

RESEARCH

Open Access



Next-generation phylogeography reveals unanticipated population history and climate and human impacts on the endangered floodplain bitterling (*Acheilognathus longipinnis*)

Keisuke Onuki^{1*} , Ryosuke K. Ito^{1,9}, Tappei Mishina^{1,2}, Yasuyuki Hashiguchi³, Koki Ikeya⁴, Kazuhiko Uehara⁵, Masaki Nishio⁶, Ryoichi Tabata⁷, Seiichi Mori⁸ and Katsutoshi Watanabe^{1*}

Abstract

Background Floodplains harbor highly biodiverse ecosystems, which have been strongly affected by both past climate change and by recent human activities, resulting in a high prevalence of many endangered species in these habitats. Understanding the history of floodplain species over a wide range of timescales can contribute to effective conservation planning. We reconstructed the population formation history of the Itasenpara bitterling *Acheilognathus longipinnis*, an endangered floodplain fish species in Japan, over a broad timescale based on phylogenetic analysis, demographic modeling, and historical demographic analysis using mitogenome and whole-genome sequences. A genome sequence was newly assembled as a reference for the resequencing analysis. This bitterling is distributed in three plains separated by high mountain ranges and exhibits ecological characteristics well adapted to floodplain environments.

Results Our analyses revealed an unexpected population branching pattern, gene flow, and timing of the differentiation that occurred within a few hundred thousand years, i.e., long after the mountain uplift that was assumed to be the primary geological cause of the population differentiation. The analyses also showed that all local populations experienced a severe decline during the last glacial and post-glacial periods.

Conclusions Our results suggest that the floodplain bitterling was able to disperse through unknown routes after mountain uplift and that its populations were strongly influenced by climatic and geographic changes in glacial–interglacial cycles and subsequent human activities, probably related to its floodplain-dependent ecology. The genomic data highlight the unanticipated distribution process of this species and the magnitude of the impact of human activities, with important implications for its conservation.

*Correspondence:

Keisuke Onuki
fwfonk0238@terra.zool.kyoto-u.ac.jp
Katsutoshi Watanabe
watanak@terra.zool.kyoto-u.ac.jp

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Keywords Next-generation phylogeography, Floodplain, Historical demography, Freshwater fish, Endangered species, Whole-genome resequencing

Background

Floodplains are one of the most important ecosystems that temporarily form on the plains around rivers and lakes and often harbor high biodiversity in their spatio-temporally heterogeneous environments, with seasonal and reciprocal transitions between terrestrial and aquatic habitats [1–3]. Floodplain ecosystems are more likely to develop in humid climates and are therefore influenced by geohistorical conditions such as climatic changes. In addition, floodplain environments were modified for human agricultural and residential use during the Holocene, followed by drastic alterations in the Anthropocene for flood control, land reclamation, and urbanization [1, 4, 5]. Human-induced environmental changes have endangered many wildlife species living in floodplains [4, 5]. To evaluate such human impacts on floodplain environments and biodiversity, compared to the impacts of geohistorical environmental changes, understanding the long-term population history of floodplain inhabitants is important. However, estimating population history over various timescales is methodologically challenging.

Next-generation phylogeography can be used to address this challenge. Access to genome-wide single nucleotide polymorphism (SNP) information using next-generation sequencing technology has enabled robust phylogenetic and demographic inferences using a large number of loci, considering gene flow, which cannot be adequately examined by a few conventional genetic markers [6–9]. The population history of target organisms can now be inferred over a wide range of timescales (several to hundreds of thousands of generations) with unprecedented resolution using whole-genome data [10]. Studies on the origin and dynamics of populations of floodplain organisms under geoclimatic and anthropogenic influences will benefit greatly from next-generation phylogeographic approaches.

Here, we focus on a typical floodplain species, the Itasenpara bitterling, *Acheilognathus longipinnis* (Acheilognathinae; Cyprinidae), a strictly freshwater fish endemic to Japan. We aimed to reconstruct the past evolutionary history of local populations of this species using whole-genome population analyses. This species has a disjunct distribution in three plains, clearly separated by mountain ranges: the Toyama, Nobi, and Osaka Plains [11, 12] (Fig. 1). This species is listed as endangered by the International Union for Conservation of Nature (IUCN) and local red lists [13, 14], being threatened in all plains by the deterioration of their floodplain habitats, which they depend on for growth during the juvenile phase [15]. Like other bitterlings, this species spawns in the mantle of

unionoid bivalves, and early development occurs within the shell [11, 12]. While most bitterling species spawn in spring or early summer, and the juveniles emerge from the bivalve within a few weeks, this species spawns in autumn, and the hatchlings stop developing in the shell to overwinter; they resume development in spring and emerge in early summer [16, 17]. The biology and life cycle of this fish species are highly adapted to the floodplain environment in the East Asian monsoon climate zone, where juveniles can utilize the abundant temporary water formed on the floodplain by spring-to-summer rainfall before other species begin to use them [15, 18, 19].

The genetic population structure and demographic history of *A. longipinnis* had been investigated in previous studies due to its biogeographical interests and conservation needs. Analyses of partial mitochondrial DNA sequences and microsatellite DNA polymorphisms revealed that each of the three local populations (Toyama, Nobi, and Osaka) of this species formed a homogeneous genetic cluster [20–22]. However, reliable results on the order and timing of the divergence of the three local populations have not been obtained so far. Okazaki et al. [21] and Kitanishi et al. [20] found a close relationship between the Nobi and Osaka populations using mitochondrial DNA data. They speculated that population differentiation reflects their isolation by the uplift of the Central Highlands (for Toyama vs. Nobi+Osaka, ca. 0.8–5 million years ago [mya]) and the Suzuka Mountains (for Nobi vs. Osaka, ca. 1.0 mya), which separate these plains (Fig. 1). In contrast, microsatellite DNA analysis (10 loci) showed no such pattern of genetic differentiation but rather a closer genetic relationship between the Toyama and Osaka populations [22]. This uncertainty may be attributed to the inaccurate reconstruction of population history owing to the use of only a single or few genetic markers [23–25]. The microsatellite data were also used to estimate the recent demographic history of *A. longipinnis*, suggesting the influence of a recent bottleneck over the last few decades to several hundred years on each local population [22]. However, the reliability of estimating the pattern and timing of the bottleneck should be re-evaluated because of the small number of loci used and uncertain mutation rates [26, 27]. Unlike floodplains in continental regions, floodplains in the Japanese Archipelago, where *A. longipinnis* occurs, are small alluvial plains on the margins of islands. Therefore, geohistorical conditions, such as sea level and climate changes, would strongly impact the demography of this species. Genome-wide data with a large number

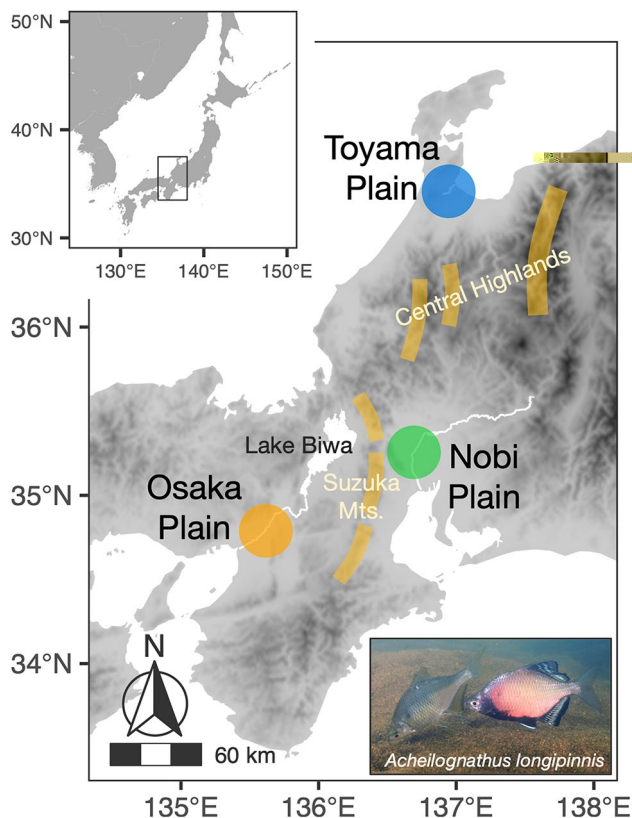


Fig. 1 The three plains where *Acheilognathus longipinnis* is distributed. Yellow lines indicate mountain ranges separating the plains. The photo: *A. longipinnis* in the Toyama Plain in autumn, the breeding season (left, a female extending an ovipositor; right, a male with nuptial color)

of loci would be useful for more reliable inference of demography over a long timescale.

This study aimed to elucidate the evolutionary history and demography of *A. longipinnis* over broad timescales based on whole-genome and mitogenome data to deepen our understanding of the distribution mechanism of floodplain fishes on islands and to contribute to the conservation of this endangered species. First, to clarify the degree, order, and timing of the differentiation of the three local populations of this species, we analyzed genetic population structure and phylogeny and conducted demographic modeling. Next, to explore the historical changes in population size experienced by each local population from their differentiation to the Anthropocene, we performed historical demography estimation using genome-wide SNP data. Finally, we discuss the relationships of the results of these analyses to climate change, the timing of the formation of mountain ranges that divide the present habitat, and human activities, comparing our results with previous hypotheses. By inferring the patterns and factors that shape the local populations of *A. longipinnis* through a next-generation phylogeographic approach, we highlight the potential of such an approach to understand more complex processes

of freshwater fish distribution than could be achieved by conventional methods.

Methods

Ethics statement

Acheilognathus longipinnis is a legally protected species designated as a Natural Monument of Japan and a “nationally rare species of wild fauna and flora” in Act on Conservation of Endangered Species of Wild Fauna and Flora, Japan. We obtained permission to conduct this research from the Agency for Cultural Affairs and the Ministry of the Environment, Japan. We used captive and wild fish tissues (fin clips) collected by non-invasive sampling or immediately after death. When sampling from live fish, a minimal portion of the fin tissue was clipped after anesthesia with 2-phenoxyethanol. The wild fish were gently returned to the river after collecting the fin tissue following a careful process. To extract high-quality genomic DNA for *de novo* whole-genome sequencing, a captive individual was euthanized by 2-phenoxyethanol overdose with utmost care following American Veterinary Medical Association Guidelines for the Euthanasia of Animals (2020).

De novo whole-genome sequencing

De novo whole-genome sequencing of *A. longipinnis* was performed on a captive individual from the Nobi population (Table 1). Total DNA was extracted from the testes of the individual euthanized by 2-phenoxyethanol overdose using a MagAttract HMW DNA Kit (Qiagen, Germany). Gel Bead-In-Emulsions were prepared, DNA fragments were barcoded with the 10x Genomics Chromium Controller instrument (10x Genomics, USA) and sequenced using a HiSeq X Ten (Illumina, USA). Library preparation and sequencing were outsourced to Macrogen Japan. The acquired 450 M linked reads were assembled using Supernova v2.0.0 [28] with default parameters. This assembler is specifically designed for genome assembly using linked reads and includes all necessary processes for assembly (e.g., trimming adapters and barcodes from reads, sorting reads by barcode, efficient k-mer estimation, assembly, edge correction, and bridging of contigs based on barcodes). The integrity of the draft genome was assessed using BUSCO v5.4.3 [29] with 3,640 actinopterygian core genes (-l actinopterygii_odb10).

Sampling and whole-genome resequencing

Whole-genome resequencing was performed on a total of 23 samples collected from all the three local populations of *A. longipinnis* from 2017 to 2020 (eight individuals from the Toyama population, the Busshouji River system; seven from the Nobi population, the Kiso River system; eight from the Osaka population, the Yodo River

Table 1 Sample information of *Acheilognathus longipinnis* and *Acheilognathus typus* (outgroup) used in the study

Sample ID	Species	Population	Type	Locality	Collecting Date
T-01	<i>Acheilognathus longipinnis</i>	Toyama	wild	Busshouji River, Himi, Toyama, Japan	8/23/2020
T-02	<i>Acheilognathus longipinnis</i>	Toyama	wild	Busshouji River, Himi, Toyama, Japan	8/23/2020
T-03 ‡	<i>Acheilognathus longipinnis</i>	Toyama	wild	Busshouji River, Himi, Toyama, Japan	8/23/2020
T-04	<i>Acheilognathus longipinnis</i>	Toyama	wild	Busshouji River, Himi, Toyama, Japan	8/23/2020
T-05 ‡	<i>Acheilognathus longipinnis</i>	Toyama	wild	Busshouji River, Himi, Toyama, Japan	8/23/2020
T-06 ‡	<i>Acheilognathus longipinnis</i>	Toyama	wild	Busshouji River, Himi, Toyama, Japan	8/23/2020
T-07 ‡	<i>Acheilognathus longipinnis</i>	Toyama	wild	Busshouji River, Himi, Toyama, Japan	8/23/2020
T-08	<i>Acheilognathus longipinnis</i>	Toyama	wild	Busshouji River, Himi, Toyama, Japan	8/23/2020
N-01	<i>Acheilognathus longipinnis</i>	Nobi	wild	Kiso River, Inazawa, Aichi, Japan	6/14/2018
N-02	<i>Acheilognathus longipinnis</i>	Nobi	wild	Kiso River, Inazawa, Aichi, Japan	6/14/2018
N-03 ‡	<i>Acheilognathus longipinnis</i>	Nobi	wild	Kiso River, Inazawa, Aichi, Japan	6/14/2018
N-04	<i>Acheilognathus longipinnis</i>	Nobi	wild	Kiso River, Inazawa, Aichi, Japan	6/14/2018
N-05 ‡	<i>Acheilognathus longipinnis</i>	Nobi	wild	Kiso River, Inazawa, Aichi, Japan	6/14/2018
N-06 † ‡	<i>Acheilognathus longipinnis</i>	Nobi	captive	Gifu World Fresh Water Aquarium, Kakamigahara, Gifu, Japan	11/3/2017
N-07	<i>Acheilognathus longipinnis</i>	Nobi	captive	Gifu World Fresh Water Aquarium, Kakamigahara, Gifu, Japan	11/3/2017
N-08 ‡	<i>Acheilognathus longipinnis</i>	Nobi	captive	Gifu World Fresh Water Aquarium, Kakamigahara, Gifu, Japan	11/3/2017
O-01	<i>Acheilognathus longipinnis</i>	Osaka	wild	RIEAFODiversity Center, Neyagawa, Osaka, Japan	11/29/2018
O-02	<i>Acheilognathus longipinnis</i>	Osaka	wild	RIEAFODiversity Center, Neyagawa, Osaka, Japan	11/29/2018
O-03	<i>Acheilognathus longipinnis</i>	Osaka	wild	RIEAFODiversity Center, Neyagawa, Osaka, Japan	11/29/2018
O-04	<i>Acheilognathus longipinnis</i>	Osaka	wild	RIEAFODiversity Center, Neyagawa, Osaka, Japan	11/29/2018
O-05 ‡	<i>Acheilognathus longipinnis</i>	Osaka	wild	RIEAFODiversity Center, Neyagawa, Osaka, Japan	11/29/2018
O-06 ‡	<i>Acheilognathus longipinnis</i>	Osaka	wild	RIEAFODiversity Center, Neyagawa, Osaka, Japan	11/29/2018
O-07 ‡	<i>Acheilognathus longipinnis</i>	Osaka	captive	Lake Biwa Museum, Kusatsu, Shiga, Japan	12/3/2019
O-08 ‡	<i>Acheilognathus longipinnis</i>	Osaka	captive	Lake Biwa Museum, Kusatsu, Shiga, Japan	10/21/2020
-	<i>Acheilognathus typus</i>	-	-	Izunuma Lake, Tome, Miyagi, Japan	2/25/2019

Note † Used for *de novo* whole genome sequencing; ‡ Used for MSMC analysis; RIEAFO: Research Institute of Environment, Agriculture and Fisheries, Osaka Prefecture

system; Table 1). For captive individuals at the Lake Biwa Museum (Osaka population; $n=2$), whole-body specimens were preserved in 99% ethanol or frozen immediately after death, and the right pelvic fin or a portion of the caudal fin was excised and preserved in 99% ethanol. For the other specimens, a portion of the right pelvic or caudal fin was minimally invasively excised and preserved in 99% ethanol.

Whole genomic DNA was extracted from the fin samples using the DNeasy Blood & Tissue Kit (Qiagen, Germany) or the QIAamp DNA Micro Kit (Qiagen, Germany). Libraries were prepared using the NEBNext Ultra II FS DNA Library Prep Kit for Illumina (New England Biolabs, USA). Indexing was performed using NEBNext Multiplex Oligos for Illumina (New England Biolabs, USA). Sequencing was performed using a HiSeq X Ten (Illumina, USA) in the 150-bp paired-end mode. Sequencing was outsourced to Macrogen Japan.

To be used as an outgroup for molecular phylogenetic analysis, a wild specimen of the Zenitanao bit-terling, *Acheilognathus typus*, the most closely related species of *A. longipinnis* [16, 30], was also collected and sequenced in the same manner (Table 1).

Mapping and SNP calling

For all acquired data (short-read sequencing data from 24 *A. longipinnis* specimens, including one individual used for *de novo* whole-genome sequencing, and one *A. typus*; Table 1), read quality control was performed using fastp v0.20.1 [31]. Reads with >40% unqualified bases (i.e., bases with a quality score of <20; -q 20) were excluded (-u 40). The window (four bases) was slid from the 3' and 5' ends of the reads, and the window with a mean quality of <20 was removed (-3 -5). The filtered reads were mapped to the reference genome using the bwa-mem2 v2.0 [32]. The draft genome assembled in this study was used as reference genome, excluding scaffolds of <15 kb. Alignments in SAM format were sorted and compressed in BAM format using Samtools v1.15.1 [33]. All possible multi-mapped reads were excluded (grep -v -e 'XA: Z:' -e 'SA: Z:'). PCR duplicates created during library preparation were marked using MarkDuplicates in Picard v2.25.0 (<http://broadinstitute.github.io/picard>).

Three different variant calling strategies were performed on the aligned reads according to the downstream analysis requirements. First, we called SNPs using HaplotypeCaller and GenotypeGVCFs in GATK v4.2.0.0 [34] for analyses that required only SNPs. We excluded bases with quality scores of <20 (-mbq 20). SNPs were

extracted by the GATK SelectVariants and filtered by GATK VariantFiltration with “QD<2.0 || FS>60.0 || MQ<40.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0.” In addition, we filtered out sites with SNP call quality too low, depths too low/high, non-biallelic sites, minor allele frequencies too low, and sites containing missing data using vcftools v0.1.16 [35], with the following parameters: minQ=30, max meanDP=46.76 (twice the mean depth in GATK calling), minDP=8, max-alleles=2, min-alleles=2, maf=0.05 (when using only the *A. longipinnis* data to avoid filtering out the unique alleles of *A. typus*), and max-missing=1. Finally, we used Plink v1.9 [36] (www.cog-genomics.org/plink/1.9/) to remove linked SNPs from the SNP-only dataset. We scanned windows of 50 variants, shifting the window by 10 variants each time, and greedily pruned variants with a squared correlation greater than 0.35 until no such pairs remained in the window (--indep-pairwise 50 10 0.35). This resulting dataset using only *A. longipinnis* data ($n=24$) is referred to as SNP-only dataset 1, and using *A. longipinnis* and *A. typus* bitterling data ($n=25$) is referred to as SNP-only dataset 2.

Second, for allele frequency spectrum (AFS) analysis, we called SNPs and invariant sites using mpileup in bcftools v1.15.1 [33, 37] on the *A. longipinnis* data ($n=24$). In addition, we filtered out sites with SNP calls of too low quality, too low/high depth, and multiallelic sites using vcftools with the following parameters: minQ=30, max meanDP=46.53 (twice the mean depth in bcftools calling), minDP=8, and max-alleles=2. We also filtered sites with missing data using vcftools under the following parameters: max-missing=0.8. This resulting dataset referred to as all-sites dataset 1. In addition, to eliminate the estimation bias introduced by simultaneous analysis of samples with low and high depths in the downstream analysis (i.e., demographic modeling), the quality-controlled reads were downsampled to a depth of approximately 10×, and the same mapping, sorting, compressing, and marking duplication as above were performed, and SNPs and invariant sites were called using bcftools. We filtered out sites with SNP calls of too low quality, too low/high depth, and multiallelic sites using vcftools with the following parameters: minQ=30, max meanDP=16.82 (twice the mean depth in bcftools calling), minDP=4, and max-alleles=2. We also filtered sites with missing data using vcftools under the following parameters: max-missing=1. This resulting dataset referred to as all-sites dataset 2.

Third, for historical demographic analyses, we called SNPs in the *A. longipinnis* data ($n=24$) using bcftools and bamCaller.py in MSMC Tools (<https://github.com/stschiff/msmc-tools>) (PSMC dataset). We filtered out bases with a minimum mapping quality<20 (--min-MQ

20), minimum base quality<20 (--min-BQ 20), or a depth greater than twice the scaffold-wide average per individual.

Mitochondrial genome assembling

De novo mitochondrial genome (mitogenome) assembling was performed using GetOrganelle v1.7.5.0 [38] on short-read sequencing data from 23 *A. longipinnis* specimens, excluding the specimen used for the *de novo* whole-genome sequencing, and one *A. typus* specimen. For the specimen used for *de novo* whole-genome sequencing, the mitogenome could not be assembled because short DNA fragments were removed during library preparation. Assembly was conducted using 10–30 M reads per individual. The animal_mt option was used. All the runs used k-mer sizes of 21, 55, 85, and 115. This resulted in complete mitogenome sequences of all individuals (approximately 16,770 bp). All sequences were annotated using the MitoAnnotator v3.67 [39].

Population structure

To elucidate the population structure of *A. longipinnis*, we performed principal component analysis (PCA), genome-wide F_{ST} , genetic divergence (d_{XY}), nucleotide diversity (π) calculation, and clustering analysis using genome-wide SNP data. PCA and clustering analyses were performed using the SNP-only dataset 1. Principal component analysis was conducted using Plink based on the variance-standardized relationship matrix. We calculated Weir and Cockerham's F_{ST} [40], d_{XY} , and π [41] in non-overlapping 10 kb windows with pixy [42] using the all-sites dataset 1. We estimated individual admixture proportions with the likelihood model-based program ADMIXTURE 1.3.0 [43], performing unsupervised clustering assuming a number of distinct genetic populations (K) from 1 to 10, with the convergence criterion set to 10^{-4} (-C 0.0001). This analysis was repeated 100 times using random seeds. The minimum cross-validation error values for each K were compared among the Ks, and the K with the smallest cross-validation error value was selected as the optimal K value.

We also investigated the genetic diversity of mitogenomes of each local population. We used Arlequin v3.5 [44] to calculate nucleotide diversity (π) and haplotype diversity (h) for each local population.

Phylogenetic analysis

Phylogenetic inferences were performed using genome-wide SNP and mitogenome data. For genome-wide SNP data, we used SNP-only dataset 2 and constructed a neighbor-net using SplitsTree v4.18.2 [45] with 1,000 bootstrap runs. Hamming distance was used as the genetic distance. Maximum likelihood phylogenetic tree estimation was performed as follows: SNP-only dataset

2 (multiple VCF) was converted to the PHYLIP format removing all missing sites using *vcf2phyli* v2.0 [46], and invariant sites were removed using *raxml_ascbias* (https://github.com/btmartin721/raxml_ascbias) to create concatenated SNP sequences (43,843 SNPs). For these concatenated sequences, a maximum likelihood phylogenetic tree was estimated using IQ-TREE v2.0.7 [47]. We used TVM+F+ASC+R2 as the nucleotide substitution model, which was selected with BIC using ModelFinder [48] implemented in IQ-TREE. Only the models that included ascertainment bias corrections (ASC) were explored [49]. The Shimodaira–Hasegawa-like approximate likelihood ratio test (SH-aLRT) [50] and ultrafast bootstrap approximation (UFBoot2) [51] were used for statistical evaluation of the branches. In these phylogenetic analyses, all heterozygous sites are labeled using the IUPAC nomenclature.

For mitochondrial phylogeny, the mitogenome sequences of 42 species and subspecies of the bitterling subfamily Acheilognathinae and six other cypriniform outgroups were obtained from the National Center for Biotechnology Information (NCBI) database and analyzed with *A. longipinnis* and *A. typus* sequences newly generated in the present study (Table S1). Their 13 protein-coding genes (PCGs; ND1, ND2, COI, COII, ATPase 8, ATPase 6, COIII, ND3, ND4L, ND4, ND5, ND6, and Cyt *b*) were used to estimate maximum likelihood (ML) and Bayesian phylogenetic trees. The sequence alignment of each gene was performed using *mafft* v7.490 [52]. Unique haplotypes were selected for the analysis: 12 haplotypes for *A. longipinnis*, 2 haplotypes for *A. typus*, 1 haplotype each for the other 40 bitterling species/subspecies, and 6 outgroup species, totaling 60 haplotypes (Table S1). For ML tree estimation, PCGs were partitioned, and base substitution models were selected by BIC using ModelFinder implemented in IQ-TREE (Table S2). SH-aLRT and UFBoot2 were used as the statistical evaluation values for the branches.

The divergence times of the bitterling species were estimated using the Bayesian relaxed clock model in BEAST v2.7.0 [53]. The 13 PCGs were partitioned; the site and clock models were unlinked for all partitions, and the tree was linked. For the site model, we used the model-averaging approach with the bModelTest [54]. As time calibration points, we used 20 mya, corresponding to the oldest fossil record of Acheilognathinae from the Early Miocene, as the minimum age [16, 55], and 33 mya, the divergence age between Acheilognathinae and the sister subfamily Gobioninae [56–60], as the maximum age of the time to the most recent common ancestor (TMRCA) of Acheilognathinae. The median divergence age of the results from several relevant studies [56–60] was obtained from TimeTree 5 [61] (accessed September 5, 2022). To include them in the 95% confidence interval,

the prior probability of the TMRCA for Acheilognathinae was given as a normal distribution with mean=26.5 mya and standard deviation=3.3. We used an optimized relaxed clock model and a birth-death model as the tree prior. The MCMC was set at 500,000,000 steps, and sampling was performed every 2,000 steps. Tracer v1.7.2 [62] was used to verify that the effective sample size (ESS) of the parameters was >200. After excluding the first 20% as burn-in, all trees were combined using TreeAnnotator v2.7.0 [53].

Demographic analyses

The historical demography of *A. longipinnis* was estimated using the pairwise sequential Markovian coalescent approach (PSMC) [63], multiple sequential Markovian coalescent approach (MSMC) [64], and approximate Bayesian computation approach (PopSizeABC) [65]. It is empirically suggested that these methods differ in the time periods for which historical demography can be adequately estimated, i.e., for PSMC, hundreds to hundreds of thousands of generations ago; for MSMC, one hundred to hundreds of thousands of generations ago; and for PopSizeABC, several to hundreds of thousands of generations ago, although they also depend on the data used and the true demographic history [10]. PSMC was performed as PSMC' by *msmc* v2.1.3 [66], using the data on a scaffold of >100 kb in the PSMC dataset with default settings for each individual. MSMC was performed using *msmc* v2.1.3, on the data on a scaffold of >100 kb in the PSMC dataset phased by the SHAPEIT2 [67]. The VCFs of each individual in the PSMC dataset were merged (*bcftools merge --missing-to-ref*) and phased by read-aware phasing [68] using SHAPEIT2, after removing multiallelic sites. The phased VCFs were split individually (*bcftools split*), and sites with only the same bases as the reference were removed. For each local population, MSMC was performed using the phased VCFs of the four individuals with the highest depth in SNP calling with *bcftools* (Table 1) and default settings, except for sites with ambiguous phasing, which were skipped (-s). For MSMC, 30 bootstrap runs were conducted using *multihetsep_bootstrap.py* from *msmc-tools* (<https://github.com/stschiff/msmc-tools>) with the default settings. For both PSMC' and MSMC, a mappability bed file was created by calculating the (150, 2)-mappability using GenMap v1.3.0 [69]. The generation time and mutation rate per site per generation were assumed to be 1 year [11, 15, 17] and 3.50×10^{-9} [70], respectively.

A more recent historical demography was estimated using PopSizeABC with the merged PSMC dataset. Two summary statistics, AFS and linkage disequilibrium, were used in the ABC analysis. These summary statistics were calculated for 21 discrete time windows (the oldest time window started 30,000 generations before the present)

based on the empirical dataset and compared with the corresponding statistics calculated from the simulated datasets. The minor allele frequency for calculating linkage disequilibrium statistics was set to 0.2 to minimize prediction error. Linkage disequilibrium was summarized for each segment of 2,000,000 bp. For the simulated datasets, the number of segments was set to 30 (the size of each segment was 2,000,000 bp), and the number of simulated datasets was set to 1,000,000. Log-uniform priors on the effective population size (N_e) ranging from 10 to 500,000 ($\log_{10}$ -scale: 1–5.7) and on the recombination rate ranging from 0.1×10^{-8} to 1×10^{-8} were used. The generation time and mutation rate per site per generation were assumed to be identical to those described above. ABC estimates of population sizes were obtained using neural network regression with a tolerance of 0.001.

Demographic modeling was performed using fastsimcoal v2.7 [71] to examine population differentiation and gene flow patterns. A multi-dimensional AFS among three local populations was generated using easySFS (<https://github.com/isaacovercast/easySFS>) from the data on a scaffold of >100 kb in the all-sites dataset 2. A total of 48 demographic models were generated, focusing on the order of local population differentiation and the timing and extent of gene flow (Fig. S1), and compared using the AIC with the composite likelihood. We simultaneously estimated the time of population differentiation, the N_e of the present and ancestral populations, migration rates among the differentiated populations, and the time at which gene flow ceased among local populations. The last parameter was incorporated because, based on the geography, connectivity of freshwater systems, and the results of population structure analysis using nuclear genome data (see Results), there appeared to be no gene flow among the current local populations. The generation time and mutation rate per site per generation were assumed to be identical to those described above. To confirm convergence, fastsimcoal2 was run independently, 100 times for each model, and the results for the run with the highest log-likelihood were obtained. Each run consisted of 200,000 coalescent simulations (-n 200,000) and 50 Expectation/Conditional Maximization (ECM) cycles (-L 50). The threshold for observed SFS entry count set as 10 (-C 10). Following the method in Meier et al. [72], the likelihood estimation for each model was repeated 100 times using the parameter values estimated in the run with the highest log-likelihood to compute the AIC distribution. The AIC distributions were compared across the models to estimate the best model. If the 95% interval of the AIC distributions did not overlap, the two models were considered significantly distinct. Non-parametric bootstrapping was used to obtain 95% confidence intervals for each parameter in the model with the smallest AIC, as follows: first, the genome was divided into 1 Mb

windows and 100 bootstrap genomes were created by randomly sampling the windows allowing for duplicates, to create a bootstrapped AFS. Using this AFS, fastsimcoal2 was run 10 times independently for the AIC minimum model to estimate the maximum likelihood of the parameters. Using the 100 parameter estimates obtained, 95% confidence intervals were calculated.

Results

De novo whole-genome sequencing and resequencing

We obtained 1,159 M Chromium linked reads from a male *Acheilognathus longipinnis* genome and performed a draft genome assembly with 450 M linked reads using Supernova v2.0.0. The assembly size, number of scaffolds, N50, L50, and coverage for scaffolds >1,000 b were 813 Mb, 26,074, 3.4 Mb, 58, and 65×, respectively. The BUSCO scores were 84.2%, 2.9%, and 4.4% for single-copy, duplicate, and fragmented genes, respectively, with 8.5% missing.

As a result of whole genome resequencing of 24 *A. longipinnis* and one *A. typus*, on average 179.6 M reads (s.d. 64.8 M reads) were obtained per individual, and the average coverage was 24.7 (s.d. 11.1) (Table S3). The number of sites in each dataset was as follows: SNP-only dataset 1: 49,348 SNPs; SNP-only dataset 2: 97,456 SNPs; all-sites dataset 1: 103,441,976 sites with 2,211,353 SNPs; all-sites dataset 2: 595,883,484 sites with 6,299,888 SNPs; PSMC dataset: 1,191,189 SNPs on average.

Population structure

The PCA results showed that *A. longipinnis* had a distinct population structure, in which the three local populations were almost equally differentiated from each other (Fig. 2a). The mean F_{ST} and d_{XY} values of 10 kb windows across genome were also almost the same, ranging from 0.22 to 0.34 for the F_{ST} and from 0.00721 to 0.00789 for the d_{XY} for each population pair (Fig. 2b). The mean π value was highest in the Osaka population (0.00699 ± 0.00841 s.d.), followed by the Toyama population (0.00556 ± 0.00915) and the Nobi population (0.00484 ± 0.00707) (Fig. 2a).

ADMIXTURE analysis showed that the cross-validation error was the lowest for $K=2$, followed by $K=1$, and $K=3$ (Fig. S2). For $K=2$, the Toyama and Nobi populations were separated into independent populations, and the Osaka population was estimated to be a mixture of the two (Fig. 2c). The admixture coefficient of the Osaka population was almost constant, at approximately 25% for the Toyama component and 75% for the Nobi component. For $K=3$, each local population was estimated as independent (Fig. 2c).

The mean π value of the mitogenomes was much higher in the Osaka population (0.00112 ± 0.00063 s.e.) than the Toyama population (0.00028 ± 0.00018) and

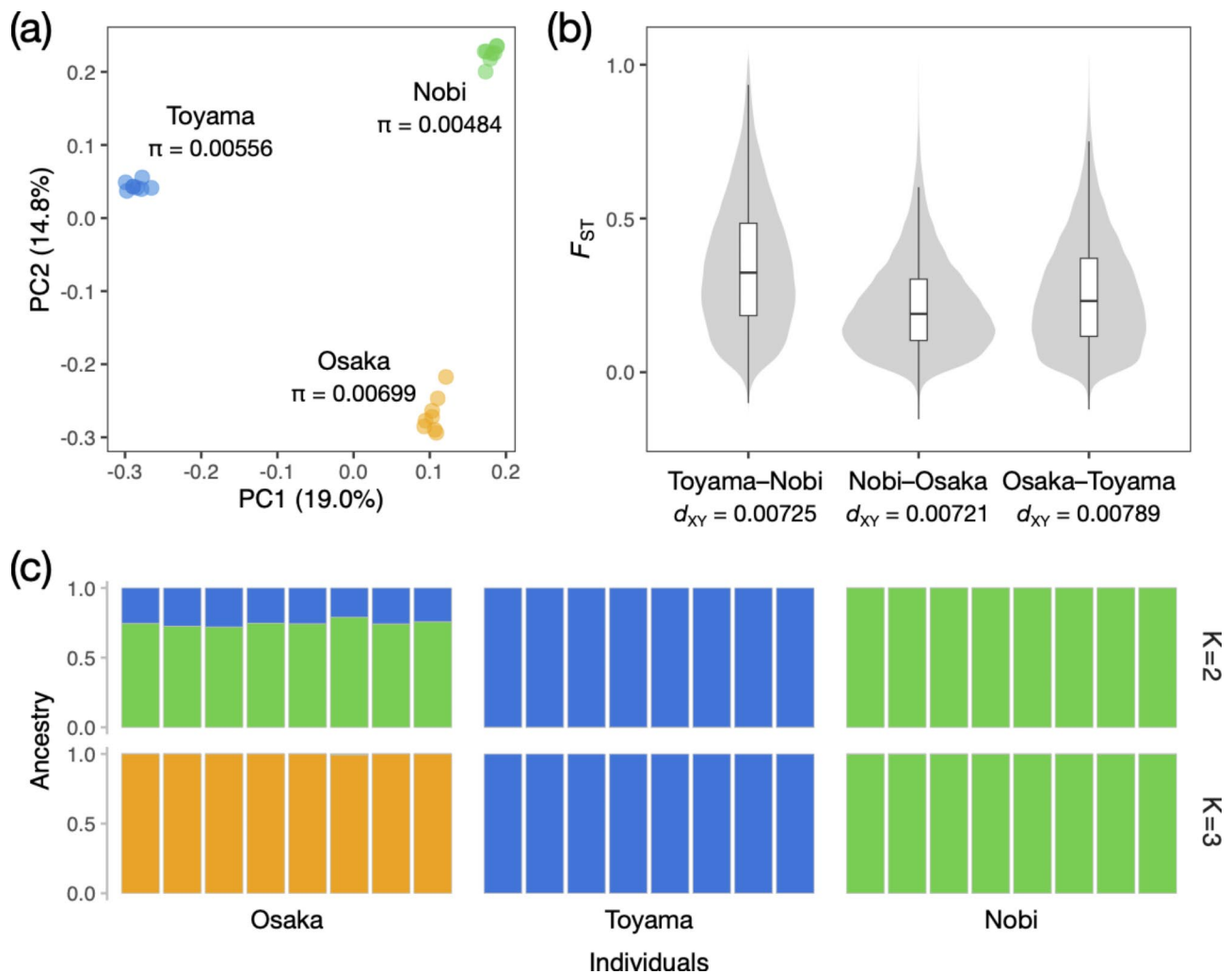


Fig. 2 Results of the population structure analyses for *Acheilognathus longipinnis*. **(a)** Scatter plots for the principal components 1 and 2 with the value of genome-wide mean nucleotide diversity (π) for the three local populations. **(b)** Estimated pairwise genome-wide F_{ST} and d_{XY} for the three local population pairs. **(c)** Estimated individual admixture proportions using ADMIXTURE at $K=2$ and 3

the Nobi population (0.00007 ± 0.00006), whereas the mean h values were similar but highest in the Toyama population (0.9286 ± 0.0844), followed by the Osaka population (0.8214 ± 0.1007) and the Nobi population (0.8095 ± 0.1298) (Fig. 3).

Phylogenetic analysis

The nuclear genome phylogenetic tree supported clear differentiation among the three local populations of *A. longipinnis*. The neighbor-net showed that the position of the root by the outgroup (*A. typus*) was not clear, and the three populations were shown to have a nearly trifurcated relationship (Fig. 4a). The ML tree, reconstructed by SNP concatenated sequences, showed that the Toyama, Nobi, and Osaka populations were each monophyletic, with a sister relationship between the Toyama and Nobi populations. The tree was highly supported statistically

(UFBoot=100, SH-aLRT=100) except for some terminal branches (Fig. 4b).

The mitogenome phylogenetic trees reconstructed using the ML and Bayesian methods showed very similar topologies, although there were some discrepancies around the branches with low statistical support (Fig. S3). The phylogenetic position of *A. longipinnis* within Acheilognathinae was similar to that reported by Kawamura et al. [16] and C. H. Chang et al. [30], with *A. typus* as a sister species (Fig. S3). The divergence time between *A. longipinnis* and *A. typus* was estimated to be 4.4 mya (95% credible interval: 3.15–5.74; Fig. S3), and the TMRCA of *A. longipinnis* to be 0.17 mya (0.11–0.26; Fig. 3). The relationships among the haplotypes of *A. longipinnis* were highly supported (UFBoot $\geq 94\%$, SH-aLRT $> 83\%$, BPP > 0.98), except in some terminal parts of the tree (Fig. 3; Fig. S3). Haplotypes from the Toyama and Nobi populations were each monophyletic, whereas those

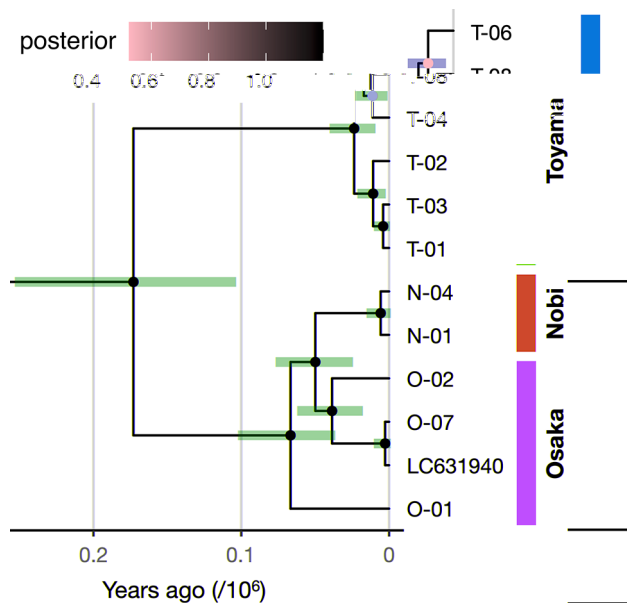


Fig. 3 A time tree estimated using mitochondrial 13 protein-coding gene data of *Acheilognathus longipinnis*. The colors of the nodes indicate the posterior probabilities. Blue bars at nodes indicate 95% credible intervals. Only the *A. longipinnis* clade is shown. See Fig. S3 for the entire tree for *Acheilognathinae*

from the Osaka population were paraphyletic and formed a monophyletic group together with those from the Nobi population (Fig. 3). The divergence time of the Nobi haplotypes from the Osaka haplotypes was estimated to be 0.05 mya (0.03–0.08; Fig. 3).

Demographic analyses

The PSMC' results showed that each local population of *A. longipinnis* experienced a reduction in population size during the last glacial period (14–115 kilo years ago [73]; kya; Fig. 5). This trend was also supported by the MSMC results (Fig. 5). The Toyama and Nobi populations continued to decrease during the post-glacial period, whereas the Osaka population showed an increase of approximately 3–10 kya and decreased again between 1 and 3 kya. The PopSizeABC results indicate a post-glacial population size reduction in all the local populations: for the Toyama population, after ca. 0.5 kya, for Nobi population, during ca. 1–10 kya and after ca. 0.5, and for Osaka population, after ca. 0.5 kya (Fig. 5). The latest N_e values were estimated to be larger for the Toyama, Nobi, and Osaka populations, in this order. The results from PopSizeABC showed good agreement with the MSMC results for the Nobi and Osaka populations but exhibited some discrepancies for the Toyama population in the 1–10 kya range, where the former showed a population increase.

In demographic modeling, Model 48 was selected as the best fit model based on the smallest AIC value among all the 48 models, when the approximately 10×

downsampled dataset was used to account for heterogeneous coverage between samples (Fig. 6, S4). The same result was also observed from the undownsampled dataset (Fig. S5). This model assumed gene flow among the differentiated populations at every assumed time point and suggested that the Osaka population was the first to differentiate. However, in the downsampled dataset, several other models (Models 11, 27, 29, 41, and 47) also provided comparable fits to the data, with the 95% AIC interval overlapping with that of the best model ($\Delta\text{mean AIC}=2,026\text{--}6,957$; Fig. S4). These models differed in terms of which population differentiated first. That is, although the Osaka population was the first to differentiate in Models 40 and 47 as well, the Toyama population was the first in Model 11 and the Nobi population was the first in Models 27 and 29. All these models included the assumption of gene flow among the populations (Fig. S4).

In Model 48 (Fig. 6), the Osaka and Nobi–Toyama ancestral populations were initially differentiated at 109 kya (T2; 95% confidence interval 2.48–252 kya), followed by a divergence between the Nobi and Toyama populations 3.75 kya (T1; 1.82–16.1 kya; Fig. 6; Table 2). Migration among the three local populations was estimated to have lasted up to 1.49 kya (Tm; 0.73–4.86 kya; Fig. 6; Table 2). The N_e of the ancestral population of the three local populations was estimated to be very large (NAnCOT: 10.2×10^5), while the relatively small N_e were estimated for the Nobi–Toyama ancestral population (NAnCOT: 2.05×10^5) and those of the Osaka, Nobi, and Toyama populations (NO: 0.193×10^5 , NN: 0.0943×10^5 , NT: 0.104×10^5), although the 95% confidence intervals were all broad (Fig. 6; Table 2). While all the estimated $N_e \times$ migration rates were generally low, those from the Toyama population to the Nobi population (MIG3NT: 2.39 /generation), from the Toyama population to the Osaka population (MIG6NT: 2.16), and from the Nobi population to the Osaka population (MIG8NK: 6.20) were comparatively higher than the rest, although the 95% confidence intervals were also wide (Fig. 6; Table 2). The estimated parameters of the other five best models exhibited similar trends (Table S4).

Discussion

Population divergence

Our genomic study confirmed that specimens from each of the three local populations of *Acheilognathus longipinnis* clustered into three distinct, homogeneous genetic groups (Figs. 2 and 4). This finding is consistent with the results of previous studies based on partial mtDNA sequences and microsatellites [20, 22]. However, our results suggest a different scenario for the evolutionary history of populations compared to previous studies,

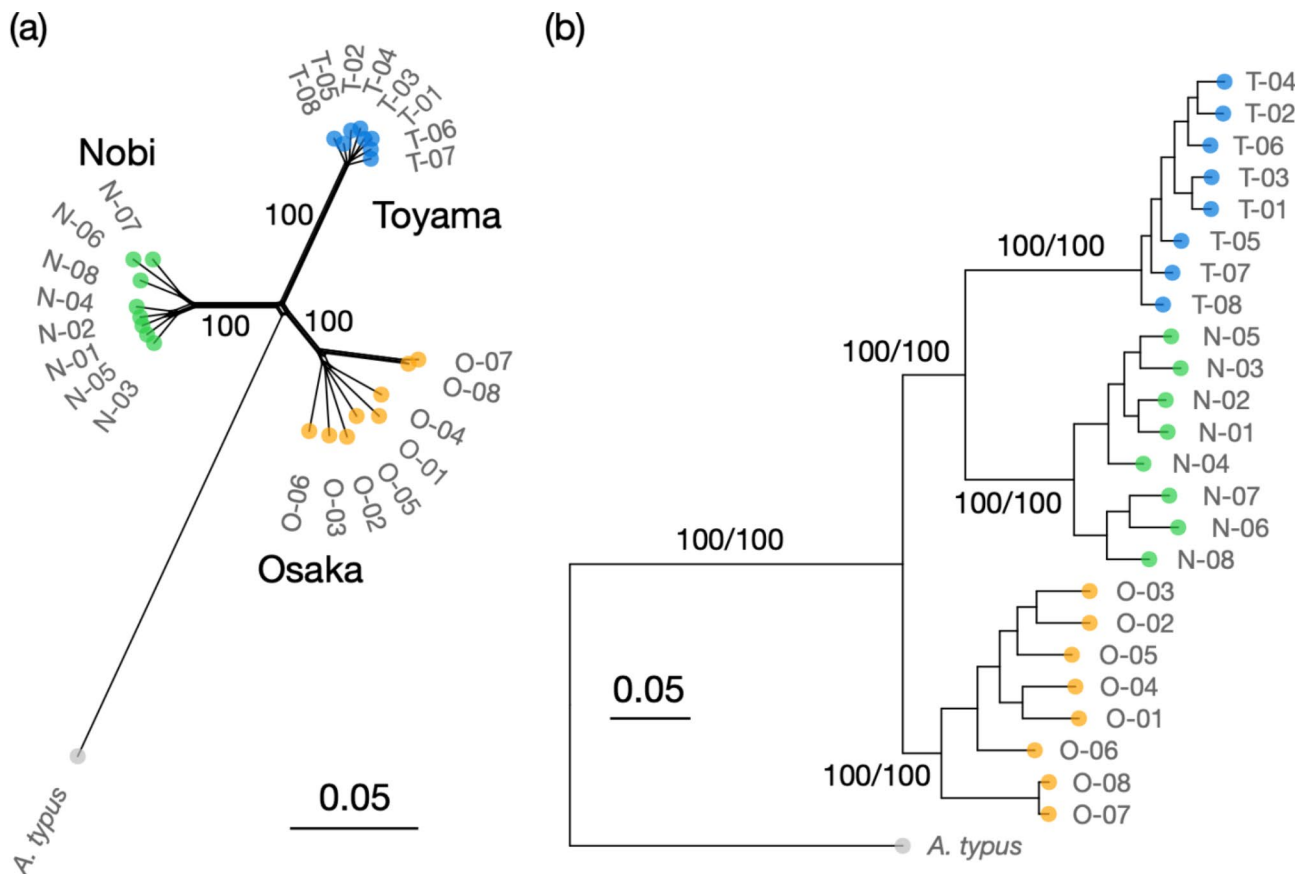


Fig. 4 Inferred phylogenetic relationships among *Acheilognathus longipinnis* using genome-wide 97,456 SNP data. **(a)** Neighbor-net based on the Hamming distance. The values on the branches indicate bootstrap values. **(b)** Maximum-likelihood tree. The values on the branches indicate statistical support values of Shimodaira–Hasegawa-like approximate likelihood ratio test (left) and ultrafast bootstrap approximation (right)

particularly with respect to the order and timing of population divergence and the effects of gene flow.

Regarding the order of differentiation of the three local populations, they were almost trifurcated. However, the nuclear genome ML tree indicated that the Osaka population diverged the earliest with strong statistical support. This is also supported by the demographic modeling involving gene flow among the populations, although different tree topologies could not be statistically rejected. These results contradict those of Okazaki et al. [21] and Kitanishi et al. [20] based on partial mitochondrial DNA sequences (ND2 and 12 S), in which the Toyama population differentiated first, followed by a divergence between the Nobi and Osaka populations. Our mitogenome phylogeny using 13 PCGs supported the results of previous mtDNA studies. Therefore, the discordance between the nuclear and mitochondrial DNA results is not attributable to insufficient data, but rather strongly suggests that the phylogeny of the mitochondrial DNA, which is inherited as a single linkage group, does not reflect the population phylogeny.

The discordance between nuclear and mitochondrial DNA trees in *A. longipinnis* includes two aspects: the

order of population differentiation (explained above) and the paraphyly of haplotypes from the Osaka population. The discordance between nuclear and mitochondrial (organelle) DNA phylogenies can generally be caused by hybridization, incomplete lineage sorting, or some mechanisms related to maternal inheritance of mitochondrial DNA (e.g., sex-specific differences in migration) [25, 74], although the latter is unlikely to apply to this species [11, 12, 75]. In our mitogenome tree, although the parts of the Osaka population haplotypes formed a monophyletic group with all the Nobi population haplotypes, recent migration or hybridization can be ruled out because of the distinct differentiation of the Nobi haplotypes from the Osaka haplotypes, corresponding to 0.05 mya (Fig. 3). However, neither the possibility of traces of past gene flow (as discussed below) nor incomplete lineage sorting can be completely ruled out as causes of nuclear and mtDNA discordance in this case.

Regarding the timing of local population differentiation, the demographic modeling analysis suggested that differentiation of the three populations began with the separation of the Osaka population dating back from 2.48 to 252 kya (95% confidence interval; maximum likelihood

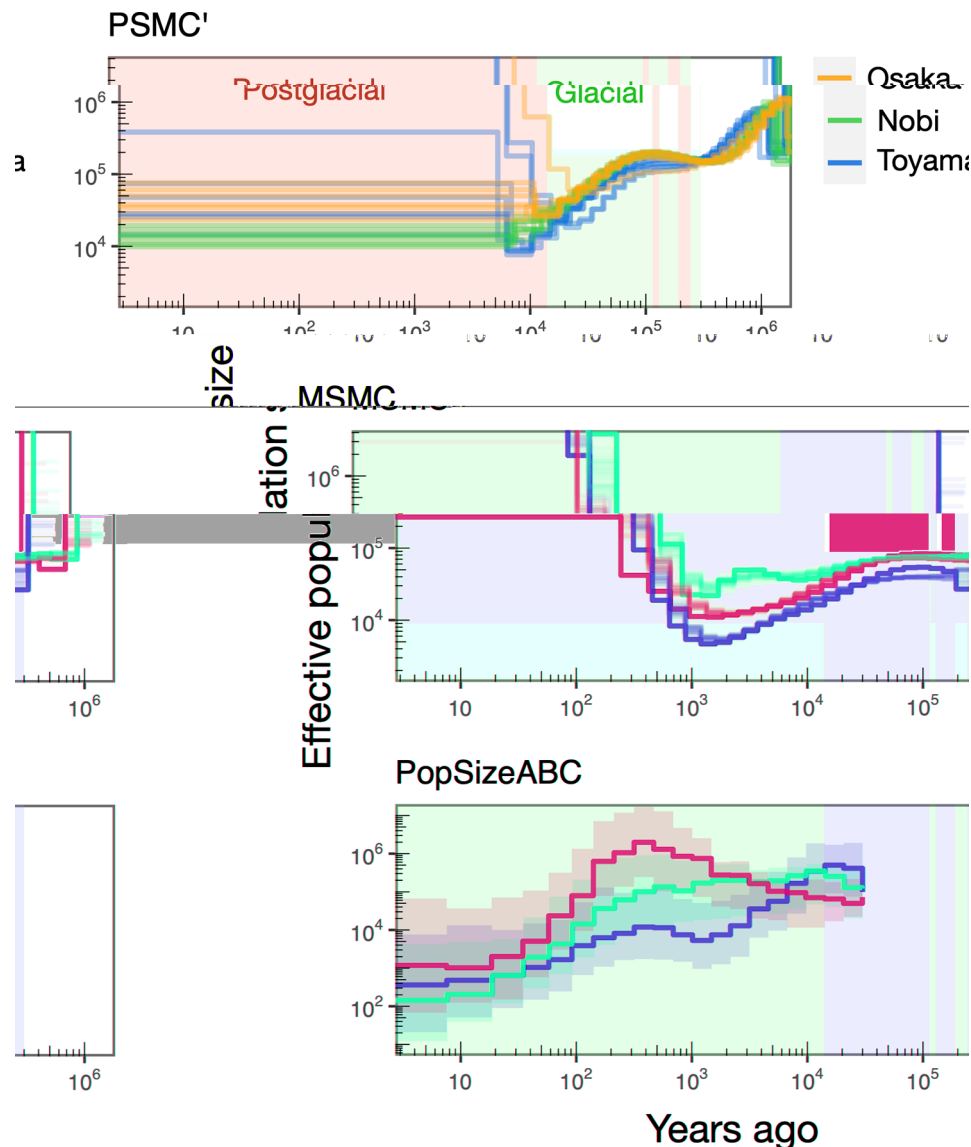


Fig. 5 Historical demography of *Acheilognathus longipinnis* estimated by pairwise sequentially Markovian coalescent approach (using msmc2; PSMC), multiple sequentially Markovian coalescent approach (MSMC), and approximate Bayesian computation approach (PopSizeABC). In the MSMC results, the light lines represent the results of 30 block bootstrapping runs; in the PopSizeABC results, the lines represent the median values and the ribbons at 95% credible intervals. In all panels, the green background indicates glacial periods, and the red background indicates interglacial or post-glacial periods after 300 kya

estimate, 109 kya), followed by the divergence of the Toyama and Nobi populations from 1.82 to 16.1 kya (3.75 kya), under conditions of restricted gene flow. The confidence intervals were wide, and the precision of the estimates was low. However, the former interval value for the onset of population divergence included the TMRCA of the *A. longipinnis* mitogenomes, estimated to be 180 kya (median; 95% confidence interval: 115–254 kya). Despite the wide confidence range of the estimates, our results contradict the evolutionary history scenario of *A. longipinnis* proposed in previous studies [20, 21], in which the uplift of the Central Highlands (ca. 800–5,000 kya) and Suzuka Mountains (ca. 1,000–1,500 kya) triggered

the detected genetic differentiation between local populations. These mountain uplift events have also been considered to be the cause of population differentiation patterns in more than a dozen freshwater fish species, and thus have been frequently used to calibrate the divergence times between vicariant populations (and closely related species), including Acheilognathinae species (e.g., [76–79]). However, it has also been found that the degree of differentiation between local populations across these mountainous terrains varies [79], with some cases, such as those involving relict lineages that diverged long before the mountain uplift, not being attributed solely to vicariance caused by the uplift (e.g., [80]). Our results suggest

Model 48

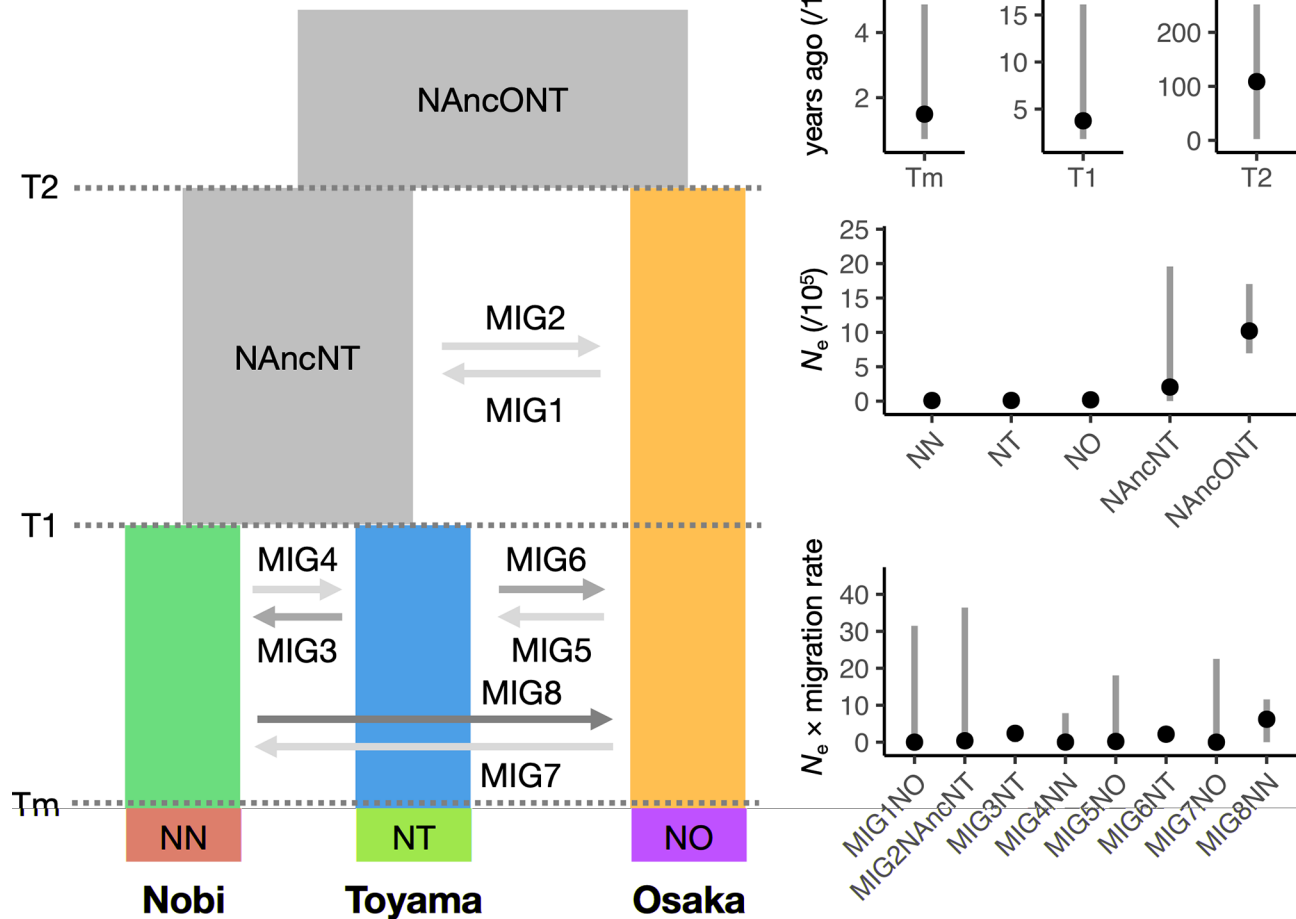


Fig. 6 The model with the smallest AIC in the demographic modeling for *Acheilognathus longipinnis* and the estimated parameters with a 95% confidence interval. Tm: time point at which migration among the populations stopped; T1, time point at which the Nobi and Toyama populations differentiated; T2, time point at which the Osaka population and the ancestral population of the Nobi and Toyama populations differentiated; NN, N_e of the Nobi population; NT, N_e of the Toyama population; NO, N_e of the Osaka population; NAncNT, N_e of the ancestral population of the Nobi and Toyama populations; NAncONT, N_e of the ancestral population of the three local populations; MIG1–8, migration rate among the three local populations or the ancestral population of the Nobi and Toyama populations

that the formation of the current population structure of *A. longipinnis* was also influenced by other factors that occurred after mountain uplift (see the “Distribution history” section below). This highlights the importance of considering various distribution processes (i.e., dispersal, vicariance, and secondary contact), even among strictly freshwater fishes that cohabit.

Population demography and distribution history

Each local population of *A. longipinnis* was suggested to have experienced at least one population contraction during the last glacial period (14–115 kya; [73]). Historical demography analyses showed that the N_e of each local population declined by a factor of 0.1–0.5 during this period. As speculated by Kitanishi et al. [20], this may be due to the low temperatures and aridity during the glacial period [81, 82], which impeded the formation of

floodplain environments to which *A. longipinnis* is highly dependent. In addition, the decline in population size may have been caused by the interruption of migration processes between the studied populations and other presumably existing populations.

Even in the post-glacial period, with a warmer and more precipitation-rich climate, floodplain species such as *A. longipinnis* are expected to be hindered in their population persistence and expansion because of the loss of plain areas due to sea-level rise. Our MSMC results suggest that the Toyama and Nobi populations experienced a continuous population decline after the last glacial period. In contrast, the Osaka population increased immediately after the end of the last glacial period. This may be attributed to the influence of Lake Biwa, the largest lake in Japan (surface area, 670.3 km²; volume, 27.5 km³), which is located in the Lake Biwa–Yodo River

Table 2 The estimated parameters and a 95% confidence interval for the Model 48

Parameter	Estimated value	95% L	95% U
Tm	1,493	730	4,863
T1	3,754	1,817	16,119
T2	108,735	2,482	251,621
NN	9,431	5,758	28,086
NT	10,496	5,766	30,774
NO	19,341	8,787	65,862
NAncNT	205,020	1,380	1,958,167
NAncONT	1,021,134	695,645	1,703,133
MIG1NO	0.005	0.000	31.478
MIG2NAncNT	0.378	0.000	36.409
MIG3NT	2.394	0.000	3.329
MIG4NN	0.025	0.000	7.857
MIG5NO	0.194	0.000	18.060
MIG6NT	2.158	0.000	3.255
MIG7NO	0.005	0.000	22.547
MIG8NN	6.204	0.000	11.578

Note 95% L/U means the lower/upper bound of the 95% CI

system that the present Osaka population inhabits. Lake Biwa is located in a basin at an elevation of approximately 86 m above sea level and has maintained a lake environment as large as the present one since 430 kya [83, 84]. This lake may have functioned as a habitat that allowed population expansion even when floodplain environments were greatly reduced on the Japanese islands due to global sea level rise during the post-glacial period (5–7 kya; [85, 86]). The Osaka population showed the highest π values for both the genome-wide SNP data and the mitogenome among the three populations, which is consistent with this hypothesis. However, the temporary increase in the Toyama population after the last glacial period shown by PopSizeABC was inconsistent with the MSMC results; the reasons for this are unclear.

The demographic analysis for the more recent past suggested a continuous population decline in all the local populations after 0.5 kya, comparable in magnitude to the decline during the last glacial period. There is a possibility that this inferred decline could be an artifact of the analysis, particularly due to the low levels of gene flow present, which could produce signatures of recent or ongoing bottlenecks [87]. However, the history of human migration and land modification, as well as the recent observation of the population decline of *A. longipinnis*, support the idea that our results reflect the demography of this species. Humans, widely distributed in the Japanese Archipelago (first migration, 37–38 kya; semi-sedentary Jomon culture, since 12 kya; Yayoi “wet rice” culture, since 2.3 kya; [88–90]), may have contributed to the population declines of *A. longipinnis* by converting floodplains into rice paddies and settlements, especially after the 1600s, when the human population exploded [91]. Over the past 150 years, humans have significantly

altered floodplain environments through river modification and land reclamation. Specifically, the Osaka population of *A. longipinnis* was inferred to have experienced a drastic decline from 500 to 50 years ago, implying the strong influence of these factors on the Osaka Plain, which has been one of the largest urban areas in Japan for several hundred years. The low mean mitogenome h value in contrast to the high mean π value (for genome-wide SNPs and mitogenomes) in the Osaka population suggests haplotype dropout and may reflect the recent population decline in this population. While all populations experienced rapid declines and were found to be very small in size, the observed nucleotide diversity was relatively high (0.00484–0.00699) compared to other threatened freshwater and marine fishes (0.00006–0.0053) [92], possibly reflecting large population sizes in the past. Rapidly declining populations may accumulate deleterious mutations and therefore require specific conservation actions [93, 94].

Regarding the distribution history of *A. longipinnis* based on our results, the following processes are hypothesized: (1) when *A. longipinnis* had a very large population size and was more widely distributed than at present (10–1,000 kya), migration between plains and population establishment were possible at low frequencies, even after the uplift of the Central Highlands and the Suzuka Mountains, through some biogeographic corridors (e.g., low watersheds, bypassing the mountains, or the coastline during low salinity timing); (2) the present distribution was established as a result of continuous habitat shrinkage due to environmental changes during the glacial period and human activities during the post-glacial period. Although the dispersal range of strictly freshwater organisms is undoubtedly restricted to contiguous water systems, our results suggest that the establishment of geographic barriers dividing the present water system may not necessarily have caused their population differentiation; i.e., their dispersal capacity may have been underestimated.

Conclusion

The present study, based on mitogenome and whole-genome data for the floodplain bitterling *Acheilognathus longipinnis*, provides a novel scenario for population differentiation and demographic history and a fundamental basis for future conservation studies of this endangered species. The initial separation of the populations was inferred to have occurred within several hundred thousand years, therefore more recently than previously assumed. This evolutionary scenario was unexpected, as mountain uplift in older times was assumed to be the main driver of population differentiation so far based on the species’ strong dependence on the floodplain environment in the lowlands. Our results force us to consider

the migration between plains isolated by mountains that occurred after the mountain uplift. Our findings suggest that such migratory potential in freshwater fishes, which has often been underestimated in analyses with limited genetic markers, is responsible for the differences in distribution and divergence patterns among other plain freshwater fishes [79]. The trends in population decline, possibly related to climate in the last glacial period and human activities, were revealed by reconstructing historical demography over a wide range of timescales using whole genome information. In particular, the post-glacial decline, which was possibly related to human activities, was estimated to be as severe as that in the last glacial period, suggesting that the modification of floodplain environments by human activities had a significant impact on the population size of this species. Further genome-based comparative phylogeographic studies of co-distributed floodplain species could provide a more comprehensive framework to understand the evolutionary history and distribution mechanisms of freshwater organisms, detail human impacts on floodplain ecosystems, and contribute to their conservation.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12862-024-02326-y>.

Supplementary Material 1

Supplementary Material 2

Acknowledgements

We thank Dr. Yo Y. Yamasaki for his important help with the genetic analysis. Computations were partially performed on the NIG supercomputer at ROIS National Institute of Genetics and the supercomputer of ACCMS, Kyoto University.

Author contributions

KO and KW designed this study, analyzed the data, and wrote the early manuscript; TM and RI constructed the reference genome of *Acheilognathus longipinnis* and performed preliminary analyses; YH obtained the genomic data of *A. typus*; KI, KU, MN, RT, and SM performed *A. longipinnis* sampling; SM, KW, and KO obtained funding for this study. All authors reviewed and contributed to revisions of the manuscript.

Funding

This research was supported by the River Works Technology Research and Development Program of The Ministry of Land, Infrastructure, Transport and Tourism (to SM), JSPS KAKENHI Grant Number 20H03009 (KW), Grant-in-Aid for JSPS Fellows Grant Number 22J23327 (KO), and the Environment Research and Technology Development Fund (JPMEERF20224M02) of the Environmental Restoration and Conservation Agency of Japan (KW).

Data availability

The draft genome sequences of *Acheilognathus longipinnis* are deposited to the DNA data bank of Japan (DDBJ) (BTCV01000001–BTCV01026061; BioProject PRJDB15958). Raw sequence reads are deposited in the DDBJ Sequence Read Archive (DRA) (BioProject PRJDB15958 for all *A. longipinnis*; Experiment DRX490417 in BioProject PRJDB16614 for *A. typus*). Unique mitogenome haplotype data are deposited to DDBJ (LC712739–LC712753).

Declarations

Ethics approval and consent to participate

Acheilognathus longipinnis is a legally protected species designated as a Natural Monument of Japan and a “nationally rare species of wild fauna and flora” in Act on Conservation of Endangered Species of Wild Fauna and Flora, Japan. We obtained permission to conduct this research from the Agency for Cultural Affairs and the Ministry of the Environment, Japan. The described methods were conducted in compliance with the Regulation on Animal Experimentation at Kyoto University and were approved by the Kyoto University Animal Experimentation Committee. This article does not contain any studies with human participants.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Laboratory of Animal Ecology, Graduate School of Science, Kyoto University, Kitashirakawa-oiwakecho, Sakyo, Kyoto 606-8502, Japan

²Faculty of Agriculture, Kyushu University, Moto-oka, Nishi, Fukuoka 819-0395, Japan

³Department of Biology, Faculty of Medicine, Osaka Medical and Pharmaceutical University, Takatsuki 569-0801, Japan

⁴Gifu World Freshwater Aquarium, 1453 Kawashimakasada-cho, Kakamigahara, Gifu 501-6021, Japan

⁵Bioaffiliation Center, Research Institute of Environment, Agriculture and Fisheries, Osaka Prefecture, 10-4, Koyamoto-machi, Neyagawa, Osaka 572-0088, Japan

⁶Board of Education in Himi City, Himi, Toyama 935-8686, Japan

⁷Lake Biwa Museum, 1091 Oroshimo-cho, Kusatsu, Shiga 525-0001, Japan

⁸Gifu Kyoritsu University, 5-50, Gifu, Kitakata, Ogaki 503-8550, Japan

⁹Present address: Laboratory of Forest Biology, Graduate School of Agriculture, Kyoto University, Kitashirakawa-oiwakecho, Sakyo, Kyoto 606-8502, Japan

Received: 7 April 2024 / Accepted: 21 October 2024

Published online: 12 November 2024

References

1. Jakubínský J, Prokopová M, Raška P, Salvati L, Bezak N, Cudlín O, et al. Managing floodplains using nature-based solutions to support multiple ecosystem functions and services. *WIREs Water*. 2021;8:Article e1545.
2. Salo J, Kalliola R, Häkkinen I, Mäkinen Y, Niemelä P, Puhakka M, et al. River dynamics and the diversity of Amazon lowland forest. *Nature*. 1986;322:254–8.
3. Ward JV, Tockner K, Schiemer F. Biodiversity of floodplain river ecosystems: ecotones and connectivity¹. *Regul Rivers Res Manag*. 1999;15:125–39.
4. Harvolk S, Symmank L, Sundermeier A, Otte A, Donath TW. Human impact on plant biodiversity in functional floodplains of heavily modified rivers: a comparative study along German federal waterways. *Ecol Eng*. 2015;84:463–75.
5. Sievers M, Hale R, Parris KM, Swearer SE. Impacts of human-induced environmental change in wetlands on aquatic animals. *Biol Rev*. 2018;93:529–54.
6. Edwards SV, Shultz AJ, Campbell-Staton SC. Next-generation sequencing and the expanding domain of phylogeography. *Folia Zool*. 2015;64:187–206.
7. Edwards SV, Robin VV, Ferrand N, Moritz C. The evolution of comparative phylogeography: putting the geography (and more) into comparative population genomics. *Genome Biol Evol*. 2022;14:Article evab176.
8. McCormack JE, Hird SM, Zellmer AJ, Carstens BC, Brumfield RT. Applications of next-generation sequencing to phylogeography and phylogenetics. *Mol Phylogenet Evol*. 2013;66:526–38.
9. Nater A, Burri R, Kawakami T, Smeds L, Ellegren H. Resolving evolutionary relationships in closely related species with whole-genome sequencing data. *Syst Biol*. 2015;64:1000–17.
10. Nadachowska-Brzyska K, Konczal M, Babik W. Navigating the temporal continuum of effective population size. *Methods Ecol Evol*. 2022;13:22–41.
11. Kitamura J, Uchiyama R. Bitterling fishes of Japan: natural history and culture. Yama-kei; 2020.

12. Nakamura M. Cyprinid fishes of Japan: studies on the life history of cyprinid fishes of Japan. Research Institute Natural Resources; 1969.
13. International Union for Conservation of Nature. The IUCN Red List of Threatened Species. Version 2022-2. 2022. <https://www.iucnredlist.org>. Accessed 25 Mar 2024.
14. Japan Ministry of the Environment. Publication of the Japanese red list 2020. 2020. <https://www.env.go.jp/press/107905.html>. Accessed 25 Mar 2024.
15. Nishio M, Kawamoto T, Kawakami R, Edo K, Yamazaki Y. Life history and reproductive ecology of the endangered itasenpara bitterling *Acheilognathus longipinnis* (Cyprinidae) in the Himi region, central Japan. *J Fish Biol*. 2015;87:616–33.
16. Kawamura K, Ueda T, Arai R, Smith C. Phylogenetic relationships of bitterling fishes (Teleostei: Cypriniformes: Acheilognathinae), inferred from mitochondrial cytochrome *b* sequences. *Zoolog Sci*. 2014;31:321–9.
17. Uehara K, Kawabata K, Ohta H. Low temperature requirement for embryonic development of Itasenpara bitterling *Acheilognathus longipinnis*. *J Exp Zool Comp Exp Biol*. 2006;305A:823–9.
18. Ogawa R, Nagata Y. Symbolic fish in the river floodplain: Itasenpara bitterling. In: Mori S, editor. Conservation ecology of freshwater organisms: toward restoration ecology. Shinzansha Sci-Tech; 1999. pp. 9–18.
19. Ogawa R, Aya S, Kawai N, Takebayashi H, Takemon Y, Uehara K, et al. Re-introduction of the Itasenpara bitterling to the Yodo River in Osaka Prefecture, Japan. In: Soorae PS, editor. Global re-introduction perspectives 2011: more case studies from around the globe. IUCN/SSC Re-introduction Specialist Group & Environment Agency; 2011. pp. 49–53.
20. Kitanishi S, Nishio M, Sagawa S, Uehara K, Ogawa R, Yokoyama T, et al. Strong population genetic structure and its implications for the conservation and management of the endangered itasenpara bitterling. *Conserv Genet*. 2013;14:901–6.
21. Okazaki T, Watanabe M, Inamura O, Kitagawa T, Tabe M, Nagata Y. Genetic relationships among regional populations of the deep-bodied bitterling, *Acheilognathus longipinnis*, inferred from mitochondrial DNA analysis. *DNA Polymorph*. 2006;14:276–80.
22. Yamazaki Y, Uehara K, Ikeya K, Nishio M. Interpopulational and intrapopulational genetic diversity of the endangered itasenpara bitterling (*Acheilognathus longipinnis*) with reference to its demographic history. *Conserv Genet*. 2020;21:55–64.
23. Ballard JWO, Whitlock MC. The incomplete natural history of mitochondria. *Mol Ecol*. 2004;13:729–44.
24. Emerson KJ, Merz CR, Catchen JM, Hohenlohe PA, Cresko WA, Bradshaw WE, et al. Resolving postglacial phylogeography using high-throughput sequencing. *Proc Natl Acad Sci*. 2010;107:16196–200.
25. Toews DPL, Brelsford A. The biogeography of mitochondrial and nuclear discordance in animals. *Mol Ecol*. 2012;21:3907–30.
26. Peery MZ, Kirby R, Reid BN, Stoeltig R, Doucet-B  er E, Robinson S, et al. Reliability of genetic bottleneck tests for detecting recent population declines. *Mol Ecol*. 2012;21:3403–18.
27. Putman AI, Carbone I. Challenges in analysis and interpretation of microsatellite data for population genetic studies. *Ecol Evol*. 2014;4:4399–428.
28. Weisenfeld NI, Kumar V, Shah P, Church DM, Jaffe DB. Direct determination of diploid genome sequences. *Genome Res*. 2017;27:757–67.
29. Manni M, Berkeley MR, Seppey M, Sim  o FA, Zdobnov EM. BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol*. 2021;38:4647–54.
30. Chang C-H, Li F, Shao K-T, Lin Y-S, Morosawa T, Kim S, et al. Phylogenetic relationships of Acheilognathidae (Cypriniformes: Cyprinoidea) as revealed from evidence of both nuclear and mitochondrial gene sequence variation: evidence for necessary taxonomic revision in the family and the identification of cryptic species. *Mol Phylogenet Evol*. 2014;81:182–94.
31. Chen S, Zhou Y, Chen Y, Gu J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018;34:884–90.
32. Vasimuddin M, Misra S, Li H, Aluru S. Efficient architecture-aware acceleration of BWA-MEM for multicore systems. In: 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS). IEEE; 2019. pp. 314–24.
33. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO et al. Twelve years of SAMtools and BCFtools. *GigaScience*. 2021;10:Article giab008.
34. Auwerda GAVder, O'Connor BD. Genomics in the cloud: using Docker, GATK, and WDL in Terra. O'Reilly Media; 2020.
35. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics*. 2011;27:2156–8.
36. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*. 2015;4:Article 7.
37. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*. 2011;27:2987–93.
38. Jin J-J, Yu W-B, Yang J-B, Song Y, dePamphilis CW, Yi T-S et al. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol*. 2020;21:Article 241.
39. Iwasaki W, Fukunaga T, Isagozawa R, Yamada K, Maeda Y, Satoh TP, et al. MitoFish and MitoAnnotator: a mitochondrial genome database of fish with an accurate and automatic annotation pipeline. *Mol Biol Evol*. 2013;30:2531–40.
40. Weir BS, Cockerham CC. Estimating *F*-statistics for the analysis of population structure. *Evolution*. 1984;38:1358–70.
41. Nei M, Li WH. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci*. 1979;76:5269–73.
42. Korunes KL, Samuk K. Pixy: unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Mol Ecol Resour*. 2021;21:1359–68.
43. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. 2009;19:1655–64.
44. Excoffier L, Lischer HEL. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour*. 2010;10:564–7.
45. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol*. 2006;23:254–67.
46. Ortiz EM. vcf2phylip v2.0: Convert a VCF matrix into several matrix formats for phylogenetic analysis. 2019. <https://doi.org/10.5281/zenodo.2540861>. Accessed 1 Apr 2024.
47. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol*. 2020;37:1530–4.
48. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermini LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 2017;14:587–9.
49. Leach   AD, Banbury BL, Felsenstein J, Oca A, de nieto-Montes A. Short tree, long tree, right tree, wrong tree: new acquisition bias corrections for inferring SNP phylogenies. *Syst Biol*. 2015;64:1032–47.
50. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*. 2010;59:307–21.
51. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol*. 2018;35:518–22.
52. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30:772–80.
53. Bouckaert R, Vaughan TG, Barido-Sottani J, Duch  ne S, Fourment M, Gavryushkina A, et al. BEAST 2.5: an advanced software platform for bayesian evolutionary analysis. *PLOS Comput Biol*. 2019;15:Article e1006650.
54. Bouckaert RR, Drummond AJ. bModelTest: bayesian phylogenetic site model averaging and model comparison. *BMC Evol Biol*. 2017;17:Article 42.
55. Yasuno T. Fossil pharyngeal teeth of the Rhodinae fish from the Miocene Katabira formation of the Kani Group, Gifu Prefecture, Japan. *Bull Mizunami Foss Mus*. 1984;11:101–5.
56. Betancur -RR, Ort   G, Pyron RA. Fossil-based comparative analyses reveal ancient marine ancestry erased by extinction in ray-finned fishes. *Ecol Lett*. 2015;18:441–50.
57. Chen W-J, Lavou   S, Mayden RL. Evolutionary origin and early biogeography of otophysan fishes (Ostariophysi: Teleostei). *Evolution*. 2013;67:2218–39.
58. Kusuma WE, Ratmuangkhwang S, Kumazawa Y. Molecular phylogeny and historical biogeography of the Indonesian freshwater fish *Rasbora lateristriata* species complex (Actinopterygii: Cyprinidae): cryptic species and west-to-east divergences. *Mol Phylogenet Evol*. 2016;105:212–23.
59. Tang KL, Lumbantobing DN, Mayden RL. The phylogenetic placement of *Oxygaster* Van Hasselt, 1823 (Teleostei: Cypriniformes: Cyprinidae) and the taxonomic status of the family-group name Oxygastrinae Bleeker, 1860. *Copeia*. 2013;2013:13–22.
60. Wang X, Gan X, Li J, Mayden RL, He S. Cyprinid phylogeny based on bayesian and maximum likelihood analyses of partitioned data: implications for Cyprinidae systematics. *Sci China Life Sci*. 2012;55:761–73.
61. Kumar S, Suleski M, Craig JM, Kasprzewicz AE, Sanderford M, Li M et al. TimeTree 5: an expanded resource for species divergence times. *Mol Biol Evol*. 2022;39:Article msac174.

62. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in bayesian phylogenetics using Tracer 1.7. *Syst Biol*. 2018;67:901–4.
63. Li H, Durbin R. Inference of human population history from individual whole-genome sequences. *Nature*. 2011;475:493–6.
64. Schiffels S, Durbin R. Inferring human population size and separation history from multiple genome sequences. *Nat Genet*. 2014;46:919–25.
65. Boitard S, Rodríguez W, Jay F, Mona S, Austerlitz F. Inferring population size history from large samples of genome-wide molecular data: an approximate bayesian computation approach. *PLOS Genet*. 2016;12:Article e1005877.
66. Schiffels S, Wang K, MSMC, MSMC2. The multiple sequentially markovian coalescent. *Statistical Population Genomics*. New York, NY: Humana; 2020. pp. 147–66.
67. Delaneau O, Marchini J, Zagury J-F. A linear complexity phasing method for thousands of genomes. *Nat Methods*. 2012;9:179–81.
68. Delaneau O, Howie B, Cox AJ, Zagury J-F, Marchini J. Haplotype estimation using sequencing reads. *Am J Hum Genet*. 2013;93:687–96.
69. Pockrandt C, Alzamel M, Iliopoulos CS, Reinert K, GenMap. Ultra-fast computation of genome mappability. *Bioinformatics*. 2020;36:3687–92.
70. Malinsky M, Svardal H, Tyers AM, Miska EA, Genner MJ, Turner GF, et al. Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nat Ecol Evol*. 2018;2:1940–55.
71. Excoffier L, Marchi N, Marques DA, Matthey-Doret R, Gouy A, Sousa VC. *fastsimcoal2*: Demographic inference under complex evolutionary scenarios. *Bioinformatics*. 2021;37:4882–5.
72. Meier JJ, Sousa VC, Marques DA, Selz OM, Wagner CE, Excoffier L, et al. Demographic modelling with whole-genome data reveals parallel origin of similar *Pundamilia* cichlid species after hybridization. *Mol Ecol*. 2017;26:123–41.
73. Lisiecki LE, Raymo ME. A pliocene-pleistocene stack of 57 globally distributed benthic $\delta^{18}\text{O}$ records. *Paleoceanography*. 2005;20:Article PA1003.
74. Firrreno TJ, O'Neill JR, Portik DM, Emery AH, Townsend JH, Fujita MK. Finding complexity in complexes: assessing the causes of mitonuclear discordance in a problematic species complex of mesoamerican toads. *Mol Ecol*. 2020;29:3543–59.
75. Nagayama S, Oota M, Fujita T, Kitamura J, Minamoto T, Mori S, et al. Autumn dispersal and limited success of reproduction of the deepbody bitterling (*Acheilognathus longipinnis*) in terrestrialized floodplain. *Knowl Manag Aquat Ecosyst*. 2022;423:4.
76. Kitazima J, Matsuda M, Mori S, Kokita T, Watanabe K. Population structure and cryptic replacement of local populations in the endangered bitterling *Acheilognathus cyanostigma*. *Ichthyol Res*. 2015;62:122–30.
77. Miyake T, Nakajima J, Umemura K, Onikura N, Ueda T, Smith C, et al. Genetic diversification of the Kanehira bitterling *Acheilognathus rhombeus* inferred from mitochondrial DNA, with comments on the phylogenetic relationship with its sister species *Acheilognathus barbatulus*. *J Fish Biol*. 2021;99:1677–95.
78. Tominaga K, Nagata N, Kitamura J, Watanabe K, Sota T. Phylogeography of the bitterling *Tanakia lanceolata* (Teleostei: Cyprinidae) in Japan inferred from mitochondrial cytochrome *b* gene sequences. *Ichthyol Res*. 2020;67:105–16.
79. Watanabe K, Tominaga K, Nakajima J, Kakioka R, Tabata R. Japanese freshwater fishes: Biogeography and cryptic diversity. In: Motokawa M, Kajihara H, editors. *Species diversity of animals in Japan*. Springer Japan; 2017. pp. 183–227.
80. Taniguchi S, Bertl J, Futschik A, Kishino H, Okazaki T. Waves out of the Korean Peninsula and inter- and intra-species replacements in freshwater fishes in Japan. *Genes*. 2021;12:303.
81. Ono Y. Last glacial paleoclimate reconstructed from glacial and periglacial landforms in Japan. *Geogr Rev Jpn Ser B*. 1984;57:87–100.
82. Tsukada M. Vegetation and climate during the last glacial Maximum in Japan. *Quat Res*. 1983;19:212–35.
83. Meyers PA, Takemura K, Horie S. Reinterpretation of late quaternary sediment chronology of Lake Biwa, Japan, from correlation with marine glacial-interglacial cycles. *Quat Res*. 1993;39:154–62.
84. Satoguchi Y. Geological history of Lake Biwa. In: Kawanabe H, Nishino M, Maehata M, editors. *Lake Biwa: interactions between nature and people*. Springer; 2012. pp. 9–16.
85. Ito Y, Oguchi T, Masuda F. Late quaternary depositional sequences and landforms in relation to sea-level changes in the Osaka intra-arc basin, Japan: a borehole database analysis. *Quat Int*. 2018;471:298–317.
86. Yasuhara M, Irizuki T, Yoshikawa S, Nanayama F. Holocene sea-level changes in Osaka Bay, western Japan: Ostracode evidence in a drilling core from the southern Osaka Plain. *J Geol Soc Jpn*. 2002;108:633–43.
87. Rodríguez W, Mazet O, Grusea S, Arredondo A, Corujo JM, Boitard S, et al. The IICR and the non-stationary structured coalescent: towards demographic inference with arbitrary changes in population structure. *Heredity*. 2018;121:663–78.
88. Jin H-J, Kwak K-D, Hammer MF, Nakahori Y, Shinka T, Lee J-W, et al. Y-chromosomal DNA haplogroups and their implications for the dual origins of the Koreans. *Hum Genet*. 2003;114:27–35.
89. Ono A, Sato H, Tsutsumi T, Kudo Y. Radiocarbon dates and archaeology of the late Pleistocene in the Japanese islands. *Radiocarbon*. 2002;44:477–94.
90. Pope KO, Terrell JE. Environmental setting of human migrations in the circum-pacific region. *J Biogeogr*. 2008;35:1–21.
91. Saito O, Takashima M. Population, urbanisation and farm output in early modern Japan, 1600–1874: a review of data and benchmark estimates. *RCESR Discuss Pap Ser*. 2015;DP15–3.
92. Naito T, Nakayama K, Takeshima H, Hashiguchi Y, Akita T, Yamasaki YY, et al. The detailed population genetic structure of the rare endangered latid fish akame *Lates japonicus* with extremely low genetic diversity revealed from single-nucleotide polymorphisms. *Conserv Genet*. 2023;24:523–35.
93. Lynch M, Gabriel W. Mutation load and the survival of small populations. *Evolution*. 1990;44:1725–37.
94. Lynch M, Conery J, Burger R. Mutation accumulation and the extinction of small populations. *Am Nat*. 1995;146:489–518.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.