The origins of limnetic forms and cryptic divergence in *Gnathopogon* fishes (Cyprinidae) in Japan

Ryo Kakioka · Tomoyuki Kokita · Ryoichi Tabata · Seiichi Mori · Katsutoshi Watanabe

R. Kakioka · R. Tabata · K. Watanabe

Graduate School of Science, Kyoto University, Kitashirakawa-Oiwake-cho, Sakyo, Kyoto 606-8502, Japan

e-mail: kakioka@terra.zool.kyoto-u.ac.jp

T. Kokita

Department of Marine Bioscience, Fukui Prefectural University, 1-1 Gakuen-cho, Obama, Fukui 917-0003, Japan

S. Mori

Biological Laboratory, Gifu-Keizai University, 5-50 Kitagata, Ogaki, Gifu 503-8550, Japan

Corresponding author: Ryo Kakioka; Graduate School of Science, Kyoto University, Kitashirakawa-Oiwake-cho, Sakyo, Kyoto, 606-8502 Japan; Tel.: +81-75-753-4077; Fax: +81-75-753-4100; e-mail: kakioka@terra.zool.kyoto-u.ac.jp

Running title: Divergence in *Gnathopogon* fishes

Abstract

The cyprinid species of the genus *Gnathopogon*, exhibiting flexible morphological and ecological variation, include limnetic life forms. We examined the origin of the limnetic forms and the population divergence of the Japanese *Gnathopogon* species, using molecular phylogenetic and phylogeographic analyses. A Bayesian phylogenetic inference approach based on mtDNA cytochrome *b* sequence data revealed three major lineages in *G. elongatus*. One of them formed a monophyletic group with the limnetic species *G. caerulescens*, which is endemic to an ancient lake, Lake Biwa. The divergence of the *G. caerulescens* lineage was estimated to date back to the early Pleistocene. This precedes the formation of the extensive pelagic environment in the present Lake Biwa. However, the recent genetic divergence of *G. caerulescens* was inferred to originate in the present Lake Biwa in the late Pleistocene. Another lacustrine population in the Mikata Lakes was shown to belong to a different lineage from *G. caerulescens.* The majority of the population possessed unique, but non-monophyletic, haplotypes, suggesting a short evolutionary history. One of the cryptic lineages of *G. elongatus* discovered in the Ina Valley, the lower area of Lake Suwa, might be related to the extinct lacustrine subspecies *G. elongatus suwae*, which has been replaced by introduced congeners. The previous and ongoing introductions of *Gnathopogon* fishes would have produced genetic disturbance to the indigenous populations.

Keywords

Lacustrine form · Lake Biwa · Mikata Lakes · Lake Suwa · divergence time · Bayesian random local clock model

Introduction

The family Cyprinidae is the most speciose group of freshwater fish. This group includes fishes with a highly diverse morphology, ecology, and physiology that are adapted to the vast range of habitats and resources they utilize (Winfield and Nelson 1991; Eschmeyer and Fricke 2011; Froese and Pauly 2011). The Gobioninae is a monophyletic group within the family (Tang et al. 2011), and, with rare exceptions, they primarily live on the bottom of streams. One such exception is the limnetic *Gnathopogon caerulescens* (Bănărescu and Nalbant 1973; Kotellat and Freyhof 2007). The genus *Gnathopogon* consists of nine species occurring in East Asia. The range of the genus includes the Russian Far East, China, the Korean Peninsula, and the Japanese Archipelago (Eschmeyer and Fricke 2011; Froese and Pauly 2011). Recent molecular phylogenetic studies (e.g., Yang et al. 2006; Saitoh et al. 2006, 2011; Mayden et al. 2009; Tang et al. 2011) consistently support the traditional taxonomic placement of *Gnathopogon* in the Gobioninae (e.g., Jordan and Fowler 1903; Bănărescu and Nalbant 1973), although in certain literature (e.g., Hosoya 1986, 1987, 2000) the genus is classified in the Barbinae based on its jaw structure. Two *Gnathopogon* species are endemic to Japan: *Gnathopogon elongatus*, found in the central to western regions of Honshu Island and Shikoku Island, and *Gnathopogon caerulescens*, which is endemic to Lake Biwa in central Honshu (Hosoya 2000, 2001).

Gnathopogon elongatus is a common and widespread species found in rivers and ponds. This species is also known to show substantial morphological variation in its swimming- and foraging-related apparatus (Hosoya 1987). In contrast, *G. caerulescens* is known to have a set of morphological features specialized to the limnetic lifestyle in Lake Biwa (e.g., a slender body, an upward-pointing mouth, and fine gill rakers; Hosoya 1987, 2000; Nakajima 1994). With its pelagic lifestyle, *G. caerulescens* has been hypothesized to be derived from *G. elongatus*, the morphologically flexible generalist species, and to have adapted to the extensive pelagic zone of Lake Biwa (Hosoya 1987; Nakajima 1994; Kawanabe 1996). Lake Biwa consists of a large, deep

northern basin (surface area 617.8 km^2; mean and maximum depths 43 and 103.6 m, respectively) and a small, shallow southern basin (area 52.5 km^2; mean and maximum depths 4 and 7 m, respectively; see Fig. 1, inset). It is the largest lake in Japan and is well known as an ancient lake with a history of over 4 million years (Myr). However, the present northern basin, with its developed pelagic area, appeared at the most recent stage of the lake, approximately 0.4 million years ago (mya) (Yokoyama 1984; Kawabe 1989, 1994). Accordingly, *G. caerulescens* is hypothesized to have originated during the middle to late Pleistocene, after the development of the northern basin of Lake Biwa (Tomoda 1978; Nakajima 1994; Kawanabe 1996). This species has attracted attention as a typical case of adaptive speciation in a novel environment. Such adaptive speciation is also known from the divergences of the limnetic forms of sticklebacks or charrs in postglacial lakes (e.g., Schluter et al. 1992; Snorrason et al. 1992; Schluter 1998). The adaptation of *Gnathopogon* species to the pelagic environment has also been hypothesized in other lakes. The Mikata Lakes, located northwest of Lake Biwa, are inhabited by a *G. caerulescens*-like fish (Hosoya 1987). Their origin and relationship to *G. caerulescens* have not been clarified. Moreover, another *Gnathopogon* population presumably adapted to pelagic life, *Gnathopogon elongatus suwae* is known from Lake Suwa and Lake Kizaki, located in central Honshu (Jordan and Hubbs 1925). This fish is, however, believed to have become extinct during the 1960s.

It is probable that *Gnathopogon* includes several limnetic forms. The genus is a potential model system for the study of adaptive population divergence and speciation. However, no contemporary approaches (e.g., molecular phylogenetics and geometric morphometrics) have been applied to the study of the evolution of *Gnathopogon* fishes. Indeed, although Lake Biwa is the definitive example of an ancient lake in East Asia (Kawanabe 1996), studies based on contemporary approaches for other endemic animals and plants in Lake Biwa are lacking. An exception is a molecular phylogenetic study of the goby *Gymnogobius isaza,* which is endemic to Lake Biwa. The study suggested that this goby was derived from its amphidromous sister group in the late Pliocene, prior to the development of the vast, deep northern basin in the middle to late

Pleistocene (Harada et al. 2002), despite this goby's present dependence on this environment in the northern basin.

The primary purpose of this study was to reveal the genetic relationships and divergence times among Japanese *Gnathopogon* species and regional populations, especially focusing on lacustrine populations. We used molecular phylogenetic and population genetic approaches with specimens collected from their entire native ranges in Japan. The nucleotide sequence of the mitochondrial cytochrome *b* gene was used as the molecular marker because of the substantial accumulation of data in fishes. Based on the phylogeny, we examined the previous hypotheses on the origin and speciation of the limnetic forms of Japanese *Gnathopogon*.

Materials and methods

Specimen collection

The mtDNA sequence data were obtained from 513 specimens of *Gnathopogon elongatus* from 43 localities (locality code #1–43) and 56 samples of *G. caerulescens* from four sites (#44–47) in Lake Biwa (Table 1; Fig. 1). These samples included populations that have been affected by artificial introductions, as inferred from the mtDNA data and records of introductions by fishery activities (e.g., Takei 2007; Sakai 1995).

Laboratory procedures and analyses

The genetic divergence and population structure were evaluated using the nucleotide sequences of the 3′-half of the mitochondrial cytochrome *b* gene [cyt*b*; 598 base pairs (bp); hereafter, the "short sequence"]. Nearly complete cyt*b* sequences (1,125 bp) were also determined for a number of the

specimens (n = 16) to obtain more robust phylogenetic relationships (hereafter, the "long sequence"; these haplotypes are denoted with an "L").

Total genomic DNA was isolated from fin clips preserved in 100% ethanol, using a Genomic DNA Purification Kit (Promega, Madison, WI, USA). Polymerase chain reaction (PCR) amplification was performed using the primer pair L14724 (5′-TGA CTT GAA RAA CCA YCG YYG-3′) (Palumbi et al. 1991) and H15915 (5′-ACC TCC GAT CTY CGG ATT ACA AGA C-3′) (Aoyama et al. 2000) to obtain the sequence of the entire cyt*b* gene. The PCR conditions consisted of 30 cycles of denaturation (94°C, 15 s), annealing (48°C, 15 s), and extension (72°C, 30 s), using a PC808 thermal cycler (ASTEC, Shime, Fukuoka, Japan). After purifying the PCR products by treatment with ExoSAP-It (USB Corporation, Cleveland, OH, USA) at 37°C, they were sequenced using an automated DNA sequencer (ABI Prism GA310 or 3130xl; Applied Biosystems, Foster City, CA, USA) with the above amplification primers and using a BigDye Terminator Cycle Sequencing FS Ready Reaction Kit ver. 1.1 or 3.1 (Applied Biosystems). The obtained sequences were deposited in DDBJ/GenBank/EMBL (accession numbers AB677321–AB677453). The haplotype frequencies of each population were deposited in GEDIMAP (http://gedimap.zool.kyoto-u.ac.jp; Watanabe et al. 2010) with population IDs P1382–P1428.

A phylogenetic analysis was conducted for two data sets of mtDNA, namely, the short-sequence data for all of the specimens and the long-sequence data for selected specimens. The latter were chosen to represent each of the lineages suggested by phylogenetic analysis with the short sequences. For the former data, an unrooted tree was reconstructed by the neighbor-joining algorithm (NJ; Saitou and Nei 1987) using PAUP*4.0b10 (Swofford 2002). The genetic distances were calculated under a TIM + G model selected by Akaike's information criterion (AIC), as implemented in Modeltest 3.7 (Posada and Crandall 1998). The robustness of the NJ tree was assessed using the bootstrap method (BP) with 1,000 replicates by PAUP*. In addition, statistically parsimonious networks were constructed using TCS 1.2.1 (Clement et al. 2000). There were no insertions/deletions in our dataset.

For the long-sequence data set, the cyt*b* sequences of three congeneric species, *Gnathopogon nicholsi* (AY952997), *Gnathopogon imberbis* (AY952998), and *Gnathopogon strigatus* (AY952999; referred to as *Paraleucogobio strigatus*), all reported by Yang et al. (2006), were used as the outgroup. In addition, the sequences of *Sarcocheilichthys variegatus microoculus* (AB054124; Saitoh et al. 2003), *Pseudorasbora parva* (AB677449; this study), *Pseudorasbora pumila pumila* (AB677452, AB677453; this study), and *Pseudorasbora pumila* subsp. (sensu Hosoya 2000; AB677450, AB677451; this study) were used as the outgroup of *Gnathopogon* species, because they are all included in the tribe Sarcocheilichthyini in the Gobioninae, together with *Gnathopogon* (Tang et al. 2011). The evolutionary genetic distance and the maximum likelihood (ML) tree were estimated using PAUP* with the GTR + G + I model selected by AIC, implemented in Modeltest. The robustness of the ML tree was assessed using the BP with 500 replicates.

A Bayesian approach was used to estimate the phylogenetic tree for the long-sequence data set and the divergence times of lineages with the GTR + G + I models and the Yule (speciation) tree prior using BEAST v1.6.2 (Drummond and Rambaut 2007). We adopted the random local clock model, which assumes one or more independent rates on different branches (Drummond and Suchard 2010). To estimate the time of the most recent common ancestors (tMRCA), two constraints on the node ages were applied. First, the uplift of the Central Highland of Honshu Island in the Pliocene–early Pleistocene (Yonekura et al. 2001; Machida et al. 2006) is thought to have caused the divergence between two *Pseudorasbora pumila* subspecies (outgroup), which show a vicariant distribution in the eastern (*P. pumila pumila*) and western (*P. pumila* subsp.) areas across the highland (Watanabe et al. 2000). The highland, or the great valley (Fossa Magna) within the highland, represents one of the most important geographic barriers for freshwater fish fauna in Japan (see Watanabe 2010). The node of the MRCA of those subspecies was constrained following a lognormal prior distribution, ranging from approximately 2 to 5 mya [mean = 3.5 mya, log(SD) = 0.3, offset = 0]. We found a distinct lineage in the upper region of the

Tenryu River system (Ina Valley, Loc. # 2, 3; see "Results"). Therefore, as the second constraint, the isolation of the lineage is thought to have occurred with or preceded the uplift of the Kiso and Akaishi Mountains, which formed the valley in the middle to early Pleistocene (ca. 0.8 mya; Matsushima 1995; Moriyama 2001). The constraint was specified as an inverse-gamma prior distribution, with the shape parameter = 2, scale = 3, and offset = 0. Both of the prior distributions for the node ages involve a wide range. Therefore, they should act only as lax constraints for determining the tMRCA and give conservative results. All of the other model parameters used default priors. For each Markov Chain Monte Carlo (MCMC) analysis, we performed two independent runs of 50 million generations. We sampled every 1,000th generation and removed 10% of the initial samples as burn-in. The convergence of the chains to the stationary distribution was confirmed using Tracer v1.5 (Rambaut and Drummond 2009). The consensus tree was calculated by TreeAnnotator v.1.6.1 in the BEAST package, and the tree was visualized using FigTree v1.3.1 (Rambaut 2009).

To describe the genetic diversity of each population, the following indices were calculated, based on the short-sequence data set using ARLEQUIN 3.5 (Excoffier and Lischer 2010): the number of haplotypes (A), the haplotype diversity (h), and the nucleotide diversity (π). To estimate the demographic history of *G. caerulescens*, we applied a Bayesian skyline plot (BSP) analysis (Drummond et al. 2005), implemented in BEAST. We used the short-sequence data of *G. caerulescens* (n = 54) with several related haplotypes of *G. elongatus* (n = 5) as the outgroup, and performed two independent runs with an MCMC chain length of 50 million generations. We sampled every 1,000th generation and removed 10% of the initial samples as burn-in. The substitution model used was HKY + I, selected by Modeltest, and the time to expansion was estimated using the mutation rate obtained in the above Bayesian phylogenetic analysis with the long-sequence data [lognormal prior distribution, mean = 0.0183/Myr, log (SD) = 0.5, covering 0.0070–0.0368/Myr in the 95% range; see the Results]. The BSP result with the stepwise (constant) model was summarized using Tracer. In addition, we conducted neutrality tests by

calculating Tajima's *D* (Tajima 1989) and Fu's F_S (Fu 1997) for the same dataset to explore its demographic change, using ARLEQUIN 3.5. The significance for the estimates was tested by 10,000 permutations.

Results

Divergence of *Gnathopogon* and distribution

A total of 112 haplotypes of the short sequences were obtained from Japanese *Gnathopogon* fishes. The mtDNA phylogeny revealed two major lineages in these sequences, with substantial divergence between the two lineages [0.077 ± 0.004 (mean ± standard deviation) in uncorrected *p* distance, and 0.109 ± 0.009 in GTR + I + G distance for the 1,125-bp data set; Figs. 2, 3, 4; Table 2]. One lineage, with 72 haplotypes, included clade C (23 haplotypes) corresponding to *G. caerulescens*, and clade E1 (49 haplotypes) consisting of haplotypes primarily from the Lake Biwa area and the western ranges of *G. elongatus* (Fig. 3). The other major lineage, with 40 haplotypes, consisted of haplotypes obtained from the eastern populations of *G. elongatus* and was divided into two sub-lineages. One of these sub-lineages consisted of widely distributed haplotypes (E2; 33 haplotypes; Fig. 3). The distribution of haplotypes belonging to the other sub-lineage was restricted to the upper region of the Tenryu River (Ina Valley), flowing from Lake Suwa, central Honshu (E3; 7 haplotypes; Loc. #1, #2) (Table 1; Figs. 1, 3). Overall, the mtDNA phylogeny indicated that *G. elongatus* consists of paraphyletic lineages with allopatric distribution, one of which is more closely related to the limnetic species *G. caerulescens*.

Although these haplotype groups showed an essentially allopatric distribution, both the E1 and E2 haplotypes were found in the eastern side of Lake Biwa (Fig. 3). In this area, most of the non-lacustrine populations essentially possessed either E1 (1 of 9 populations) or E2 (7 of 9), with

one exception that showed both types (Loc. #24). The E2 haplotypes (the majority in this area) were identical to or very close to those detected in the Ise Bay area beyond the Suzuka Mountains. Certain populations with E2 haplotypes in the Lake Biwa area exhibited a low genetic diversity (Table 1) and were sporadically distributed in the network (closed circles in Fig. 2).

Some haplotypes exhibited irregular geographical distributions. For example, the haplotypes of clades C and E1 were found in Lake Suwa (Loc. #1; Fig. 3), which was consistent with the documented introductions of *G. caerulescens* stocks putatively from Lake Biwa into Lake Suwa (Kurasawa et al. 1981). A number of haplotypes, such as haplotypes e1-01 and e1-17 of clade E1, were detected from dispersed sites (Fig. 3) [see Electronic Supplementary Material (ESM) Appendix Table S1], another indication of their artificial distribution.

Genetic characteristics of limnetic forms

The clade C haplotypes were found almost exclusively from *G. caerulescens* in Lake Biwa (and Lake Suwa, via introduction). Exceptionally, two clade C haplotypes were detected in the *G. elongatus* populations around Lake Biwa at a low frequency (1.6%; 2 of 125 specimens). Conversely, a clade E2 haplotype (e2-01) was found in *G. caerulescens* (3.6%; 2 of 56). In contrast, another known extant lacustrine population from the Mikata Lakes possessed the non-monophyletic haplotypes included in clade E1 (star symbol in Fig. 2). The majority of the haplotypes are, however, relatively close to each other, except for haplotype e1-01, which is widely distributed.

As mentioned above, we did not find any unique haplotypes from Lake Suwa and its inlets, the type locality of the "extinct" *G. elongatus suwae*. However, haplotypes of the distinct clade (E3) were found exclusively from the tributaries of the outlet of the lake. In one of their two localities (Loc. #2), the clade E3 haplotypes co-occurred with the clade E1 haplotypes commonly found around Lake Biwa.

254

255 Divergence time

256

257 The number of changes in the substitution rate across the phylogeny was inferred to be 1.17 ± 0.02
258 times from the random local clock model. This value corresponded to a slightly slower rate in the
259 Japanese *Gnathopogon* clade (0.0164–0.0195/Myr) than in the other clades (0.0243–0.0251/Myr)
260 (Fig. 4; Table 2), but the difference was not drastic.

261 The tMRCA of the Japanese *Gnathopogon* populations was estimated at 4.01 Myr
262 [1.34–7.95 Myr, 95% highest posterior density (HPD)] (Fig. 4; Table 2). The tMRCA of the
263 lineage leading to clades C (*G. caerulescens*) and E1 was inferred as 1.68 Myr (0.47–3.53 Myr),
264 comparable with that of E2 and E3 (1.88 Myr; 0.62–3.83 Myr). These age estimates were smaller
265 than the tMRCA of *Pseudorasbora pumila* subspp., which was assumed to correspond to the Fossa
266 Magna vicariance, inferred as 2.53 Myr (1.28–4.01 Myr).

267 The tMRCA of *G. caerulescens* was estimated at 0.23 Myr (0.05–0.53 Myr, 95% HPD)
268 based on 54 short sequences. The BSP analysis indicated that the population expansion of this
269 limnetic species began 0.05 mya (Fig. 5). Neutrality tests also indicated a population expansion
270 (Tajima's $D = -1.75$, $p = 0.020$; Fu's $F_S = -7.99$, $p = 0.001$).

271

272

273 **Discussion**

274

275 Credibility of mutation rate and divergence time estimates

276

277 The mutation rate of the mtDNA cytochrome *b* gene for the Japanese *Gnathopogon* fishes was
278 estimated to be 0.016–0.025/Myr/lineage for GTR + I + G distances. This rate appears to be faster
279 than those in previous studies (0.003–0.015/Myr/lineage for cytochrome *b* in fishes; see Burridge

et al. 2008; Watanabe and Takahashi 2010). However, many of previous studies estimated mutation rates simply using the proportion of sequence differences (*p* distance), while we estimated them based on a molecular evolutionary model (GTR + I + G). Indeed, the mutation rates based on *p* distance were estimated for our data at 0.007–0.015/Myr/lineage (see Table 2 for the major clades), which agree with those from previous studies.

The credibility intervals of the tMRCA estimates were generally large because of the lax constraints used in dating the phylogeny. Also, our estimation of tMRCAs might be biased because it was based on single mtDNA gene sequences. However, because the phylogenic tree used for the analyses had high statistical support, we here consider that the estimates can be used as conservative values for a discussion of the population divergence and origin of limnetic forms in *Gnathopogon* fishes. The estimations need to be tested in the future with increased data, especially multilocus nuclear sequences, and with denser taxon sampling.

Cryptic differentiation within *Gnathopogon elongatus*

Monophyly of *G. elongatus* was not supported by our phylogenetic analyses. This species included two deeply diverged cryptic lineages, one of which is closer to *G. caerulescens* than to the other. The Suzuka Mountains roughly bounded the two lineages to the east and west. The Suzuka Mountains are known as one of the major geographical boundaries of freshwater fish fauna in Japan (Watanabe 1998, 2010), which started uplifting during the early Pleistocene (Yokoyama 1988; Kawabe 1994).

The eastern lineage was further divided into two allopatrically distributed sub-lineages, E2 and E3. Clade E2 was found across a widespread area, while E3 was restricted to the upper reaches of the Tenryu River in Ina Valley flowing from Lake Suwa. The Bayesian tMRCA analysis for E2 and E3 yielded an estimation of 1.88 Myr (0.62–3.83 Myr, 95% HPD), which tends to precede the uplift of the Kiso Mountains (~0.8 mya) used as a calibrating point. The wide

credibility interval may prevent ruling out the vicariance by the uplift of the Kiso Mountains, but the preceding geological events, such as the formation of Ina Valley (~2 mya; Machida et al. 2006), could have caused the divergence between E2 and E3.

The distribution range and genetic distinctness of E3 suggest that this mtDNA lineage may be related to the "extinct" *G. elongatus suwae*, which was the local representative in an area around Lake Suwa (Jordan and Hubbs 1925; Miyadi 1930). In other words, we may have discovered an unknown lineage of *G. elongatus* closely related to *G. elongatus suwae*, or rediscovered this subspecies itself. *G. elongatus suwae* was described from lacustrine populations; therefore, detailed morphological comparisons are necessary to determine the taxonomic status of the present populations from creeks in the Ina Valley.

We showed that *G. elongatus* is a paraphyletic species. In addition, the type locality of *G. elongatus* is unspecified (Temminck and Schlegel 1846). All three lineages (E1, E2, and E3) of this species should be taxonomically re-examined through detailed morphological comparisons, including inspection of the type series of this group.

Origins of limnetic forms

Adaptive divergence in an ancient lake is usually considered to begin with the invasion of a new habitat, followed by ecological adaptations to novel environments, and the derivation of new taxa from the ancestors (Martens 1997; Kornfield and Smith 2000; Kontula et al. 2003). This process of adaptive evolution has been hypothesized for the origin of some endemic species of Lake Biwa (e.g., Tomoda 1978; Tokui and Kawanabe 1984; Kawanabe 1996; Yuma et al. 1998). The endemic species of Lake Biwa are often divided into two categories, namely, "relic species" and "species evolved in the lake" (Kawanabe 1978, 1996). Particularly for the latter, their origins have been presumed to be the ancestral species occurring around the lake following adaptation to novel environments (e.g., the extensive pelagic area of the northern basin and the locally developed

rocky shores). Such environments developed after the middle Pleistocene (ca. 0.4 mya or later; Yokoyama 1984; Meyers et al. 1993); therefore, the species that evolved in the lake are believed to have originated in the same or later periods (e.g., Takahashi 1989). Indeed, the Lake Biwa endemic gudgeon, *Sarcocheilichthys*, exhibits clear trophic-resource polymorphism but shows no genetic divergence between morphs. These characteristics suggest a recent origin of the adaptive population divergence (Komiya et al. 2011).

However, our results suggest that such recent speciation does not hold for *G. caerulescens*. This species has been considered as a typical species that evolved in Lake Biwa from the riverine ancestor (*G. elongatus*) after the establishment of the present Lake Biwa (Hosoya 1987; Nakajima 1994; Kawanabe 1996) because *G. caerulescens* is specialized in feeding apparatus for planktivory (e.g., an upward-directed mouth and 13–20 gill rakers vs. subterminal mouth and 6–12 gill rakers in *G. elongatus*) and body shape for efficient swimming in open water (e.g., a low body depth and caudal peduncle; Hosoya 1987, 2000). However, the estimated tMRCA of *G. caerulescens* and E1 of *G. elongatus* indicated that their divergence dates to the early Pleistocene (1.68 Myr; 0.47–3.53 Myr, 95% HPD). Even with the wide credibility interval, it is unlikely that the *G. caerulescens* lineage derived at 0.4 mya or more recently. Molecular phylogenetic studies have also suggested an earlier origin (Late Pliocene) for the Lake Biwa pelagic goby, *Gymnogobius isaza* (Harada et al. 2002), which was similarly presumed to have evolved in the present Lake Biwa (Takahashi 1989; Kawanabe 1996).

In contrast, the tMRCA and BSP analyses focused on *G. caerulescens* suggested a more recent beginning of diversification in the present mtDNA lineage (0.23 mya) and a population expansion in the late Pleistocene (0.05 mya). These results agree well with the expected scenario in which *G. caerulescens* has thrived in the present environment of Lake Biwa. The adaptation to the limnetic lifestyle with the acquisition of specialized morphological features probably enabled its population expansion in the lake. It remains possible, however, that limnetic features had evolved in an extinct lake at the earlier stage of Paleo-Lake Biwa, and were retained as standing

variation in the populations having survived in rivers or marshes.

Our data clearly rejected the monophyletic origin of *G. caerulescens* and another lacustrine population in the Mikata Lakes. Most of the mtDNA haplotypes in the Mikata Lakes were endemic and close to each other, but were not monophyletic. The morphological specialization of the Mikata Lakes population to pelagic life is considered to be limited (Hosoya 1987). These findings suggest a short evolutionary history of the population in the lakes and/or confined adaptation to the less-developed pelagic environment in the lakes. These circumstances might have allowed gene flow with neighboring populations in their inlets. These hypotheses are supported by the geological history of the Mikata Lakes. The lakes have a relatively long history of at least 0.1 Myr (Takemura et al. 1994), but all the lakes, except one, are saline or brackish at present. Moreover, the freshwater lake has experienced seawater incursions during periods of high sea level because of their low altitude (0 m above sea level).

Our results and a previous report (Hosoya 2003) strongly suggest that *G. elongatus suwae* in Lake Suwa has been extirpated from the lake. The extinction of this population is considered to have resulted from habitat degradation and the hybridization with introduced *G. caerulescens* (and possibly *G. elongatus*) since 1925 (Kurasawa et al. 1981; Hosoya 1997, 2003; Takei 2007). Another known population of *G. elongatus suwae* from Lake Kizaki (60 km north of Lake Suwa) is also suggested to have become extinct through a similar process (Kohno et al. 2006). Lake Suwa was formed in the early (1.5–1.2 mya) or middle (0.2 mya) Pleistocene (see Machida et al. 2006). In this long-standing lake, *G. elongatus suwae* might have evolved adapting to the lacustrine environment as in other limnetic populations.

The present study provided phylogenetic evidence for the multiple origins of the limnetic forms of *Gnathopogon* fishes. Pelagic adaptation should have required a series of novel morphological, physiological, and ecological traits. In addition to the morphological variability of *G. elongatus*, which might serve as a preadaptation (Hosoya 1987), the variety furnished by the long-standing lineages might have contributed to the evolution of pelagic forms in this genus.

384

385 Natural and artificial hybridization

386

387 We found a low-level (~2%) of mtDNA introgression in both directions between *G. caerulescens*
388 and *G. elongatus*. Although they generally show a parapatric distribution in and around Lake Biwa,
389 they may have the chance to hybridize, because they produce fertile offspring and share spawning
390 sites (i.e. emergent plants at the lakeshore, lagoons and inlets; Nakamura 1969). Indeed, hybrid
391 offspring have been found near the spawning sites at low frequency (Kokita, unpublished data). A
392 hybrid disadvantage may serve to effectively prevent introgression between them in the natural
393 habitats, because their lifestyles (entirely pelagic vs. benthopelagic) substantially differ.
394 For several decades, *Gnathopogon* fishes have been intensively introduced to establish
395 fisheries (Nakamura 1969; Biodiversity Center of Japan 2002). Moreover, *G. elongatus* may have
396 been transplanted accidentally via contaminations to the stocks of, for example, the crucian carp
397 *Carassius cuvieri* and the common carp *Cyprinus carpio*, which are commonly stocked for fishery
398 and game fishing from ponds sometimes inhabited by *G. elongatus* (Okada and Nakamura 1948;
399 Yada 1977). Widespread introductions of *Gnathopogon* fishes may have affected the native fish
400 assemblages and native populations of *Gnathopogon* fishes. Some of the E1 haplotypes were
401 distributed widely from Lake Suwa to southwestern Shikoku Island. It is believed that the native
402 range of *G. elongatus* includes southwestern Shikoku Island (Hosoya 2001; Biodiversity Center of
403 Japan 2002). However, we found only a single widespread E1 haplotype in four localities in this
404 area. Similarly, on the eastern side of Lake Biwa, several E2 haplotypes were shared with
405 populations in the Ise Bay basin beyond the Suzuka Mountains. The presence of widespread
406 haplotypes that cross known biogeographic boundaries (Watanabe et al. 2010) strongly suggests
407 that *Gnathopogon* populations have been established in many localities out of their original ranges.
408 In addition, hybridization or replacement of the native *Gnathopogon* fish with introduced fish is
409 probable. *Gnathopogon caerulescens* and *G. elongatus* are known to form a hybrid swarm in a

nonnative habitat, despite their reproductive isolation in their native habitat (Sakai 1995). As mentioned above for *G. elongtus suwae*, artificial introductions would result in losses of endemic lineages and, hence, a reduction in the biodiversity of natural communities.

Acknowledgments

We are very grateful to T. Abe, T. Asaka, T. Karube, K. Kodama, T. Komiya, H. Kumada, T. Mukai, H. Ogawa, H. Sakai, M. Sugimura, N. Suzuki, and K. Tominaga for providing a portion of specimen, M. Nishida for lending us experimental instruments, and K. Hosoya for giving us valuable information. This study was partly supported by Grants-in-Aid from the Ministry of Education, Culture, Sports, Science and Technology of Japan (nos. 18570086, 21370035, and 2155282, and "Formation of a Strategic Base for Biodiversity and Evolutionary Research: from Genome to Ecosystem" of the GCOE).

References

Aoyama J, Watanabe S, Ishikawa S, Nishida M, Tsukamoto K (2000) Are morphological characters distinctive enough to discriminate between two species of freshwater eels, *Anguilla celebesensis* and *A. interioris*? Ichthyol Res 47:157–161

Bănărescu P, Nalbant TT (1973) Pisces, Teleostei, Cyprinidae (Gobioninae). Das Tierreich, Lieferung 93. Walter de Guryter, Berlin

Biodiversity Center of Japan (2002) The national survey on the natural environment report of the distributional survey of Japanese animals (freshwater fishes). Japan Wildlife Research Center, Tokyo (in Japanese)

Burridge, CP, Craw D, Fletcher D, Waters JM (2008) Geological dates and molecular rates: fish DNA sheds light on time dependency. Mol Biol Evol 25:624–633

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. Mol Ecol 9:1657–1660

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7:214

Drummond AJ, Suchard MA (2010) Bayesian random local clocks, or one rate to rule them all. BMC Biol 8:114

Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. Mol Biol Evol 22: 1185–1192

Eschmeyer WN, Fricke R. (2011) Catalog of Fishes electronic (version 5 May 2011). http://research.calacademy.org/ichthyology/catalog/fishcatmain.asp. Accessed 20 May 2011

Excoffier L, Lischer H E L (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10:564–567

Froese R, Pauly D (2011) FishBase. World Wide Web electronic publication.http://www.fishbase.org. version 2011/2. Accessed 20 May 2011

449 Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking,
450 and background selection. Genetics 147:915–925
451 Fujioka Y (2001) Thermolabile sex determination in honmoroko. J Fish Biol 59:851–861
452 Fujioka Y (2006) Patterns of sex ratio response to water temperature during sex determination in
453 honmoroko *Gnathopogon caerulescens*. Fish Sci 72:1034–1041
454 Harada S, Jeon SR, Kinoshita I, Tanaka M, Nishida M (2002) Phylogenetic relationships of four
455 species of floating gobies (*Gymnogobius*) as inferred from partial mitochondrial cytochrome *b*
456 gene sequences. Ichthyol Res 49:324–332
457 Hosoya K (1986) Interrelationships of the Gobioninae (Cyprinidae). In: Uyeno T, Arai R, Taniuchi
458 T, Matsuura K (eds) Indo-Pacific Fish Biology: Proceeding of the Second International
459 Conference on Indo-Pacific Fishes. Ichthyological Society of Japan, Tokyo, pp 484–501
460 Hosoya K (1987) Phylogeny and character displacement in *Gnathopogon* fishes. In: Mizuno N,
461 Goto A (eds) Freshwater fishes in Japan: their distribution, variation and speciation. Tokai
462 University Press, Tokyo, pp 31–40 (in Japanese)
463 Hosoya K (1997) The endangered Japanese freshwater fishes. In: Nagata Y, Hosoya K (eds)
464 Circumstances in endangered Japanese freshwater fishes and their protection. Midori Shobo,
465 Tokyo, pp 3–21 (in Japanese)
466 Hosoya K (2000) Cyprinidae. In: Nakabo T (ed) Fishes of Japan with pictorial keys to the species,
467 2nd edn. Tokai University Press, Tokyo, pp 253–271 (in Japanese)
468 Hosoya K (2001) *Gnathopogon*. In: Kawanabe H, Mizuno N, Hosoya K (eds) Freshwater fishes of
469 Japan 3rd edn Yama-Kei Publishers, Tokyo, pp 297–299 (in Japanese)
470 Hosoya K (2003) *Gnathopogon elongatus suwae*. In: Japan Ministry of the Environment (ed)
471 Threatened Wildlife of Japan, Red Data Book. 2nd ed. Japan Wildlife Research Center, Tokyo,
472 pp 26–27 (in Japanese)
473 Jordan DS, Fowler HW (1903) A review of the cyprinid fishes of Japan. Proc U S Natn Mus
474 26(1334):811–862

475 Jordan DS, Hubbs CL (1925) Record of fishes obtained by David Starr Jordan in Japan, 1922.
476 Mem Carneg Mus10:93–346
477 Kawabe T (1989) Stratigraphy of the lower part of the Kobiwako group around the Ueno Basin,
478 Kinki District, Japan. J Geoscience, Osaka City Univ 32:39–52
479 Kawabe T (1994) Chapter 1. Biwako no Oitachi (formation of Lake Biwa). In: Research Group for
480 Natural History of Lake Biwa (ed) Biwako no Shizenshi (The natural history of Lake Biwa).
481 Yasaka Shobo, Tokyo, pp 24–72 (in Japanese)
482 Kawanabe H (1978) Some biological problems. Verh Internat Ver Limnol 20:2674–2677
483 Kawanabe H (1996) Asian great lakes, especially Lake Biwa. Environ Biol Fish 47:219–234
484 Kohno N, Hosoe A, Ogawa S (2006) Species composition of fish caught by shore seine in Lake
485 Kizaki. Bull Nagano Pref Fish Exp Stn 8:35–38 (in Japanese)
486 Komiya T, Fujita S, Watanabe K (2011) A novel resource polymorphism in fish, driven by
487 differential bottom environments: an example from an ancient lake in Japan. PLoS ONE 6:
488 e17430
489 Kontula T, Kirilchik SV, Vainola R (2003) Endemic diversification of the monophyletic cottoid
490 fish species flock in Lake Baikal explored with mtDNA sequencing. Mol Phylogenetics Evol
491 58:142–147
492 Kornfield I, Smith PF (2000) African cichlid fishes: model systems for evolutionary biology.
493 Annu Rev Ecol Syst 31:163–96
494 Kottelat M, Freyhof J (2007) Gobioninae. In: Kottelat M, Freyhof J (ed) Handbook of European
495 Freshwater Fishes. Publications Kottelat, Cornol, pp 85–108.
496 Kurasawa H, Yamamoto M, Okino T (1981) Chronological changes of fish and mollusca faunae
497 and transplantation species in Lake Suwa. Ann Environ Sci Shinshu Univ 3:1–6 (in Japanese)
498 Machida H, Matsuda T, Umitsu M, Koizumi T (2006) Regional geomorphology of the Japanese
499 Islands, vol 5: Geomorphology of Chubu. University of Tokyo Press, Tokyo (in Japanese)
500 Martens K (1997) Speciation in ancient lakes. Trends Ecol Evol 12:177–182.

501 Matsushima S (1995) Morphogenetic history of the Ina basin. Res Rep Iida City Mus 3:1–145 (in
502 Japanese with English abstract)
503 Mayden RL, Chen WJ, Bart HL, Doosey MH, Simons AM, Tang KL, Wood RM, Agnew MK,
504 Yang L, Hirt MV, Clements MD, Saitoh K, Sado T, Miya M, Nishida M (2009) Reconstructing
505 the phylogenetic relationships of the earth's most diverse clade of freshwater fishes—order
506 Cypriniformes (Actinopterygii: Ostariophysi): A case study using multiple nuclear loci and
507 the mitochondrial genome. Mol Phylogenet Evol 51:500–514
508 Meyers PA, Takemura K, Horie S (1993) Reinterpretation of late Quaternary sediment chronology
509 of Lake Biwa, Japan, from correlation with marine glacial–interglacial cycles. Quat Res
510 39:154–162
511 Miyadi D (1930) Kizaki-Ko no gyorui ni tsuite (On fishes of Lake Kizaki). In: Tanaka A (ed)
512 Nippon Kita-Alps Kosho no Kenkyu (Studies on the lakes of Japanese Northern Alps).
513 Shinano Kyoiku-Kai Kitaazumi Bukai, Omachi, pp 626–630 (in Japanese)
514 Moriyama A (2001) Chronology of mountain formation in the Central Mountain region in Japan.
515 In: Yonekura N, Okada A, Moriyama A (eds) Hendou Chikeigaku (Techtonic
516 geomorphology). Kokinshoin, Tokyo, pp 87–109 (in Japanese)
517 Nakajima T (1994) Chapter 4-d. Cyprinid fishes. In: Research Group for Natural History of Lake
518 Biwa (ed) Biwako no Shizenshi (The natural history of Lake Biwa), Yasaka Shobo, Tokyo, pp
519 235–275 (in Japanese)
520 Nakamura M (1969) Cyprinid Fishes of Japan. Spec Publ Res Inst Nat Resour, Tokyo (in
521 Japanese)
522 Okada Y, Nakamura M (1948) Zoshoku (Aquaculture). In: Nippon no Tansui-Gyorui (Freshwater
523 fishes of Japan), Nippon Shuppan-sha, Osaka, pp 119–125 (in Japanese)
524 Palumbi S, Martin A, Romano S, McMillian WO, Stice L, Grabowski G (1991) The Simple Fool's
525 Guide to PCR. University of Hawaii, Honolulu
526 Posada D, Crandall KA (1998) Modeltest: testing the model of DNA substitutions. Bioinformatics

14:817–818

Rambaut A, Drummond AJ (2009) Tracer Version 1.5. http://tree.bio.ed.ac.uk/software/tracer/

Rambaut A (2009) FigTree Version 1.3.1. http://tree.bio.ed.ac.uk/software/figtree/

Saitoh K, Sado T, Mayden RL, Hanzawa N, Nakamura K, Nishida M, Miya M (2006) Mitogenomic evolution and interrelationships of the Cypriniformes (Actinopterygii: Ostariophysi): The first evidence towards resolution of higher-level relationships of the world. J Mol Evol 63:826–841

Saitoh K, Sado T, Doosey MH, Bart Jr HL, Inoue JG, Nishida M, Mayden RL, Miya M (2011) Evidence from mitochondrial genomics supports the lower Mesozoic of South Asia as the time and place of basal divergence of cypriniform fishes (Actinopterygii: Ostariophysi). Zool J Linn Soc 161:633–662

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Sakai H (1995) Natural hybrid and speciation in fish. Biol Sci 47:113–123 (In Japanese)

Schluter D, McPhail, JD (1992) Ecological character displacement and speciation in sticklebacks. Am Nat 140:85–108

Schluter D (1998) Ecological speciation in postglacial fishes. In: Grant PR (ed) Evolution on islands. Oxford University Press, Oxford

Snorrason SS, Skúlason S, Jonsson B, Malmquist HJ, Jónasson PM, Sandlund OT, Lindem T (1992) Trophic specialization in Arctic charr *Salvelinus alpinus* (Pisces; Salmonidae): morphological divergence and ontogenetic niche shifts. Biol J Linn Soc 52:1–18

Swofford DL (2002) PAUP*: phylogenetic analysis using parsimony (*and other methods), ver 4. Sinauer Associates, Sunderland

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595

Takahashi S (1989) A review of the origins of endemic species in Lake Biwa with special reference

553 to the goby fish, *Chaenogobius isaza*. J Paleolimnology 1:279–292
554 Takei K (2007) Verified the list of the fishes of Lake Suwa. Bull Nagano Pref Fish Exp Stn 9:7–21
555 (in Japanese)
556 Takemura K, Kitagawa H, Hayashida A, Yasuda Y (1994) Sedimentary facies and chronology of
557 core samples from Lake Mikata, Lake Suigetsu and Kurota Lowland, central
558 Japan—sedimentary environment in Mikata Lowland since the last interglacial time. J
559 Geography 103:233–242
560 Tang KL, Agnew MK, Chen WJ, Vincent Hirt M, Raley ME, Sado T, Schneider LM, Yang L, Bart
561 HL, He S, Liu H, Miya M, Saitoh K, Simons AM, Wood RM, Mayden RL (2011) Phylogeny
562 of the gudgeons (Teleostei: Cyprinidae: Gobioninae). Mol Phylogenet Evol 61:103–124
563 Temminck CJ, Schlegel H (1846) Pisces. Fauna Japonica, sive descriptio animalium quae in
564 itinere per Japoniam suscepto annis 1823–30 collegit, notis observationibus et
565 adumbrationibus illustravit P. F. de Siebold. Parts 10–14:173–269
566 Tokui T, Kawanabe H (1984) Fishes. In: Horie S (ed) Lake Biwa, Monographiae Biologicae
567 (volume 54). Dr W Junk Publishers, Dordrecht, pp 339–360
568 Tomoda Y (1978) Biwako to Namazu (Lake Biwa and catfish). Chobunsha, Tokyo (in Japanese)
569 Watanabe K (1998) Parsimony analysis of the distribution pattern of Japanese primary freshwater
570 fishes, and its application to the distribution of the bagrid catfishes. Ichthyol Res 45:259–270
571 Watanabe K (2010) Faunal structure of Japanese freshwater fishes and its artificial disturbance.
572 Environ Biol Fish. Doi:10.1007/s10641-010-9601-5
573 Watanabe K, Takahashi H (2010) Tansuigyorui chiri no shizenshi (Natural history of freshwater
574 fish geography). Hokkaido University Press, Sapporo (in Japanese)
575 Watanabe K, Iguchi K, Hosoya K, Nishida M (2000) Phylogenetic relationships of the Japanese
576 minnows, *Pseudorasbora* (Cyprinidae), as inferred from mitochondrial 16S rRNA gene
577 sequences. Ichthyol Res 47:43–50
578 Watanabe K, Kano Y, Takahashi H, Mukai T, Kakioka R, Tominaga K (2010) GEDIMAP: a

579 database of genetic diversity for Japanese freshwater fishes. Ichthyol Res 57:107–109
580 Winfield IJ, Nelson JS (1991) Cyprinid Fishes: Systematics, biology and exploitation. Chapman &
581 Hall, London.
582 Yada T (1977) Studies on the spawning period and number of egg spawned on “Tamoroko”,
583 *Gnathopogon elongatus elongatus*. Bull Osaka Pref Freshwater Fish Exp Stn 5:1–8 (In
584 Japanese)
585 Yang JQ, He SP, Freyhof J, Witte K, Liu HZ (2006) The phylogenetic relationships of the
586 gobioninae (Teleostei: Cyprinidae) inferred from mitochondrial cytochrome *b* gene sequences.
587 Hydrobiologia 553:255–266
588 Yokoyama T (1984) Stratigraphy of the Quaternary system around Lake Biwa and geohistory of
589 the ancient Lake Biwa. In: Horie S (ed) Lake Biwa, Monographiae Biologicae (volume 54).
590 Dr W Junk Publishers, Dordrecht, pp 43–128
591 Yokoyama T (1988) Seinan Nihon no Shizenshi (Natural history of southwestern Japan).
592 Sanwa-shobo, Kyoto (in Japanese)
593 Yonekura N, Kaizuka S, Nogami M, Chinzai K (2001) Regional geomorphology of the Japanese
594 Islands, vol 1: Introduction to Japanese geomorphology. University of Tokyo Press, Tokyo (in
595 Japanese)
596 Yuma H, Hosoya K, Nagata Y (1998) Distribution of the freshwater fishes of Japan: an historical
597 overview. Environ Biol Fish 52:97–124
598

Figure legends

Fig. 1 Sampling localities for *Gnathopogon* fishes. *Numbers* correspond to those in Table 1. *Larger ellipses* indicate the inclusion of several neighboring sites

Fig. 2 Unrooted neighbor-joining (NJ) tree of Japanese *Gnathopogon* fishes based on the 3′-half of mtDNA cytochrome *b* sequences (598 bp). The evolutionary distance is based on the TIM + G model selected by AIC, with base frequencies of A = 0.292, C = 0.149, G = 0.284, and T = 0.275, a substitution rate matrix of A ↔ C = 1.000, A ↔ G = 19.846, A ↔ T and C ↔ G = 2.182, and C ↔ T = 34.975, and a gamma shape = 0.263. The region where each haplotype was detected is shown by a *different symbol*. *Numbers at nodes* indicate NJ bootstrap probabilities (values <70% not shown)

Fig. 3 Geographic distributions and statistically parsimonious networks for the haplotypes of each *Gnathopogon* lineage in Japan. Areas of nodes in the networks are proportional to haplotype frequency; *different patterns* indicate geographic origins of a haplotype. *Filled squares* indicate unobserved hypothetical haplotypes. The sampling site where each lineage was detected is shown by a *symbol* according to a geographic region

Fig. 4 Bayesian phylogenetic tree of the Japanese *Gnathopogon* fishes with selected continental species and outgroup based on the mtDNA cytochrome *b* sequences (1,125 bp) with the GTR + I + G model. The tree is dated by the random local clock model with two node-age constraints (C1 and C2), the prior distributions of which are shown in the *upper left panels*. The *numbers at nodes* correspond to Bayesian posterior probabilities on the left and ML bootstrap probabilities on the right (values <70% not shown). The *numbers in brackets under the internodes* indicate the estimated mutation rates/Myr. *Bars* show credibility intervals as 95% HPD

Fig. 5 The Bayesian skyline plot for *Gnathopogon caerulescens* based on the HKY + I model. The *central bold line* represents the median value for the relative effective female population size, and the *narrow line* denotes the 95% upper and lower credibility limits (95% HPD)

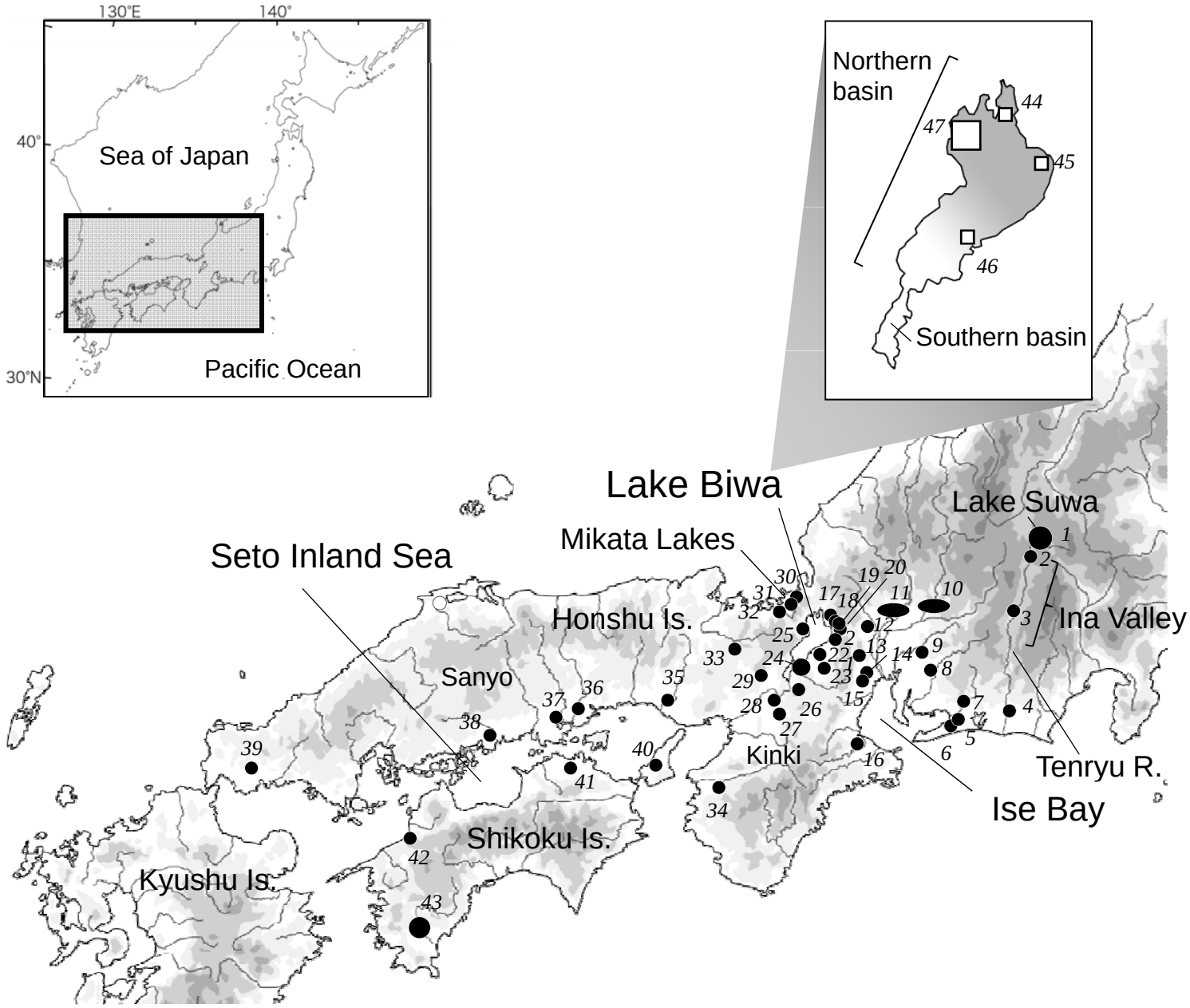
130°E
140°
40°
30°N
Sea of Japan
Pacific Ocean
Northern basin
Southern basin
44
47
45
46
Lake Biwa
Lake Suwa
Mikata Lakes
Seto Inland Sea
Honshu Is.
Ina Valley
Sanyo
Kinki
Tenryu R.
Ise Bay
Shikoku Is.
Kyushu Is.
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43

Fig. 1

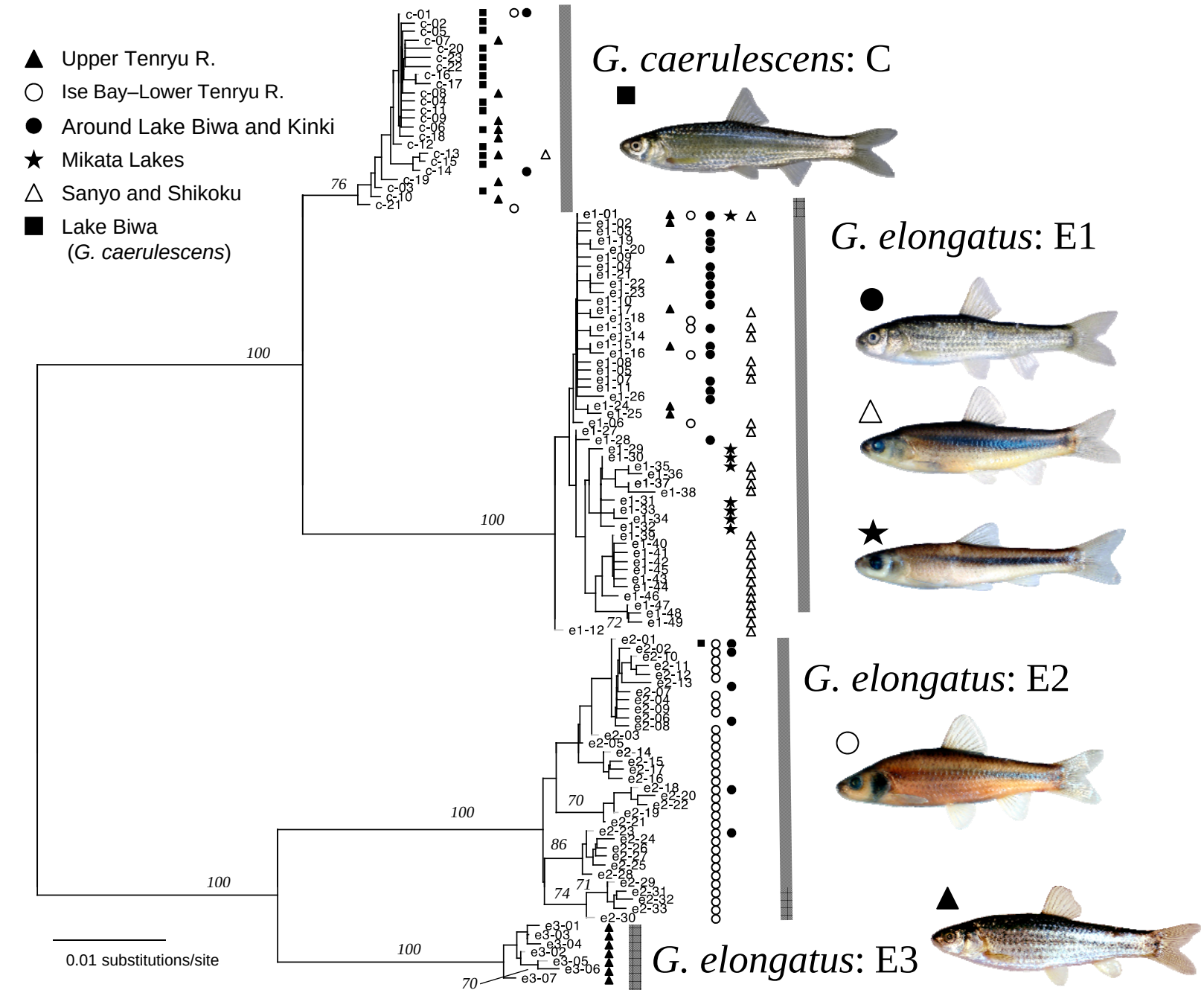
Upper Tenryu R.
Ise Bay–Lower Tenryu R.
Around Lake Biwa and Kinki
Mikata Lakes
Sanyo and Shikoku
Lake Biwa
(G. caerulescens)
G. caerulescens: C
G. elongatus: E1
G. elongatus: E2
G. elongatus: E3
0.01 substitutions/site

Fig. 2

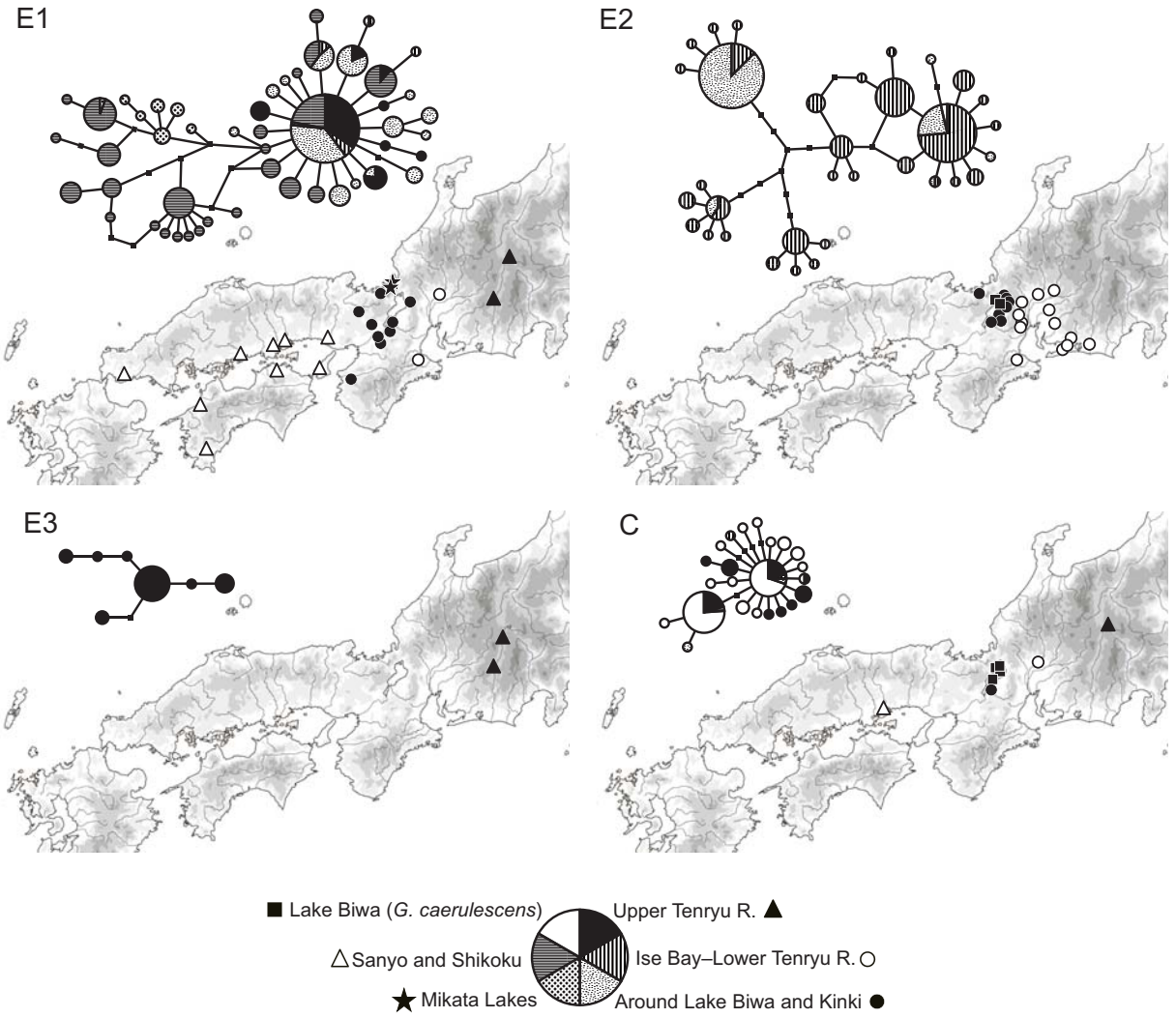
E1
E2
E3
C
■ Lake Biwa (G. caerulescens)
△ Sanyo and Shikoku
★ Mikata Lakes
Upper Tenryu R. ▲
Ise Bay–Lower Tenryu R. ○
Around Lake Biwa and Kinki ●

Fig. 3

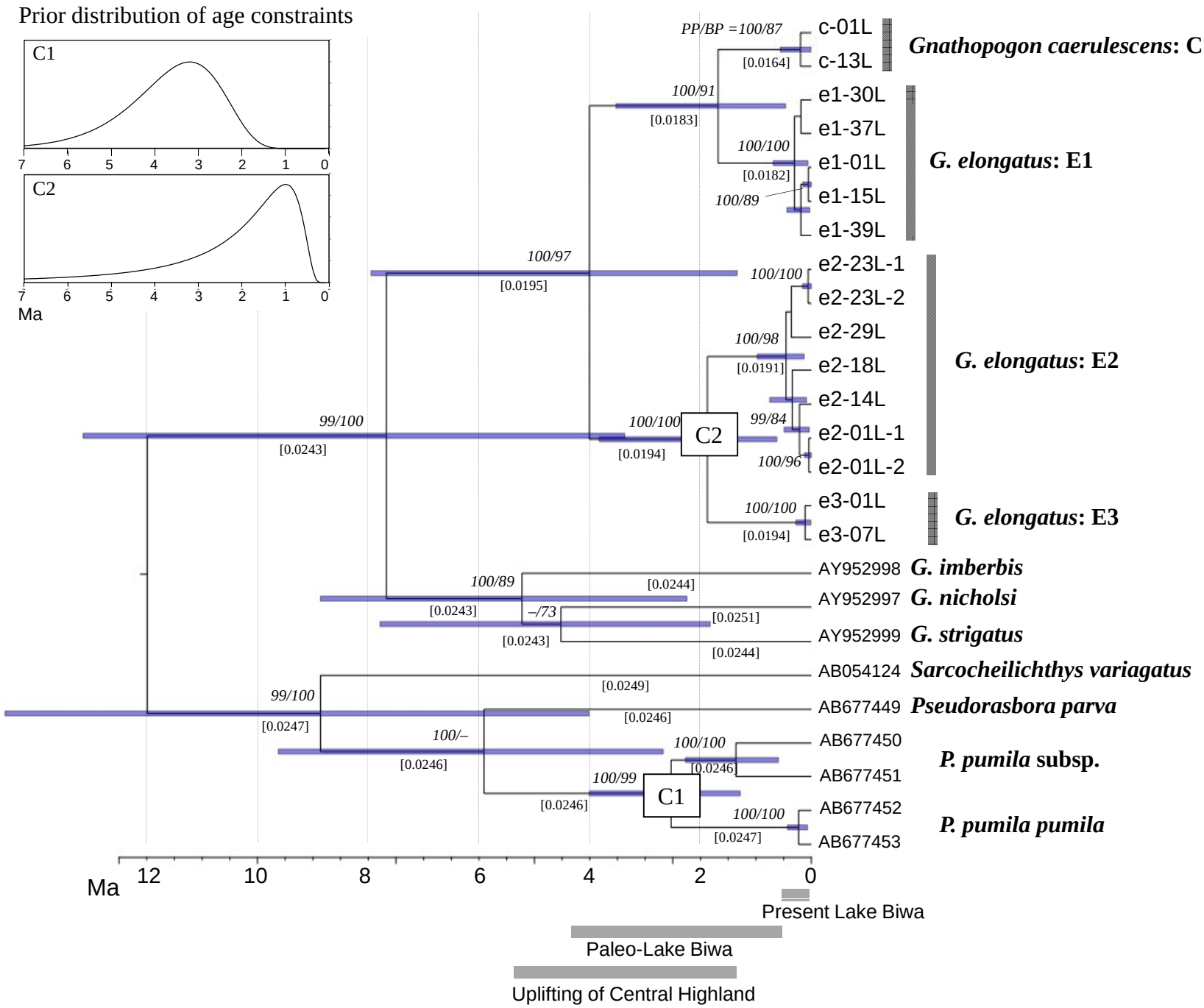
Prior distribution of age constraints
C1
C2
7 6 5 4 3 2 1 0
Ma
PP/BP =100/87
c-01L
c-13L
[0.0164]
Gnathopogon caerulescens: C
100/91
[0.0183]
e1-30L
e1-37L
100/100
[0.0182]
e1-01L
100/89
e1-15L
e1-39L
G. elongatus: E1
100/97
[0.0195]
e2-23L-1
100/100
e2-23L-2
e2-29L
100/98
[0.0191]
e2-18L
e2-14L
99/100
[0.0243]
100/100
[0.0194]
C2
99/84
e2-01L-1
100/96
e2-01L-2
G. elongatus: E2
100/100
e3-01L
[0.0194]
e3-07L
G. elongatus: E3
AY952998 G. imberbis
100/89
[0.0244]
[0.0243]
AY952997 G. nicholsi
–/73
[0.0251]
[0.0243]
AY952999 G. strigatus
[0.0244]
AB054124 Sarcocheilichthys variagatus
[0.0249]
99/100
[0.0247]
AB677449 Pseudorasbora parva
[0.0246]
100/–
[0.0246]
100/100
AB677450
[0.0246]
AB677451
P. pumila subsp.
100/99
[0.0246]
C1
AB677452
100/100
[0.0247]
AB677453
P. pumila pumila
Ma
12 10 8 6 4 2 0
Present Lake Biwa
Paleo-Lake Biwa
Uplifting of Central Highland

Fig. 4

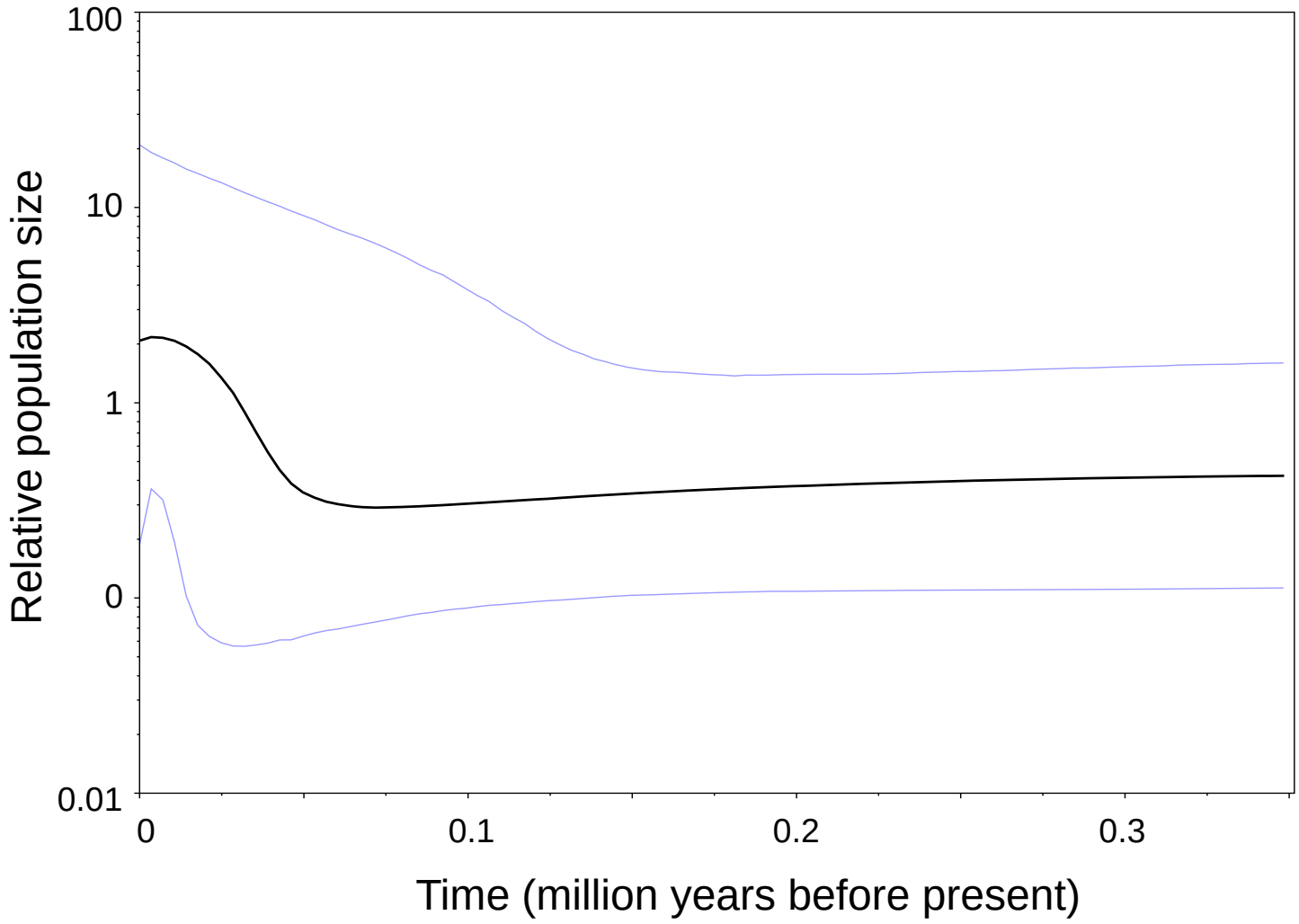
100
10
1
0
0.01
Relative population size
0
0.1
0.2
0.3
Time (million years before present)

Fig. 5

Table 1. Locality, number of specimens (*n*), and genetic diversity indices of *Gnathopogon* populations examined

Regional group	Population code	River; river/lake system	Locality	*n*	mtDNA groups				*k*	*h*	π
					E1	E2	E3	C			
Gnathopogon elongatus											
Nagano											
	1	Lake Suwa and its inlets	Suwa, Nagano Prefecture	63	41*	–	–	22*	17	0.811	0.021
	2	Creek; upper Tenryu River	Kamiina, Nagano Pref.	16	–	–	16	–	3	0.433	0.002
	3	Creek; upper Tenryu River	Iida, Nagano Pref.	15	5*	–	10	–	8	0.886	0.042
Shizuoka											
	4	Ichiunsai R.; lower Tenryu R.	Iwata, Shizuoka Pref.	13	–	13	–	–	5	0.628	0.001
Ise Bay											
	5	A creek; Umeda R.	Toyohashi, Aichi Pref.	7	–	7	–	–	3	0.667	0.001
	6	Kamida R.	Toyohashi, Aichi Pref.	4	–	4	–	–	2	0.667	0.006
	7	Toyo R.	Shinshiro, Aichi Pref.	2	–	2	–	–	2	1.000	0.002
	8	Yashita R., Yahagi-furu R.; Yahagi R.	Toyota and Nishio, Aichi Pref.	22	–	22	–	–	5	0.338	0.001
	9	Ponds; Shonai River	Nagoya and Nagakute, Aichi Pref.	19	–	19	–	–	6	0.708	0.010
	10	Ponds and streams; Kiso R.	Sofue and Ichinomiya, Aichi Pref.; Hashima, Minokamo and Yaotsu, Gifu Pref.	23	–	23	–	–	7	0.712	0.008
	11	Ponds and streams; Nagara R.	Ijira and Kaizu, Gifu Pref.	8	5	1	–	2*	5	0.786	0.037
	12	Creeks; Ibi R.	Yoro and Ogaki, Gifu Pref.	17	–	17	–	–	5	0.757	0.008
	13	A pond; Inabe R.	Inabe, Mie Pref.	3	–	3	–	–	2	0.667	0.001
	14	Kaizo R.	Yokkaichi, Mie Pref.	3	–	3	–	–	2	0.667	0.011
	15	Kabake R.; Tenpaku R.	Yokkaichi, Mie Pref.	4	–	4	–	–	2	0.500	0.008
	16	Kushida R. and Harai R.; Kushida R.	Matsusaka, Mie Pref.	14	4	10	–	–	6	0.681	0.043
Around Lake Biwa											
	17	Yogo R.	Takatsuki, Shiga Pref.	2	–	2	–	–	2	1.000	0.002
	18	Kawamichi R.	Nagahama, Shiga Pref.	5	–	5	–	–	3	0.700	0.010
	19	A pond	Nagahama, Shiga Pref.	8	8	–	–	–	4	0.821	0.002
	20	Nagahama-shinsen R.	Nagahama, Shiga Pref.	8	–	8	–	–	2	0.536	0.007
	21	Anjiki R.	Hikone, Shiga Pref.	14	–	14	–	–	1	0.000	0.000
	22	Daidoh R.	Notogawa, Shiga Pref.	12	–	12	–	–	1	0.000	0.000
	23	Hino R.	Hino, Shiga Pref.	16	–	16	–	–	2	0.325	0.004
	24	Creeks	Ritto, Moriyama, and Kusatsu, Shiga Pref.	22	16	4	–	2	6	0.788	0.036
	25	Creeks	Adogawa, Shiga	13	–	13	–	–	4	0.423	0.006

			Pref.									
Yodo River system												
	26	Daido R.	Otsu, Shiga Pref.	8	8	–	–	–	3	0.679	0.002
	27	Fugenji R.	Kyotanabe, Kyoto Pref.	10	10	–	–	–	2	0.200	0.000
	28	Nunome R.; Kizu R.	Kasagi, Kyoto Pref.	6	6	–	–	–	2	0.600	0.001
	29	Creek; Hozu R.	Kameoka, Kyoto Pref.	15	15	–	–	–	7	0.838	0.003
Northern Kinki												
	30	Mikata Lakes	Mikata, Fukui Pref.	8	8	–	–	–	5	0.857	0.003
	31	Hasu R.; Mikata L.	Mikata, Fukui Pref.	4	4	–	–	–	3	0.833	0.004
	32	Kita R.	Obama, Fukui Pref.	8	0	8	–	–	2	0.536	0.007
	33	Takaya R.; Yura R.	Mizuho, Kyoto Pref.	3	3	–	–	–	1	0.000	0.000
Southern Kinki												
	34	Creeks; Kinokawa R.	Katsuragi, Wakayama Pref.	12	12	–	–	–	4	0.455	0.001
Sanyo												
	35	Kako R.	Kakogawa, Hyogo Pref.	12	12	–	–	–	8	0.924	0.008
	36	Uryu R.; Yoshii R.	Seto, Okayama Pref.	3	2	–	–	1*	2	0.667	0.031
	37	Sasagase R.	Okayama, Okayama Pref.	15	15	–	–	–	9	0.905	0.006
	38	Takaya R.; Ashida R.	Fukuyama, Hiroshima Pref.	10	10	–	–	–	5	0.756	0.004
	39	Ono Reservoir, Koto R.	Ube, Yamaguchi Pref.	14	14*	–	–	–	3	0.473	0.001
Awaji Island												
	40	Shitoori R.; Mihara R.	Minami-awaji, Hyogo Pref.	16	16	–	–	–	4	0.758	0.006
Eastern Shikoku												
	41	Honzu R.	Takamatsu, Kagawa Pref.	16	16	–	–	–	4	0.525	0.005
	42	Otani R.	Iyo, Ehime Pref.	5	5	–	–	–	3	0.700	0.006
Southern Shikoku												
	43	Ushiro R., Uchigawa R., Mima R., and a pond; Shimanto R.	Shimanto, Kochi Pref., and Uwajima, Ehime Pref.	15	15*	–	–	–	1	0.000	0.000
Gnathopogon caerulescens												
Lake Biwa (LBW)												
	44	Lake Biwa	Kohoku, Shiga Pref.	8	–	–	–	8	3	0.607	0.003
	45	Lake Biwa	Hikone, Shiga Pref.	3	–	1	–	2	3	1.000	0.051
	46	Lake Biwa	Omihachiman, Shiga Pref.	14	–	–	–	14	7	0.857	0.003
	47	Lake Biwa	Oura and Imazu, Shiga Pref.	31	–	1	–	30	10	0.753	0.008

*Haplotypes highly probably originated from artificially introduced fish (see text)
k Number of haplotypes, *h* haplotype diversity, π nucleotide diversity

Table 2. Genetic distances and estimated divergence time between major lineages of Japanese *Gnathopogon* species based on 1,125-bp mtDNA cytochrome *b* sequences

	C + E1 vs. E2 + E3	C vs. E1	E2 vs. E3[a]	*Pseudorasbora pumila* subspp.[b]
p distance	0.0771 ± 0.0042	0.0335 ± 0.0020	0.0471 ± 0.0010	0.0742 ± 0.0017
GTR + G + I distance	0.1089 ± 0.0086	0.0369 ± 0.0021	0.0551 ± 0.0014	0.0988 ± 0.0033
tMRCA (Myr)	4.01 ± 0.10 [1.34, 7.95]	1.68 ± 0.04 [0.47, 3.53]	1.88 ± 0.05 [0.62, 3.83]	2.53 ± 0.01 [1.28, 4.01]
Mean clock rate (/Myr)	0.0195	0.0183	0.0194	0.0246
p distance/Myr/lineage	0.0096	0.0100	0.0125	0.0147

Data are shown as mean ± standard deviation
In brackets, 95% confidence interval (highest posterior density) is shown
[a] The node was used as calibration point (C2 in Fig. 3)
[b] The node was used as calibration point (C1 in Fig. 3)

Appendix Table S1. Haplotypes and their frequencies of *Gnathopogon* populations examined

Population code	Haplotype (frequency)	GEDIMAP[a] population ID
1	c-1 (5), c-6 (1), c-7 (3), c-8 (1), c-9 (1), c-10 (1), c-13 (6), c-18 (3), c-19 (1), e1-1 (26), e1-2 (4), e1-9 (4), e1-15 (2), e1-17 (2), e1-19 (1), e1-24 (1), e1-25 (1)	P1382
2	e3-1 (12), e3-6 (2), e3-7 (2)	P1383
3	e1-1 (1), e1-2 (3), e1-15 (1), e3-1 (3), e3-2 (1), e3-3 (1), e3-4 (4), e3-5 (1)	P1384
4	e2-29 (8), e2-30 (1), e2-31 (1), e2-32 (1), e2-33 (2)	P1385
5	e2-14 (2), e2-15 (1), e2-16 (4)	P1386
6	e2-14 (2), e2-18 (2)	P1387
7	e2-14 (1), e2-17 (1)	P1388
8	e2-8 (1), e2-10 (18), e2-11 (1), e2-12 (1), e2-14 (1)	P1389
9	e2-1 (10), e2-18 (2), e2-22 (1), e2-23 (2), e2-26 (3), e2-28 (1)	P1390
10	e2-1 (12), e2-4 (3), e2-7 (1), e2-18 (3), e2-20 (1), e2-23 (2),	P1391
11	c-1 (1), c-21 (1), e1-1 (4), e1-6 (1), e2-19 (1)	P1392
12	e2-1 (7), e2-9 (5), e2-18 (2), e2-21 (1), e2-23 (2)	P1393
13	e2-1 (1), e2-5 (2)	P1394
14	e2-1 (1), e2-24 (2)	P1395
15	e2-1 (3), e2-25 (1)	P1396
16	e2-1 (8), e2-2 (1), e2-3 (1), e1-13 (2), e1-16 (1), e1-18 (1)	P1397
17	e2-1 (1), e2-6 (1)	P1398
18	e2-1 (3), e2-18 (1), e2-23 (1)	P1399
19	e1-1 (1), e1-4 (2), e1-7 (3), e1-15 (2)	P1400
20	e2-1 (5), e2-18 (3)	P1401
21	e2-18 (14)	P1402
22	e2-18 (12)	P1403
23	e2-18 (13), e2-23 (3)	P1404
24	c-1 (1), c-14 (1), e1-1 (7), e1-2 (2), e1-13 (7), e2-18 (4)	P1405
25	e2-1 (1), e2-2 (1), e2-13 (1), e2-18 (10)	P1406
26	e1-1 (4), e1-15 (1), e1-22 (3)	P1407
27	e1-1 (1), e1-15 (9)	P1408
28	e1-1 (3), e1-4 (3)	P1409
29	e1-1 (4), e1-10 (2), e1-11 (1), e1-16 (1), e1-20 (1), e1-21 (5),	P1410
30	e1-1 (1), e1-30 (3), e1-33 (1), e1-31 (2), e1-35 (1)	P1411
31	e1-29 (1), e1-32 (2), e1-34 (1)	P1412
32	e2-1 (3), e2-18 (5)	P1413
33	e1-26 (3)	P1414
34	e1-1 (9), e1-3 (1), e1-15 (1), e1-28 (1)	P1415
35	e1-1 (1), e1-8 (1), e1-35 (3), e1-36 (1), e1-39 (2), e1-40 (1), e1-47 (2), e1-48 (1)	P1416
36	c-13 (1), e1-37 (2)	P1417
37	e1-1 (3), e1-6 (1), e1-27 (1), e1-37 (2), e1-38 (1), e1-39 (4), e1-41 (1), e1-42 (1), e1-43 (1)	P1418
38	e1-37 (2), e1-39 (5), e1-44 (1), e1-45 (1), e1-46 (1)	P1419
39	e1-1 (10), e1-5 (3), e1-7 (1)	P1420
40	e1-6 (3), e1-13 (6), e1-14 (2), e1-49 (5)	P1421
41	e1-1 (2), e1-8 (1), e1-35 (11), e1-47 (2)	P1422
42	e1-1 (1), e1-12 (1), e1-35 (3)	P1423
43	e1-17 (15)	P1424
44	c-1 (2), c-13 (5), c-23 (1)	P1425
45	c-1 (1), c-2 (1), e2-1 (1)	P1426
46	c-1 (4), c-3 (1), c-6 (1), c-12 (2), c-13 (4), c-15 (1), c-16 (1)	P1427
47	c-1 (9), c-3 (1), c-4 (2), c-5 (1), c-11 (1), c-13 (13), c-17 (1), c-20 (1), c-22 (1), e2-1 (1)	P1428

[a]http://gedimap.zool.kyoto-u.ac.jp

Population codes correspond to those shown in Table 1 and Fig. 1

Haplotypes begin with c are those of *G. caerulescens* ; haplotypes begin with e1–e3 are those of the E1–E3 clades of *G. elongatus*

Frequencies for each haplotype are shown in parentheses

Sequences of the haplotypes were deposited in DDBJ/EMBL/GenBank (accession numbers AB677321–AB677440)