides* is smaller than that of “Sea of Japan *P. zonoleucus*” at present, it is due to the rapid population expansion of the latter, and the effective population size of *P. elapoides* was larger when the speciation occurred (Supplementary. Fig. S2A). Therefore, Hypothesis-2 seems more likely from the evidence currently available. Furthermore, smaller estimated effective population size of “Sea of Japan *P. zonoleucus*” mtDNA than that of *P. elapoides* at the time of speciation suggests that the introgression was from the latter to the former as shown in Fig. 7B. On the other hand, the ancestral population size of *P. elapoides* is comparable to that of “Sea of Japan *P. zonoleucus*” for RAG1 (Supplementary Fig. S2B), and therefore the history of RAG1 gene might be different from that of mtDNA.

The species tree as inferred from the nucleotide sequences of RAG1, TBR1 and S7RP by using the multispecies coalescent method is shown in the Fig. 8. As discussed in Supplementary Information, the analysis also supports Hypothesis-2B shown in Fig. 7B.

A similar discordance between nuclear and mitochondrial gene trees was shown concerning the evolution of polar bears (*Ursus maritimus*). Early analyses of mtDNA suggested that polar bears evolved from within brown bears (*Ursus arctos*) (Shields et al., 2000; Lindqvist et al., 2010).

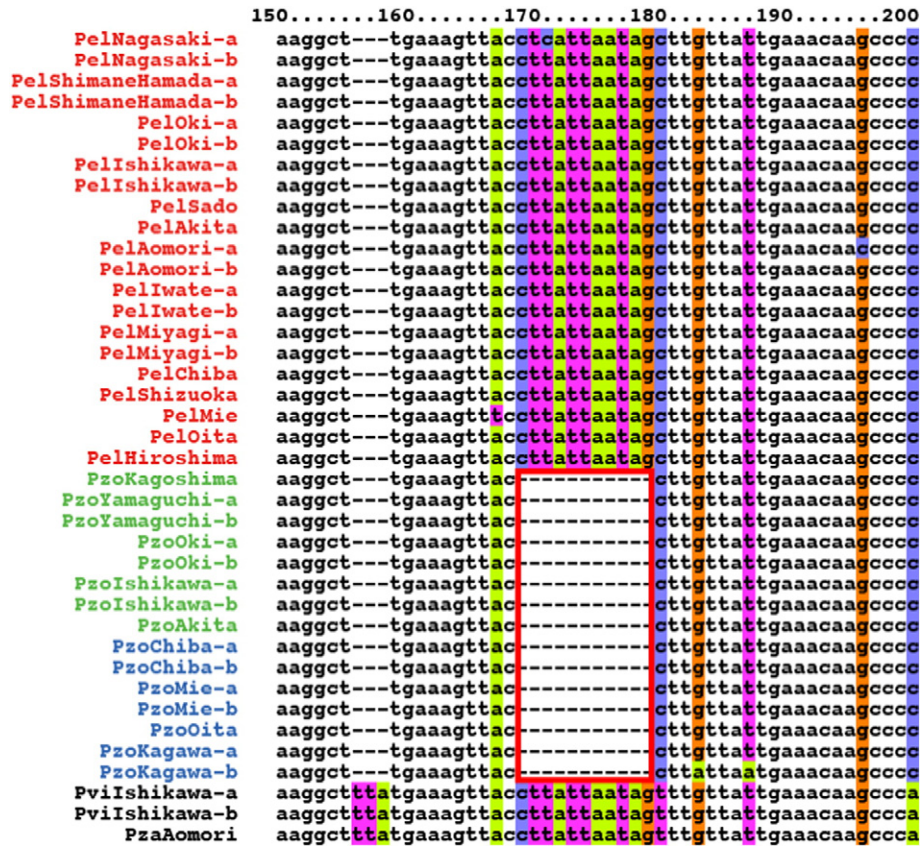


Fig. 4. Multiple alignment of TBR1 nucleotide sequence positions from 150 to 200. The red open rectangular box indicates a deletion observed only in *P. zonoleucus* examined in this study. A color was shown when the nucleotide was not conserved at a position.

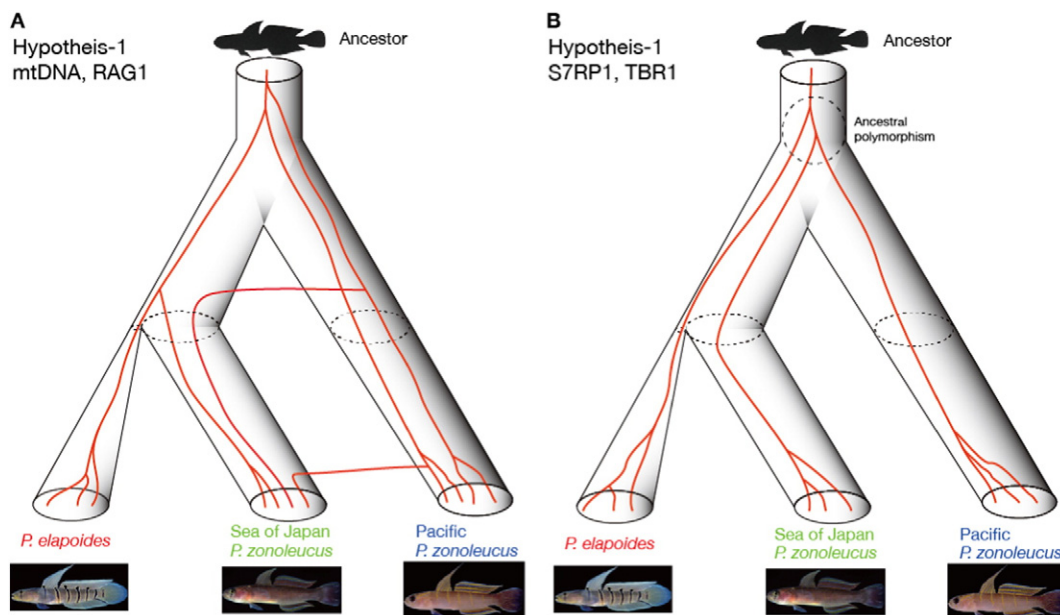


Fig. 5. Gene trees of mtDNA and RAG1 (A) and S7RP and TBR1 (B) under Hypothesis-1. The species and gene trees under hypothesis-1 were shown as cylinders and red lines, respectively. The species tree suggests that *P. elapoides* evolved from “Sea of Japan *P. zonoleucus*” consistently with the mitochondrial tree. Under this hypothesis, *P. elapoides* evolved from “Sea of Japan *P. zonoleucus*” at the time shown by the dotted flat circles. The figure also illustrates a smaller population size of *P. elapoides* than that of the mother population when the new species emerged. The pictures of *P. elapoides*, “Sea of Japan *P. zonoleucus*”, and “Pacific *P. zonoleucus*” were obtained from Akihito et al., 2008. (A) Contrary to the mitochondrial gene tree, in which “Sea of Japan *P. zonoleucus*” is monophyletic, gene flow occurred from Pacific to Sea of Japan populations of *P. zonoleucus* as represented by the two horizontal lines. (B) The ancestral polymorphism was maintained until the separation of *P. elapoides* from “Sea of Japan *P. zonoleucus*”.

Hypothesis-2 mtDNA, RAG1

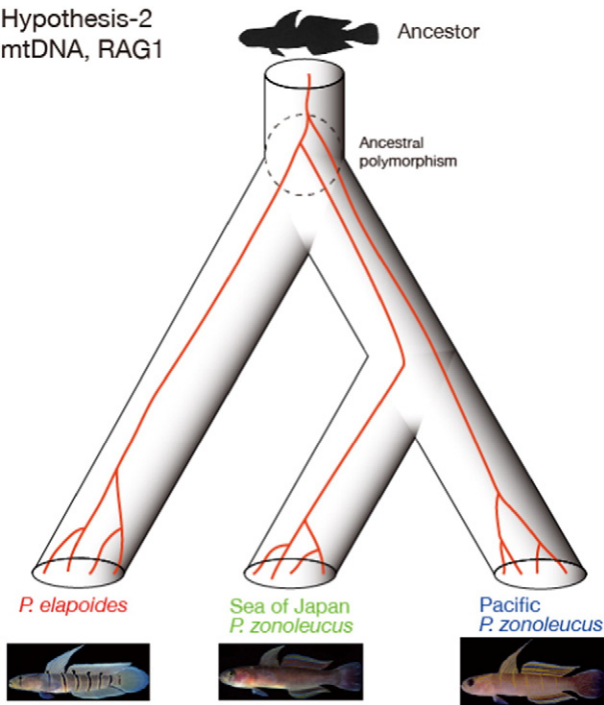


Fig. 6. Gene trees of mtDNA and RAG1 under Hypothesis-2. The ancestral polymorphism of the common ancestral population of *P. elapoides* and *P. zonoleucus* was maintained until the separation between “Sea of Japan *P. zonoleucus*” and “Pacific *P. zonoleucus*”. Since the effective population size of mtDNA is 1/4 of those of nuclear autosomal genes, the duration time of the ancestral polymorphism of mtDNA is expected to be accordingly short and it seems difficult to attribute the incongruence of mitochondrial tree to the ancestral polymorphism based on Hypothesis-2.

On the other hand, recent concatenated analyses of nuclear genes by Hailer et al. (2012) and Miller et al. (2012) suggested the monophyly of brown bears, and these authors attributed this discrepancy to the introgression of mitochondria from brown bears to polar bears. However, a population genetic analysis of Hailer et al.’s data does not support these authors’ hypothesis of the brown bear monophyly (Nakagome et al., 2013), suggesting the concatenated analysis of independent

nuclear genes with different histories is sometimes not reliable (Song et al., 2012). While the goat genus *Capra* is the sister-group to the genus *Hemitragus* in the mitochondrial gene tree, the gene tree from four nuclear loci strongly suggests that *Capra* is more distantly related to *Hemitragus* in Caprinae. Ropiquet and Hassanin (2006) attributed this incongruence to the introgression of mitochondria from *Hemitragus* to the common ancestor of *Capra*.

Another example of a similar discordance among gene trees was also found in African elephant species (Roca et al., 2005). African elephant genus *Loxodonta* consists of two species, savanna elephants, *Loxodonta africana*, and forest elephants, *Loxodonta cyclotis*. While nuclear gene trees indicated deep genetic separation between savanna and forest elephants, some individuals of savanna elephants with savanna-specific nuclear genes carry forest elephant-type mtDNA. In the case of African elephants, savanna elephants with forest elephant-type mtDNAs are intermingled with forest elephants in the tree, and savanna elephants with forest elephant-type mtDNA do not form a monophyletic clade. Probably, mtDNAs in the savanna elephant population is in the process of fixation. In the future, stochastic mtDNA lineage extinctions will occur and there is a possibility of realizing a similar pattern with that of *Pterogobius* when the savanna elephant population with forest elephant-type mtDNA and forest elephants are replaced by “Sea of Japan *P. zonoleucus*” and *P. elapoides*, respectively, in the case of the scenario shown in Fig. 7. The pattern observed in the African elephants might be regarded as a pattern just after the ancient Fig. 7B introgression occurred from *P. elapoides* to “Sea of Japan *P. zonoleucus*”.

3.5. Difference of the mitochondrial gene tree from the nuclear gene trees

In the mitochondrial tree, individuals from the same local population tend to form a clade excluding individuals from other local populations, and Sea of Japan populations and Pacific populations of both species of *P. zonoleucus* and *P. elapoides* are not intermingled (Akihito et al., 2008), while in the nuclear gene trees, genes from Sea of Japan and Pacific populations are intermingled in the trees. Although only one individual from each local population was sampled in our nuclear gene analyses, heterozygous nuclear genes from an individual tend to scatter in the tree (Figs. 1–3). Since it seems difficult to explain this observation by sexual difference of the behavior with respect to migration of the two species of *Pterogobius*, an alternative explanation is needed.

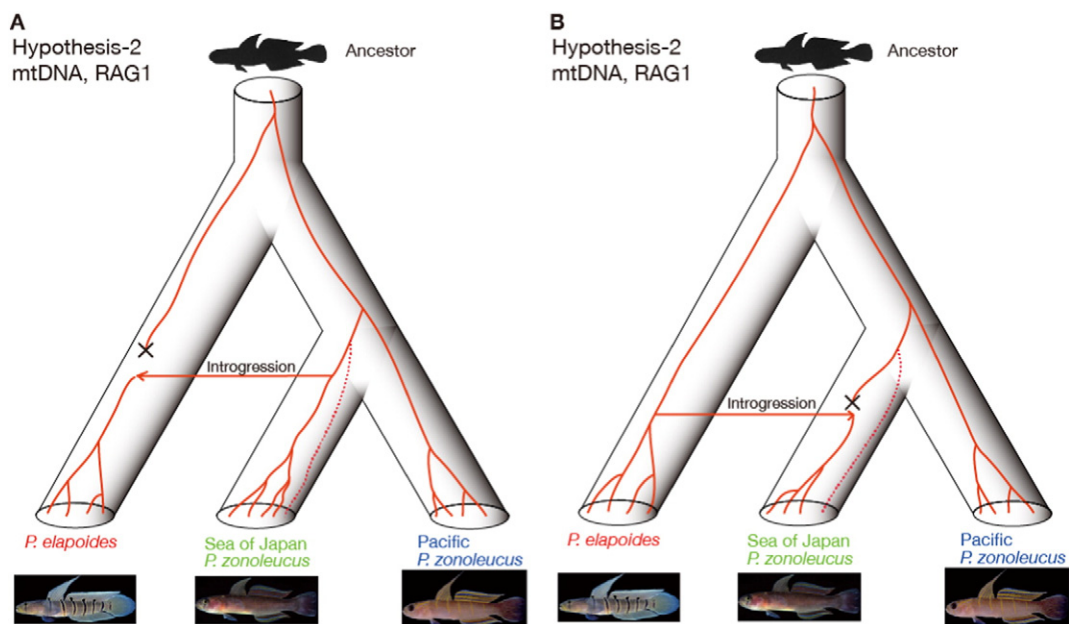


Fig. 7. Introgression of mtDNA and RAG1 under Hypothesis-2. (A) Introgression occurred from “Sea of Japan *P. zonoleucus*” to *P. elapoides*. The dotted line shows a basal lineage of “Sea of Japan *P. zonoleucus*” RAG1 separated before the hypothetical introgression. (B) Introgression of the opposite direction.

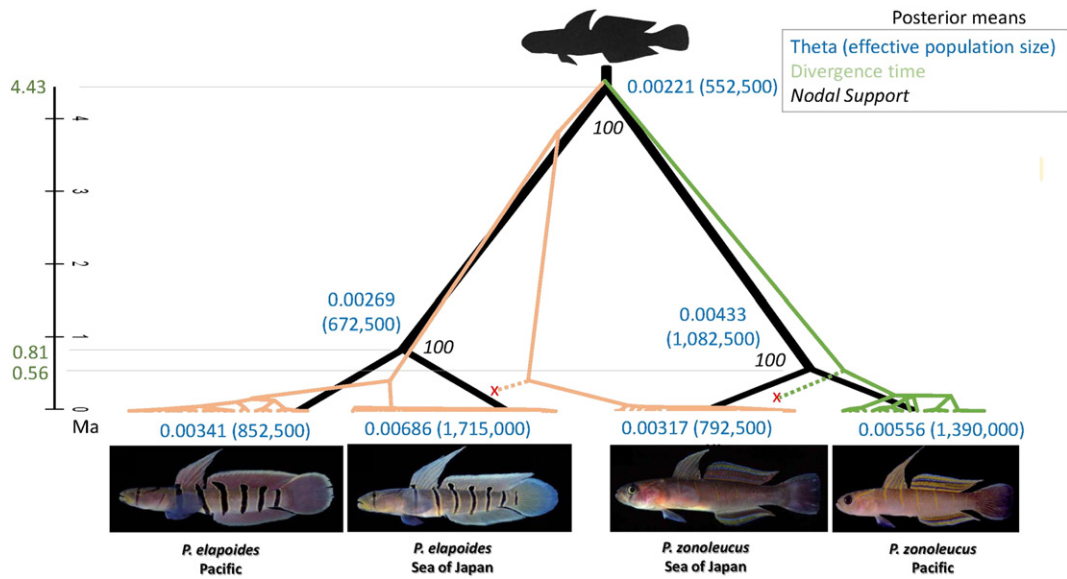


Fig. 8. The species tree (bold black line) as inferred from the nucleotide sequences of RAG1, TBR1, S7RP by using the multi-species coalescent method (Rannala and Yang, 2003; Yang and Rannala, 2010) and the time calibrated mitochondrial phylogenetic tree (thin colored line). The lengths of vertical component of each branch are proportional to the geological time scale in Mega annum (Ma: million years ago). The nodal numbers of the species tree indicate the posterior mean of the θ (numbers in parentheses indicate effective population sizes) in blue fonts, and the posterior probabilities of the nodal support in black italic fonts. The details of the parameters were summarized in the Supplementary Table S3. Note the θ s in this figure are per site per year, while the θ s in the Supplementary Table S2 are per sequence per generation. Thin green lines indicate the mitochondrial tree of *P. zonoleucus* typed lineage and thin dark pink lines indicate the mitochondrial tree of *P. elapoides* typed lineage. The short branches indicated by the dashed lines are putative extinct mitochondrial lineages of *P. elapoides* and *P. zonoleucus* probably distributed in the Sea of Japan, and the \times symbols in red color indicate the extinctions of these lineages. The absolute times in this figure should be regarded as rough estimates, because they are based on many assumptions.

As we noted before, because of the smaller N_e of mtDNA, stochastic mtDNA lineage extinction within a population is expected to occur at a rapid pace. In other words, mtDNA trees are continually “self-pruning” (Avice et al., 1987). This line of reasoning may largely account for the limited mtDNA sequence divergence observed within a local population both in *P. zonoleucus* and *P. elapoides*.

4. Conclusion

This evolutionary work of the two gobioid species highlights the need of analyzing multiple gene trees for both nuclear and mitochondrial genes as well as scrutinization of morphological characteristics to obtain a population tree representing the organismal evolutionary history.

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.gene.2015.10.014>.

Acknowledgments

We would like to thank Takako Sako of Imperial Household Agency for her aids to the research during the entire process of our activities, and Dr. Ziheng Yang at University College London for his helpful comments.

References

- Akihito, Iwata, A., Kobayashi, T., Ikeo, K., Imanishi, T., Ono, H., Umehara, Y., Hmamatsu, C., Sugiyama, K., Ikeda, Y., Sakamoto, K., Fumihito, A., Ohno, S., Gojobori, T., 2000. Evolutionary aspects of gobioid fishes based upon a phylogenetic analysis of mitochondrial cytochrome *b* genes. *Gene* 259 (1–2), 5–15.
- Akihito, Fumihito, A., Ikeda, Y., Aizawa, M., Makino, T., Umehara, Y., Kai, Y., Nishimoto, Y., Hasegawa, M., Nakabo, T., Gojobori, T., 2008. Evolution of Pacific Ocean and the sea of Japan populations of the gobioid species, *Pterogobius elapoides* and *Pterogobius zonoleucus*, based on molecular and morphological analyses. *Gene* 427 (1–2), 7–18.
- Avice, J.C., Arnold, J., Ball, R.M., Bermingham, E., Lamb, T., Neigel, J.E., Reeb, C.A., Saunders, N.C., 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematic. *Annu. Rev. Ecol. Syst.* 18, 489–522.
- Birky, C.W., Fuerst, P., Maruyama, T., 1989. Organellar gene diversity under migration, mutation and drift: equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. *Genetics* 121 (3), 613–627.

- Brown, W.M., George, M., Wilson, A.C., 1979. Rapid evolution of animal mitochondrial DNA. *Proc. Natl. Acad. Sci. U. S. A.* 76 (4), 1967–1971.
- Degnan, J.H., Rosenberg, N.A., 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol. Evol.* 24 (6), 332–340.
- Dowling, T.E., DeMarais, B.D., 1993. Evolutionary significance of introgressive hybridization in cyprinid fishes. *Nature* 362 (6419), 444–446.
- Dupuis, J.R., Roe, A.D., Sperling, F.A.H., 2012. Multi-locus species delimitation in closely related animals and fungi: one marker is not enough. *Mol. Ecol.* 21 (18), 4422–4436.
- Ebersberger, I., Gálgoczy, P., Taudien, S., Taenzler, S., Playzer, M., von Haeseler, A., 2007. Mapping human genetic ancestry. *Mol. Biol. Evol.* 24 (10), 2266–2276.
- Felsenstein, J., 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* 17 (6), 368–376.
- Funk, D.J., Omland, K.E., 2003. Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annu. Rev. Ecol. Syst.* 34, 397–423.
- Griffiths, R.C., Tavaré, S., 1994. Simulating probability distributions in the coalescent. *Theor. Popul. Biol.* 46 (2), 131–159.
- Gyllenstein, U.B., Wilson, A.C., 1987. Interspecific mitochondrial DNA transfer and the colonization of Scandinavia by mice. *Genet. Res. Camb.* 49 (1), 25–29.
- Hailer, F., Kutschera, V.E., Hallström, B.M., Klassert, D., Fain, S.R., Leonard, J.A., Arnason, U., Janke, A., 2012. Nuclear genomic sequences reveal that polar bears are an old and distinct bear lineage. *Science* 336, 344–347.
- Horai, S., Satta, Y., Hayasaka, K., Kondo, R., Inoue, T., Ishida, T., Hayashi, S., Takahata, N., 1992. Man's place in hominoida revealed by mitochondrial DNA genealogy. *J. Mol. Evol.* 35 (1), 32–43.
- Hudson, R.R., Turelli, M., 2003. Stochasticity overrules the “three-times rule”: genetic drift, genetic draft, and coalescence times for nuclear loci versus mitochondrial DNA. *Evolution* 57 (1), 182–190.
- Katoh, K., Kuma, K., Toh, H., Miyata, T., 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* 33 (2), 511–518.
- Lindqvist, C., Schuster, S.C., Sun, Y., Talbot, S.L., Qi, J., Ratan, A., Tomsho, L.P., Kasson, L., Zeyl, E., Aars, J., Miller, W., Ingólfsson, O., Bachmann, L., Wiig, Ø., 2010. Complete mitochondrial genome of a Pleistocene jawbone unveils the origin of polar bear. *Proc. Natl. Acad. Sci. U. S. A.* 107 (11), 5053–5057.
- Maddison, W.P., 1997. Gene trees in species trees. *Syst. Biol.* 46 (3), 523–536.
- Mallet, J., 2007. Hybrid speciation. *Nature* 446 (7133), 279–283.
- Miller, W., Schuster, S.C., Welch, A.J., Ratan, A., Bedoya-Reina, O.C., et al., 2012. Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change. *Proc. Natl. Acad. Sci. U. S. A.* 109 (36), E2382–E2390.
- Nakagome, S., Mano, S., Hasegawa, M., 2013. Comment on “nuclear genomic sequences reveal that polar bears are an old and distinct bear lineage”. *Science* 339 (6127), 1522.
- Nei, M., 1987. *Molecular Evolutionary Genetics*. Columbia University Press, New York, USA (512 pp.).
- Nelson, J.S., 2006. *Fishes of the World*. 4th edition. John Wiley and Sons, Inc., New York (601 pp.).
- Palumbi, S.R., Cipriano, F., Hare, M.P., 2001. Predicting nuclear gene coalescence from mitochondrial DNA: the three-times rule. *Evolution* 55 (2), 859–868.

- Pamilo, P., Nei, M., 1988. Relationships between gene trees and species trees. *Mol. Biol. Evol.* 5 (5), 568–583.
- Pezold, F.L., 2011. In: Patzner, R., Van Tassell, J.L., Kovacic, M., Kapoor, B.G. (Eds.), *Systematics of the family Gobionellidae. The biology of gobies*. Science Publishers, Channel Islands, UK, pp. 87–97.
- Piganeau, G., Gardner, M., Eyre-Walker, A., 2004. A broad survey of recombination in animal mitochondria. *Mol. Biol. Evol.* 21 (12), 2319–2325.
- Pupko, T., Huchon, D., Cao, Y., Okada, N., Hasegawa, M., 2002. Combining multiple data sets in a likelihood analysis: which models are the best? *Mol. Biol. Evol.* 19 (12), 2294–2307.
- Rannala, B., Yang, Z., 2003. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics* 164, 1645–1656.
- Roca, A.L., Georgiadis, N., O'Brien, S.J., 2005. Cytonuclear genomic dissociation in African elephant species. *Nat. Genet.* 37 (1), 96–100.
- Ropiquet, A., Hassanin, A., 2006. Hybrid origin of the Pliocene ancestor of wild goats. *Mol. Phylogenet. Evol.* 41 (2), 395–404.
- Rubinoff, D., Holland, B., 2005. Between two extremes: mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. *Syst. Biol.* 54 (6), 952–961.
- Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4 (4), 406–425.
- Shields, G.F., Adams, D., Gamer, G., Labelle, M., Pietsch, J., Ramsay, M., Schwartz, C., Titus, K., Williamson, S., 2000. Phylogeography of mitochondrial DNA variation in brown bears and polar bears. *Mol. Phylogenet. Evol.* 15 (2), 319–326.
- Song, S., Liu, L., Edwards, S.V., Wu, S., 2012. Resolving conflict in eutherian mammal phylogeny using phylogenomics and multispecies coalescent model. *Proc. Natl. Acad. Sci. U. S. A.* 109 (37), 14942–14947.
- Stamatakis, A., Hoover, P., Rougemont, J., 2008. A rapid bootstrap algorithm for the RAxML web servers. *Syst. Biol.* 57 (5), 758–771.
- Takahata, N., Slatkin, M., 1984. Mitochondrial gene flow. *Proc. Natl. Acad. Sci. U. S. A.* 81 (6), 1764–1767.
- Tamura, K., Nei, M., Kumar, S., 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc. Natl. Acad. Sci. U. S. A.* 101 (30), 11030–11035.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28 (10), 2731–2739.
- Yang, Z., Rannala, B., 2010. Bayesian species delimitation using multilocus sequence data. *Proc. Natl. Acad. Sci. U. S. A.* 107, 9264–9269.