

Multiple waves of freshwater colonization of the three-spined stickleback in the Japanese Archipelago

Ryo Kakioka

University of the Ryukyus

Seiichi Mori

Gifu Kyoritsu University

Tomoyuki Kokita

Fukui Prefectural University

Takuya K Hosoki

National Institute of Genetics

Atsushi J Nagano

Ryukoku University

Asano Ishikawa

National Institute of Genetics

Manabu Kume

Kyoto University

Atsushi Toyoda

National Institute of Genetics

Jun Kitano (✉ jkitano@nig.ac.jp)

National Institute of Genetics <https://orcid.org/0000-0001-8659-5698>

Research article

Keywords: Restriction-site Associated DNA Sequencing, Convergent evolution, Glacial relic, Interglacial refugia, Non-native population, Hybridization, Speciation

Posted Date: October 30th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-59443/v3>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published on November 3rd, 2020. See the published version at <https://doi.org/10.1186/s12862-020-01713-5>.

Abstract

Background: The three-spined stickleback (*Gasterosteus aculeatus*) is a remarkable system to study the genetic mechanisms underlying parallel evolution during the transition from marine to freshwater habitats. Although the majority of previous studies on the parallel evolution of sticklebacks have mainly focused on postglacial freshwater populations in the Pacific Northwest of North America and northern Europe, we recently use Japanese stickleback populations for investigating shared and unique features of adaptation and speciation between geographically distant populations. However, we currently lack a comprehensive phylogeny of the Japanese three-spined sticklebacks, despite the fact that a good phylogeny is essential for any evolutionary and ecological studies. Here, we conducted a phylogenomic analysis of the three-spined stickleback in the Japanese Archipelago.

Results: We found that freshwater colonization occurred in multiple waves, each of which may reflect different interglacial isolations. Some of the oldest freshwater populations from the central regions of the mainland of Japan (hariyo populations) were estimated to colonize freshwater approximately 170,000 years ago. The next wave of colonization likely occurred approximately 100,000 years ago. The inferred origins of several human-introduced populations showed that introduction occurred mainly from nearby habitats. We also found a new habitat of the three-spined stickleback sympatric with the Japan Sea stickleback (*Gasterosteus nipponicus*).

Conclusions: These Japanese stickleback systems differ from those in the Pacific Northwest of North America and northern Europe in terms of divergence time and history. Stickleback populations in the Japanese Archipelago offer valuable opportunities to study diverse evolutionary processes in historical and contemporary timescales.

Background

The presence of phylogenetically independent lineages adapting to similar environments offers great opportunities to investigate the roles of natural selection in phenotypic evolution [1]. Furthermore, such replicate systems enable us to investigate the extent to which causative alleles and genes are shared among independent lineages adapting to similar environments and what factors determine the probabilities of sharing the same alleles and genes [2-5]. Such knowledge will help to understand the repeatability and predictability of evolution [2-5]. Although several researchers distinguish between parallel and convergent evolution based on the underlying genetic mechanisms with the former caused by the same genetic mechanisms and the latter by different mechanisms, we call both parallel evolution in this study, because it is often difficult to draw a clear line between them [6].

Because transition from marine to freshwater habitats occurred in multiple lineages [7, 8], we can find replicate pairs of closely related marine and freshwater organisms. Marine and freshwater environments differ in many biotic and abiotic factors. Therefore, phylogenetically independent lineages that achieved

the marine–freshwater transition would offer great opportunities to investigate the genetic basis for parallel/convergent evolution accompanying freshwater colonization and adaptation [7, 8].

Among the organisms that have undergone the marine–freshwater transition, the three-spined stickleback (*Gasterosteus aculeatus*) are a remarkable system to study the genetic mechanisms underlying this transition [9-11]. The three-spined stickleback is a cold-water fish widely distributed in coastal marine, brackish, and freshwater habitats of the Northern hemisphere [12, 13]. Ancestral marine ecotypes of the three-spined stickleback colonized freshwater habitats across its distribution. Many of these habitats emerged following deglaciation during the Quaternary Period. Freshwater populations from different geographic regions often show similar morphology and physiology. Thus, the three-spined stickleback is an excellent system to investigate the genetic mechanisms underlying parallel evolution [9, 10, 12-17].

Previous genetic studies on the parallel evolution of sticklebacks have mainly focused on postglacial freshwater populations in the Pacific Northwest of North America and in northern Europe [9-11]. The habitats in these regions were covered by ice sheets during the last glacial period and became uncovered within the last 12,000 years. Parallel evolution of several morphological and physiological traits in these postglacial populations has been caused by repeated fixation of identical-by-descent alleles [18-20]. Freshwater environments select freshwater-adaptive alleles that previously existed as standing variations in the founding marine populations [14, 18, 20-22], whose standing allelic variation may be maintained by gene flow from another freshwater population [9]. However, cases in which independent mutations of the same genes or different genes underlie parallel evolution have been described [10, 14, 15, 20, 21, 23].

Recent studies have demonstrated that geographically distant lineages, such as East Pacific and Atlantic lineages, use different sets of standing genetic variations for parallel evolution [21, 23]. These results indicate that analysis of geographically diverse regions can help to understand the wide distribution of freshwater-adaptive alleles in *G. aculeatus* across its distribution [21]. Such analyses can also clarify the alternative solutions when standing variations are not available [24, 25].

Japanese three-spined stickleback populations in the western Pacific basin offer several unique opportunities to investigate the genetic basis of parallel evolution (Fig. 1A). First, the Japanese Archipelago is geographically distant from North America and Europe, suggesting that the Japanese populations may share a relatively small number of genetic variants with North American and European populations. Previous studies have shown that reduction in the armor plate in freshwater populations in North America and Europe is caused by repeated fixation of the same *ectodysplasin* (*Eda*) allele, whereas armor plate reduction in Japanese freshwater populations is caused by independent mutations at *Eda* [9, 14, 24, 26].

Second, there are freshwater populations with different ages of colonization. The Japanese Archipelago was not covered by ice sheets in the Quaternary glaciation, suggesting that several freshwater habitats were accessible by sticklebacks well before 12,000 years ago. A previous mitochondrial DNA phylogenetic analysis estimated the divergence time of freshwater populations in Gifu and Shiga, central

Honshu Island, termed “hariyo stickleback” in Japan [27, 28], from the rest of *G. aculeatus* as 0.37–0.43 million years before present (Ma BP) based on a molecular clock. Additionally, there are several young freshwater populations, *e.g.* those inhabiting lakes and ponds that were formed within 2,000–3,000 years in eastern Hokkaido [29, 30]. These freshwater populations are not genetically differentiated from marine *G. aculeatus* at allozyme or microsatellite loci [29, 30]. Several human-introduced populations also offer opportunities to investigate the genetic basis of rapid adaptation [31, 32]. Freshwater populations with such a diverse array of colonization ages provide opportunities to investigate how freshwater adaptation progresses over time.

Finally, the distribution range of *G. aculeatus* overlaps with that of its sister species *G. nipponicus* in northern Japan [33, 34]. Previous studies have shown that all freshwater populations examined thus far belong to *G. aculeatus* rather than *G. nipponicus* [15, 35]. *G. aculeatus* has higher copy numbers of the metabolic gene *Fads2* and can survive better on freshwater-derived diets than *G. nipponicus* [15]. Because there is past and ongoing hybridization between these two species [32, 36–38], it is important to determine the extent of introgression of freshwater-adaptive alleles between these two species to understand the genetic factors constraining the freshwater colonization of *G. nipponicus*.

As a first step towards a comprehensive understanding of the genetic basis of parallel evolution in the Japanese freshwater populations of *Gasterosteus*, we investigated their origins using phylogenomic approaches. The majority of previous phylogenetic studies on Japanese sticklebacks have used allozyme, microsatellite, and mitochondrial DNA. Mitochondrial DNA has been shown to introgress from *G. nipponicus* to *G. aculeatus*, suggesting that phylogeny based on mitochondrial DNA does not reflect the population history [37, 39, 40]. Previous phylogenetic analyses using allozyme and microsatellite were based on a small number of markers. More precise phylogenetic analysis with a large number of genome-wide single nucleotide polymorphisms (SNPs) are necessary. We have conducted phylogenetic analyses using whole genome sequences [41] and Restriction-site associated DNA (RAD) markers [15]. However, we have identified several new habitats of freshwater populations and new possible hybrid zones between *G. aculeatus* and *G. nipponicus* since then. Additionally, previous studies did not investigate the divergence time or phylogenetic relationships of the Japanese populations with North American and European populations. To solve these unanswered questions, we conducted a phylogenomic analysis of Japanese stickleback populations using RAD sequencing.

Results

Population structure

Two clusters revealed by ADMIXTURE analysis of all samples from the Japanese Archipelago at $K = 2$ reflected interspecies differentiation between *G. aculeatus* and *G. nipponicus* (Fig. 1B). All freshwater populations, including the artificially introduced non-native ones, were assigned to the *G. aculeatus* cluster. Although *G. aculeatus* and *G. nipponicus* were overall genetically differentiated, hybrids were also found at several localities. If we judge fish with Q values (admixture proportion of the ADMIXTURE

analysis) < 0.875 for either species as hybrids, such hybrids were mostly found in marine populations, although freshwater populations in Otsuchi (FW Fureai [population code: F6], FW Mast [F9]) and a non-native population, FW Kussharo (FN1), also contained hybrids.

Increasing K identified more freshwater clusters and revealed genetic distinctiveness among freshwater populations. At $K = 3$, the hariyo stickleback, FW Tsuya (Gifu; F15) and FW Shiga (F16), separated from other *G. aculeatus* populations. Two introduced populations (FW Komono [FN10], FW Kobe [FN11]) were assigned to this cluster. At $K = 4$, several freshwater populations, Aizu populations (FW Hakusan [F10], FW Inawashiro [F11], FW Kitakata [F12]), FW Nasu (F13), and FW Ono (F14), separated as a cluster from the rest of the native *G. aculeatus* populations. At $K = 5$, populations from Aizu (FW Hakusan [F10], FW Inawashiro [F11], FW Kitakata [F12]) further separated from this cluster, to which non-native populations of FW Uono (FN6), FW Nikko (FN7), and FW Kinu (FN8) also belonged. Individuals of FW Nasu (F13) formed a distinct cluster together with the non-native FW Hitachi (FN9) population at $K = 7$ and 9. The cross-validation error was the lowest at $K = 8$ (Supplementary Fig. S4). At $K \geq 6$, *G. nipponicus* contained distinct clusters, but the clustering was incongruent among different K .

Results of the ADMIXTURE analyses were supported by principal component analyses (PCA) (Supplementary Fig. S2). PC1 separated *G. aculeatus* and *G. nipponicus*. The hybrids identified in the ADMIXTURE analysis were placed between *G. aculeatus* and *G. nipponicus*. PC2 separated the hariyo stickleback, FW Tsuya (Gifu; F15) and FW Shiga (F16), from the other populations. The freshwater population of FW Nasu (F13) was distinct along the PC3 axis. PC4 splits the Aizu populations (FW Hakusan [F10], FW Inawashiro [F11], FW Kitakata [F12]), FW Ono (F14), and others. Grouping of non-native populations to native populations was concordant with the assignment of the ADMIXTURE clustering.

Phylogeny

Maximum likelihood (ML) phylogeny using concatenated SNPs of the Japanese samples clearly distinguished *G. nipponicus* and *G. aculeatus* (Fig. 2). The hariyo stickleback were monophyletic and first branched off from the rest of *G. aculeatus*. Another monophyletic clade composed of populations from the Aizu Basin (FW Hakusan [F10], FW Inawashiro [F11], FW Kitakata [F12]), FW Nasu (F13), and FW Ono (F14) split from the rest of *G. aculeatus*. Other freshwater populations were not monophyletic and nested in marine populations. The placement of non-native populations was congruent with the results of the population structure analyses. Non-native populations from adjacent sites often clustered together. These comprised FW Shikotsu (FN2) and FW Nishitappu (FN3); FW Aisaka (FN4) and FW Towada (FN5); and FW Uono (FN6), FW Nikko (FN7), and FW Kinu (FN8).

ML phylogenetic analysis including the samples from the western and eastern basins of the Pacific and northern Europe (Fig. 3) also supported the monophyly of the hariyo stickleback, which split from the rest of *G. aculeatus* earlier than any other freshwater populations examined (Fig. 3B). Next, populations from

the East Pacific and Europe branched off. All of the Japanese *G. aculeatus* populations other than the hariyo stickleback were monophyletic, although the bootstrap support was low (bootstrap value < 60%).

The topology of the species tree obtained by the SNAPP analyses was identical among all the runs and were generally congruent with the ML trees (Fig. 4, Supplementary Fig. S5). Most of the nodes were strongly supported with posterior probabilities of > 0.93, except for one (see the node with 0.76 in Fig. 4). The divergence times of each node agreed well among the runs with the same prior on divergence time and were scalable to the root divergence time with different priors. Assuming a root divergence of 680 thousand years (ka) before present (BP), which was estimated by a demographic analysis with an Approximate Bayesian Computation approach [37], the mean divergence time between *G. aculeatus* and *G. nipponicus* was estimated to be 644–653 ka (95% highest posterior density intervals [95HDI] = 395–868 ka) (for the results of other runs, see Supplementary Fig. S5B and S5C). For the results assuming divergence at 1.38 million years (Ma) BP, see Supplementary Fig. S5D–F. Hariyo stickleback diverged from the rest of *G. aculeatus* at 167–169 ka (95HDI = 114–237 ka) BP. Two lineages within the hariyo stickleback, FW Tsuya (Gifu; F15) and FW Shiga (F16), diverged at 97–99 ka (95HDI = 62–140 ka) BP. The divergence time of a Japanese freshwater population FW Nasu (F13) from the rest was 104–106 ka (95HDI = 65–148 ka), while the freshwater lineage leading to FW Hakusan (Aizu; F10) and FW Ono (F14) diverged at 96–100 ka (95HDI = 59–144 ka) BP. FW Chimikeppu (F1) diverged at 55 ka (95HDI = 35–79 ka) BP, while a younger Japanese freshwater population from Otsuchi (FW Gensui 2010 [F7]) diverged at 24 ka (95HDI = 14–34 ka) BP.

As for the North American and European populations, the divergence of an East Pacific freshwater population (Little Campbell Stream) from the rest occurred at 128–129 ka (95HDI = 76–182 ka) BP. European and Pacific populations excluding this old Pacific freshwater lineage diverged 72–73 ka (95HDI = 45–99 ka) BP. The East Pacific marine population from the estuary of Little Campbell River diverged from the Japanese Pacific Ocean marine population 38–39 ka (95HDI = 22–53 ka) BP. North European freshwater (Grosser Ploener See) and marine (Lemvig) populations diverged 13 ka (95HDI = 7–19 ka) BP.

Discussion

Multiple waves of freshwater colonization in the Japanese three-spined sticklebacks

Our results based on genome-wide SNPs with new additional populations support previous findings that all freshwater populations in the Japanese Archipelago are within the *G. aculeatus* clade [15, 34, 35]. Furthermore, our present phylogenetic analysis showed that the Japanese freshwater populations are not monophyletic, suggesting that freshwater colonization has occurred in multiple waves.

Freshwater populations called hariyo sticklebacks are the oldest extant freshwater lineages of the species reported thus far. Phylogenetic analyses revealed the monophyly of the hariyo sticklebacks. Bayesian species tree analysis showed that the divergence of the hariyo sticklebacks from the rest of *G. aculeatus* was approximately 167–169 ka BP, which largely predates the last glacial period and is the oldest extant freshwater lineage ever reported. The eastern Pacific basin harbors old freshwater lineages [42, 43] and

the present data confirmed that the divergence time of a stream population from the eastern Pacific (Little Campbell River) predates the end of the last glacial period. A previous study using a SNAPP species tree analysis based on the same calibration point with similar priors [42] estimated the divergence time of another freshwater population from the East Pacific basin (Beaver Lake on Vancouver Island) as 119 ka BP, which is close to our estimate of the divergence time of the eastern Pacific freshwater population. Nonetheless, the divergence of the hariyo stickleback lineage preceded that of the eastern Pacific stream populations. To date, no previous phylogenetic analysis at the global scale using genome-wide SNP data (e.g., [18, 43]) have included the hariyo lineage.

The freshwater lineages of Nasu (FW Nasu [F13]), Aizu (FW Hakusan [F10], FW Inawashiro [F11], FW Kitakata [F12]), and Ono (FW Ono [F14]) were estimated to have diverged at approximately 100 ka BP. This is still before the latest Pacific–Atlantic split, which has been suggested to have occurred when the Bering Strait closed somewhere between 34 and 75 ka BP during the last glacial period [44]. Other freshwater lineages in northern Japan have diverged more recently. FW Gensui 2010 (F7) from Otsuchi was estimated to have diverged at approximately 24 ka BP, which is close to the time of postglacial freshwater colonization in northern Europe [42].

Interglacial isolations can explain some of these multiple waves of freshwater colonization in the Japanese three-spined stickleback. Sticklebacks favor a cooler climate [12, 45], so they would shift the distribution southward during glacial periods and northward during interglacial periods [41]. Freshwater populations in central Honshu Island are presently restricted to springs and spring-fed streams in which water temperature is maintained below 20°C, allowing the fish to avoid heat in summer [28, 45]. Habitats of the hariyo stickleback and the Nasu population are on the Pacific slope, out of the current distribution range of marine *G. aculeatus*. The waters inhabited by the Aizu and Ono populations are drained by the rivers that flow into the Sea of Japan, where *G. aculeatus* is absent at present (Fig. 1A). In addition to global cooling, southward extension of the cold ocean current in the Pacific Ocean 12.8–21 ka BP [46], shut-off of the warm Tsushima Current from the East China Sea into the Sea of Japan during glacial periods [47], and the intrusion of the cold Oyashio Current into the Sea of Japan through the Tsugaru Strait 4.8–17.5 ka BP [48] could have shifted the range of marine *G. aculeatus* southward during the last and preceding glacial periods. The collective data support the hypothesis that these freshwater lineages are glacial relicts originating from ancient marine *G. aculeatus* that once shifted its distribution southward during the glacial periods. California in the eastern Pacific basin also houses isolated freshwater populations, which may have colonized before the last glacial period [49, 50]. Some freshwater populations from southern Europe may also be glacial relicts [42, 43, 51–55]. Therefore, freshwater colonization and subsequent isolation in the glacial–interglacial cycles likely have come in multiple waves at multiple geographical regions across the distribution range of the three-spined stickleback.

Although the hariyo sticklebacks may be the oldest extant freshwater lineage, fossils of *Gasterosteus* from both the eastern and western basins of the Pacific date back to 10 Ma BP [56–58]. This suggests that *Gasterosteus* flourished around the Pacific, including fresh waters, since at least 10 Ma BP [57].

These fossils largely predate the divergence of the hariyo lineage of *G. aculeatus* and even the split of *G. aculeatus* and *G. nipponicus* [37]. Although these ancient freshwater *Gasterosteus* are not direct ancestors of the extant freshwater populations of *G. aculeatus*, they may have served as sources of standing variation of freshwater-adaptive alleles that have facilitated freshwater adaptation in extant *G. aculeatus* [9, 18]. The ancient age (average of 6.4 Ma) of several freshwater-adaptive alleles segregating in extant *G. aculeatus* [59] is consistent with this idea. Analysis of standing genetic variation of these freshwater-adaptive alleles in the Japanese marine and freshwater populations will provide insights into how widely freshwater-adaptive alleles are shared among global populations in *Gasterosteus* and what genetic mechanisms have enabled freshwater adaptation in parallel.

Non-native populations

Recent human activities have moved sticklebacks from original habitats to non-native habitats. Our genetic analysis showed that the introduced populations were derived from nearby habitats. For example, non-native populations from FW Komono (FN10) and FW Kobe (FN11) clustered with the nearby FW Tsuya (F15). All of these are located in southwestern Honshu Island. Non-native populations from northern Japan in Hokkaido (FW Kussharo [FN1], FW Shikotsu [FN2], and FW Nishitappu [FN3]) and northern Honshu Islands (FW Aisaka [FN4] and FW Towada [FN5]) were genetically similar to *G. aculeatus* distributed in northern Japan. Non-native populations from central Honshu (FW Uono [FN6], FW Nikko [FN7], FW Kinu [FN8], and FW Hitachi [FN9]) were derived from either Aizu or Nasu populations. Although non-native populations can provide opportunities to study the process of adaptation to novel habitats on a contemporary timescale [31, 32], their spread may lead to hybridization with, or extinction, of native populations [28]. Native freshwater populations are invaluable genetic resources to study the genetic basis of adaptive phenotypic diversification generated during the last 200,000 years in the Japanese Archipelago. Thus, it is important to conserve them. Particular caution is needed to prevent translocation of sticklebacks between water systems, which can lead to genetic contamination or even population extinction due to hybridization [60].

A new sympatric habitat

In addition to previously reported sympatric habitats [30, 34, 36-38, 61, 62], we identified a new sympatric habitat of *G. aculeatus* and *G. nipponicus* at the eastern end of Hokkaido (Okinebe [Oki]). Based on the Q values of the ADMIXTURE analysis, among 32 fish analyzed, two individuals were F_1 hybrids and one was a backcross to *G. nipponicus*. Okinebe Pond is relatively small (approximately 30,000 m²) and is connected to the Pacific Ocean by a short stream approximately 200 m in length. The frequency of hybrids in this pond is relatively high compared to previously investigated sympatric habitats. Previous genomic studies have shown that sympatric habitats can differ in the magnitude of reproductive isolation and hybridization [37, 38]. This new sympatric habitat would provide an additional study system to investigate the genetic and ecological mechanisms underlying reproductive isolation between these two species.

Conclusions

Stickleback populations in the Japanese Archipelago offer valuable opportunities to study a wide spectrum of evolutionary processes in historical and contemporary timescales. First, Japanese freshwater populations provide phylogenetically independent and geographically distant replicates of stickleback freshwater populations. Using these systems, we can test the extent to which causative alleles and genes are shared among independent lineages adapting to similar environments and what factors determine the probabilities of sharing the same alleles and genes [2-5]. Second, several newly identified non-native populations will provide us opportunities to investigate the genetic and ecological mechanisms underlying rapid evolution [63]. Finally, replicates of sympatric habitats of *G. aculeatus* and *G. nipponicus* enable us to test whether the same genomic loci are resistant to introgression or likely to introgress between closely related species [38]. By characterizing these loci, we can obtain insights into the genomic patterns of divergence and introgression during speciation with gene flow [64, 65]. In conclusion, Japanese stickleback populations provide a valuable system to study the genetic basis of adaptation and speciation.

Methods

Sample collection

All sticklebacks were collected with seine nets and minnow traps as described previously [20, 30-32, 37, 66, 67] (Fig. 1A). After euthanasia with an overdose of MS-222 (0.5 g/L), the pectoral fins were dissected out and preserved in 99% ethanol until use. Supplementary Tables S1 and S2 provide details of the samples. Morphologically identified species [33] collected at the same locality were denoted as different populations, with the exception of Okinebe (Oki), where *G. aculeatus*, *G. nipponicus*, and possibly their hybrids are supposed to be included.

Laboratory experiments and sequencing

Genomic DNA was isolated using a DNeasy Blood & Tissue Kit (QIAGEN, Valencia, CA, USA). Double digest RAD sequencing (ddRAD-seq) was performed as described previously [68]. Briefly, 10 ng of genomic DNA was digested with EcoRI and BglII, followed by adapter ligation and amplification with uniquely barcoded primers. The libraries were run on HiSeq 2000 or 2500 using the 50 bp single-end or a 100 bp paired-end mode at Macrogen (Kyoto, Japan) or the Advanced Genomics Center of the National Institute of Genetics (Shizuoka, Japan). The sequence data are available from DDBJ/EMBL/EBI/NCBI Sequence Read Archive (DRA010673). Some of the ddRAD-seq data has been published previously [15] (see Table S2).

Additionally, we used publicly available whole genome sequence (WGS) data (Table S2). For *G. aculeatus* collected from PO Akkeshi (P4), *G. nipponicus* from JS Akkeshi (J4), and *Gasterosteus wheatlandi*, we used the previously reported whole genome sequences [69]. Sequence data of *G. aculeatus* from FW Aisaka (FN4) and FW Towada (FN5) were derived from a previous study [32]. For *G. aculeatus* from

northern Europe, the sequences of two randomly selected samples from the marine population reported previously [70], and those of one or two randomly selected samples from each freshwater population reported previously [71] were obtained.

Sequence data processing

The flow of bioinformatic analyses is summarized in Fig. S1. Trimming of ddRAD-seq reads was performed to remove adapter sequences and failed reads using Trimmomatic v0.39 [72] with the following parameters: “ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10:2 CROP:50 LEADING:3 TRAILING:3 MINLEN:50”. The trimmed reads were mapped to the BROADS S1 stickleback reference genome sequence (soft-masked, Ensembl 99) using NextGenMap v0.5.5 [73]. Variants were called with FreeBayes v1.3.2 [74], skipping sites with the average coverage per sample exceeding 500 and with the options: “–report-monomorphic –use-mapping-quality –use-best-n-alleles 8”. Sites of a sample with a coverage of less than five were discarded with BCFtools v1.9 [75].

We further selected RAD loci with the following criteria using BCFtools and bedtools v2.17.1 [76]. First, the sites genotyped in less than 25% of the samples and located on the mitochondrion were excluded. Next, we searched for the regions consecutively genotyped for at least 40 bp, allowing gaps not longer than 10 bp. The records within the identified RAD regions were extracted and variant representations were normalized with vt v0.5772 [77].

WGS reads were trimmed using Trimmomatic with the following settings: “ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10:2:true LEADING:3 TRAILING:3 MINLEN:30”. Overlapped paired-end reads were merged with PEAR 0.9.10 [78] and read pairing was confirmed with fastq-pair v1.0 [79]. Mate-pair reads from Feulner et al. [70] were reversed and complemented using SeqKit 0.10.0 [80]. The reads were mapped to the BROADS S1 stickleback reference sequence using NextGenMap v0.5.5. The maximum insert size for the alignments of the mate-pair reads was set to 6000. Duplicate reads were marked with Picard Tools v2.21.8 [81]. Variants within the selected contiguous RAD loci (see above) were called with FreeBayes using the same settings as that of ddRAD-seq. Sites of a sample with a coverage of less than five were discarded with BCFtool, and normalization of variants was conducted with vt.

The pre-processed variant calls from ddRAD-seq and WGS were merged using BCFtools. Block substitutions were decomposed into their constituent SNPs using vt. Indels, invariant sites, and sites on the sex chromosomes of *G. aculeatus* and *G. nipponicus* (Chromosomes IX and XIX) or those in masked regions or on ambiguous nucleotides in the reference sequence were discarded. Samples with excessively missing genotypes (> 80%) were excluded with BCFtools. This process resulted in a dataset of 97145 SNPs genotyped in a total of 310 samples.

Population structure analyses

In order to investigate genetic differentiation and potential introgression among the stickleback populations in the Japanese Archipelago, we first used a model-based likelihood clustering algorithm

implemented in ADMIXTURE v1.3.0 [82]. We selected biallelic SNPs that were genotyped in all populations with only one missing population allowed, and that were missing in less than 30% of the overall samples with VCFtools v 0.1.17 [83]. If an allele at a SNP site was found in only one sample, the SNP site was excluded regardless of whether it was identified as “singleton” or “doubleton” with VCFtools. The SNPs were subsampled with VCFtools to maintain a minimum distance of 1 kb to reduce the effect of linkage between SNPs. The input file for ADMIXTURE including 2735 SNPs was created using PLINK v1.90 [84]. ADMIXTURE was run by varying the number of evolutionary clusters K from one to nine. The results were summarized and visualized using CLUMPAK [85] on the web (<http://clumpak.tau.ac.il/index.html>).

We also conducted principal component analyses (PCA), using the adegenet v2.1.1 package [86, 87] of R [88]. The dataset for the ADMIXTURE analysis was further filtered, keeping SNPs with minor allele frequency ≥ 0.03 and individuals with missing genotypes $< 20\%$. This resulted in a dataset of 813 SNPs.

Phylogenetic analyses

Maximum likelihood (ML) phylogenetic trees were constructed with RAxML-NG v0.9.0 [89] based on two datasets of concatenated SNPs. The first includes 1,919 SNPs of all the samples from Japan excluding putative recent hybrids between *G. aculeatus* and *G. nipponicus*, which would violate basic assumptions of phylogenetic reconstruction methods and bias tree topology and branch lengths. Hybrid individuals were identified using the ADMIXTURE analysis described in section 2.4 based on Q values assuming $K = 2$. When both a Q value for the *G. aculeatus* cluster and that for the *G. nipponicus* cluster at $K = 2$ were < 0.875 , that individual was classified as a hybrid. The identified hybrids were concordant with those detected by PCA (Supplementary Fig. S2, left panel).

The second tree consists of 3,717 SNPs of non-hybrid individuals from native populations of western and eastern basins of the Pacific and Europe, which were subsampled to include at most six individuals per population, two of which had the least missing genotypes and the rest of which were randomly selected. Hybrids in the East Pacific populations were identified and excluded in a manner similar to that of Japanese samples using ADMIXTURE. We selected biallelic SNPs that were genotyped in all four populations using BCFtools. The SNPs were subsampled with VCFtools to maintain a minimum distance of 1 kb. ADMIXTURE was run by varying the number of evolutionary clusters K from one through four. Identification of hybrid individuals was conducted based on Q values from the ADMIXTURE analysis assuming $K = 3$. If the Q values of an individual for any cluster did not exceed 0.875, it was classified as a hybrid. As a result, three putative hybrids from Duwamish and Big Soos were removed (Supplementary Fig. S3). The reference sequence obtained from a freshwater stickleback collected at Bear Paw Lake, Alaska [18] was added as a sample in the second dataset. Each dataset included the SNPs genotyped in all but one population and $> 70\%$ of the overall samples, keeping the minimum distance between SNPs at 1 kb. We used the general time-reversible model of nucleotide substitution with gamma-distributed rate heterogeneity and ascertainment bias correction [90] using the conditional likelihood method [91]. We

conducted bootstrap analyses with 200 replicates and searched for the best scoring trees in each of the two runs. The tree was visualized with FigTree v1.4.4 [92].

Phylogeny and divergence time among stickleback populations was estimated with the multispecies coalescent model using the Bayesian framework of SNAPP v1.5.0 [93] implemented in Beast v2.6.2 [94]. To reduce the computational time, we selected two non-hybrid individuals with the least missing genotypes from 13 representative populations covering the distribution range and distinct lineages of the stickleback. They consisted of *G. nipponicus*, marine populations of *G. aculeatus* from the western and eastern basins of the Pacific and Europe, freshwater populations from each of the three regions, including those comprising highly supported clades in the Japanese Archipelago that were revealed by the ML tree analysis. We removed SNPs with missing genotypes, and subsampled SNPs to maintain a minimal distance of 1 kb. This resulted in a dataset of 2,022 biallelic SNPs.

Root divergence was used as the calibration point. We adopted two previously published estimates as the time of divergence between *G. aculeatus* and *G. nipponicus*. The first is 680 thousand years (ka) BP following our previous study [37], estimated by a demographic analysis with an Approximate Bayesian Computation approach. The second was the 1.38 Ma BP [43] based on a Bayesian estimation of phylogeny and divergence time with concatenated RAD sequences. Although the potential overestimation of the latter due to incomplete lineage sorting is pointed out [42], we included it to account for uncertainty in the estimation of the divergence time, since it is close to another estimate of 1.22 Ma BP based on an ML-based demographic analysis [37], and within the 95% confidence interval of the former divergence time estimate (0.18–4.1 Ma).

Prior for the divergence time was specified to follow a log-normal distribution with means in real space to the respective divergence times (*i.e.*, 0.68 and 1.38 Ma), and with a standard deviation of 0.18 so that 95% intervals of the two priors do not overlap. We fixed a population parameter theta, which is proportional to the product of effective population size and mutation rate per site, to be equal across lineages with a uniform prior, following Stange et al. [95]. It should be noted that fixed and equal population sizes among all populations could flaw divergence time estimates obtained in the coalescent analysis. Monophyly of *G. aculeatus* (*i.e.*, all the populations except *G. nipponicus*) and that of two European populations were set as constraints. We used a script by Matschiner [96] to prepare input files for SNAPP. Three independent runs were performed for each calibration scheme with a chain length of $1.54\text{--}2.22 \times 10^6$ generations starting from different initial trees. Trees were sampled every 5,000 steps and checked for convergence to the stationary distribution and a sufficient effective sample size (ESS > 200) using Tracer v1.7.1 [97]. The first 10% of the trees were discarded as burn-in and the remaining trees were visualized using DensiTree v2.2.7 [98]. Maximum clade credibility consensus trees of each run after burn-in were summarized with TreeAnnotator v2.6.2 [99] and visualized with FigTree [92].

Abbreviations

ML: Maximum likelihood; *Eda*: *Ectodysplasin*; ddRAD: Double digest Restriction-site associated DNA; SNP: single nucleotide polymorphism; PCA: principal component analysis; ka: thousand years; Ma: million years; BP: before present; HDI: highest posterior density intervals.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The sequence data generated during the current study are available in DDBJ (DRA010673).

Competing interests

The authors declare that they have no competing interests.

Funding

This research was supported by JSPS Kakenhi to JK (17KT0028 and 19H01003) and MEXT Kakenhi to AT (16H06279). The funding bodies had no role in study design, data analysis, interpretation of data, decision to publish, or preparation of the manuscript.

Authors' contributions

RK and JK conceived the project, interpreted results, and drafted the manuscript. RK analyzed data. RK, SM, TK, TKH, AI, MK, and JK contributed materials. AJN and AT performed sequencing. All authors read and approved the final manuscript.

Acknowledgements

We thank Yi-Ta Shao, Yoshiyasu Machida, Yuya Kogame, Kota Kamiyama, Natsuki Suzuki, Io Miyashita, Katie Peichel and Dolph Schluter for providing samples and/or help with sampling, Haruki Hinata for providing information about a stickleback habitat, Satoko Kondo and Lina Kawaguchi for technical assistance, and Yo Yamasaki for discussion.

Authors' information

¹Ecological Genetics Laboratory, National Institute of Genetics, Mishima, Shizuoka, 411-8540 Japan.

²Present address: Tropical Biosphere Research Center, University of the Ryukyus, Nishihara, Nakagami-

gun, Okinawa 903-0213 Japan. ³Biological Laboratory, Gifu Kyoritsu University, Ogaki, Gifu, 503-8550 Japan. ⁴Department of Marine Bioscience, Fukui Prefectural University, Obama, Fukui, 917-0003 Japan. ⁵Department of Genetics, The Graduate University for Advanced Studies, Mishima, Shizuoka, 411-8540 Japan. ⁶Faculty of Agriculture, Ryukoku University, Otsu, Shiga, 520-2194 Japan. ⁷Field Science Education and Research Center, Kyoto University, Kyoto, 606-8502 Japan. ⁸Comparative Genomics Laboratory, National Institute of Genetics, Mishima, Shizuoka, 411-8540 Japan.

References

1. Schluter D: The Ecology of Adaptive Radiation. Oxford: Oxford University Press; 2000.
2. Martin A, Orgogozo V. The loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. *Evolution* 2013;67(5):1235–1250.
3. Conte GL, Arnegard ME, Peichel CL, Schluter D. The probability of genetic parallelism and convergence in natural populations. *Proc R Soc B* 2012;279(1749):5039–5047.
4. Kopp A. Metamodels and phylogenetic replication: a systematic approach to the evolution of developmental pathways. *Evolution* 2009;63(11):2771–2789.
5. Stern DL. The genetic causes of convergent evolution. *Nat Rev Genet* 2013;14(11):751–764.
6. Arendt J, Reznick D. Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends Ecol Evol* 2008;23(1):26–32.
7. Seehausen O, Wagner CE. Speciation in Freshwater Fishes. *Annu Rev Ecol Evol Syst* 2014;45(1):621–651.
8. Lee CE, Bell MA. Causes and consequences of recent freshwater invasions by saltwater animals. *Trends Ecol Evol* 1999;14(7):284–288.
9. Schluter D, Conte GL. Genetics and ecological speciation. *Proc Natl Acad Sci U S A* 2009;106 Suppl 1:9955–9962.
10. Peichel CL, Marques DA. The genetic and molecular architecture of phenotypic diversity in sticklebacks. *Phil Trans R Soc B* 2017;372(1713):20150486.
11. Kitano J, Ishikawa A, Kume M, Mori S. Physiological and genetic basis for variation in migratory behavior in the three-spined stickleback, *Gasterosteus aculeatus*. *Ichthyol Res* 2012;59(4):293–303.
12. Wootton RJ: The Biology of the Sticklebacks. London: Academic Press; 1976.
13. Bell MA, Foster SA: The Evolutionary Biology of the Threespine Stickleback. Oxford: Oxford University Press; 1994.
14. Colosimo PF, Hosemann KE, Balabhadra S, Villarreal G, Jr., Dickson M, Grimwood J, Schmutz J, Myers RM, Schluter D, Kingsley DM. Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. *Science* 2005;307(5717):1928–1933.
15. Ishikawa A, Kabeya N, Ikeya K, Kakioka R, Cech JN, Osada N, Leal MC, Inoue J, Kume M, Toyoda A, Tezuka A, Nagano AJ, Yamasaki YY, Suzuki Y, Kokita T, Takahashi H, Lucek K, Marques D, Takehana

- Y, Naruse K, Mori S, Monroig O, Ladd N, Schubert CJ, Matthews B, Peichel CL, Seehausen O, Yoshizaki G, Kitano J. A key metabolic gene for recurrent freshwater colonization and radiation in fishes. *Science* 2019;364(6443):886–889.
16. Colosimo PF, Peichel CL, Nereng K, Blackman BK, Shapiro MD, Schluter D, Kingsley DM. The genetic architecture of parallel armor plate reduction in threespine sticklebacks. *PLoS Biol* 2004;2(5):E109.
 17. Chan YF, Marks ME, Jones FC, Villarreal G, Jr., Shapiro MD, Brady SD, Southwick AM, Absher DM, Grimwood J, Schmutz J, Myers RM, Petrov D, Jonsson B, Schluter D, Bell MA, Kingsley DM. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. *Science* 2010;327(5963):302–305.
 18. Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, Swofford R, Pirun M, Zody MC, White S, Birney E, Searle S, Schmutz J, Grimwood J, Dickson MC, Myers RM, Miller CT, Summers BR, Knecht AK, Brady SD, Zhang H, Pollen AA, Howes T, Amemiya C, Broad Institute Genome Sequencing Platform & Whole Genome Assembly Team, Baldwin J, Bloom T, Jaffe DB, Nicol R, Wilkinson J, Lander ES, Di Palma F, Lindblad-Toh K, Kingsley DM. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* 2012;484(7392):55–61.
 19. Miller SE, Roesti M, Schluter D. A single interacting species leads to widespread parallel evolution of the stickleback genome. *Curr Biol* 2019;29(3):530–537.
 20. Kitano J, Lema SC, Luckenbach JA, Mori S, Kawagishi Y, Kusakabe M, Swanson P, Peichel CL. Adaptive divergence in the thyroid hormone signaling pathway in the stickleback radiation. *Curr Biol* 2010;20(23):2124–2130.
 21. Jones FC, Chan YF, Schmutz J, Grimwood J, Brady SD, Southwick AM, Absher DM, Myers RM, Reimchen TE, Deagle BE, Schluter D, Kingsley DM. A genome-wide SNP genotyping array reveals patterns of global and repeated species-pair divergence in sticklebacks. *Curr Biol* 2012;22(1):83–90.
 22. Miller CT, Beleza S, Pollen AA, Schluter D, Kittles RA, Shriver MD, Kingsley DM. *cis*-Regulatory changes in *Kit* ligand expression and parallel evolution of pigmentation in sticklebacks and humans. *Cell* 2007;131(6):1179–1189.
 23. Fang B, Kemppainen P, Momigliano P, Feng X, Merilä J. On the causes of geographically heterogeneous parallel evolution in sticklebacks. *Nat Ecol Evol.* 2020. doi:10.1038/s41559-020-1222-6.
 24. Yamasaki YY, Mori S, Kokita T, Kitano J. Armour plate diversity in Japanese freshwater threespine stickleback (*Gasterosteus aculeatus*). *Evol Ecol Res* 2019;20:51–67.
 25. Leinonen T, McCairns RJ, Herczeg G, Merila J. Multiple evolutionary pathways to decreased lateral plate coverage in freshwater threespine sticklebacks. *Evolution* 2012;66(12):3866–3875.
 26. O’Brown NM, Summers BR, Jones FC, Brady SD, Kingsley DM. A recurrent regulatory change underlying altered expression and Wnt response of the stickleback armor plates gene *EDA*. *eLife* 2015;4:e05290.
 27. Watanabe K, Mori S, Nishida M. Genetic relationships and origin of two geographic groups of the freshwater threespine stickleback, ‘hariyo’. *Zool Sci* 2003;20(2):265–274.

28. Kitano J, Mori S. Toward conservation of genetic and phenotypic diversity in Japanese sticklebacks. *Genes Genet Syst* 2016;91(2):77–84.
29. Higuchi M, Goto A, Yamazaki F. Genetic structure of threespine stickleback, *Gasterosteus aculeatus*, in Lake Harutori, Japan, with reference to coexisting anadromous and freshwater forms. *Ichthyol Res* 1996;43(4):349–358.
30. Kitano J, Mori S, Peichel CL. Phenotypic divergence and reproductive isolation between sympatric forms of Japanese threespine sticklebacks. *Biol J Linn Soc* 2007;91(4):671–685.
31. Adachi T, Ishikawa A, Mori S, Makino W, Kume M, Kawata M, Kitano J. Shifts in morphology and diet of non-native sticklebacks introduced into Japanese crater lakes. *Ecol Evol* 2012;2(6):1083–1098.
32. Yoshida K, Miyagi R, Mori S, Takahashi A, Makino T, Toyoda A, Fujiyama A, Kitano J. Whole-genome sequencing reveals small genomic regions of introgression in an introduced crater lake population of threespine stickleback. *Ecol Evol* 2016;6(7):2190–2204.
33. Higuchi M, Sakai H, Goto A. A new threespine stickleback, *Gasterosteus nipponicus* sp. nov. (Teleostei: Gasterosteidae), from the Japan Sea region. *Ichthyol Res* 2014;61(4):341–351.
34. Higuchi M, Goto A. Genetic evidence supporting the existence of two distinct species in the genus *Gasterosteus* around Japan. *Environ Biol Fish* 1996;47(1):1–16.
35. Cassidy LM, Ravinet M, Mori S, Kitano J. Are Japanese freshwater populations of threespine stickleback derived from the Pacific Ocean lineage? *Evol Ecol Res* 2013;15:295–311.
36. Kitano J, Ross JA, Mori S, Kume M, Jones FC, Chan YF, Absher DM, Grimwood J, Schmutz J, Myers RM, Kingsley DM, Peichel CL. A role for a neo-sex chromosome in stickleback speciation. *Nature* 2009;461(7267):1079–1083.
37. Ravinet M, Yoshida K, Shigenobu S, Toyoda A, Fujiyama A, Kitano J. The genomic landscape at a late stage of stickleback speciation: High genomic divergence interspersed by small localized regions of introgression. *PLoS Genet* 2018;14(5):e1007358.
38. Ravinet M, Kume M, Ishikawa A, Kitano J. Patterns of genomic divergence and introgression between Japanese stickleback species with overlapping breeding habitats. *J Evol Biol*. 2020. doi:10.1111/jeb.13664.
39. Yamada M, Higuchi M, Goto A. Extensive introgression of mitochondrial DNA found between two genetically divergent forms of threespine stickleback, *Gasterosteus aculeatus*, around Japan. *Environ Biol Fish* 2001;61(3):269–284.
40. Yamada M, Higuchi M, Goto A. Long-term occurrence of hybrids between Japan Sea and Pacific Ocean forms of threespine stickleback, *Gasterosteus aculeatus*, in Hokkaido Island, Japan. *Environ Biol Fish* 2007;80(4):435–443.
41. Yoshida K, Ravinet M, Makino T, Toyoda A, Kokita T, Mori S, Kitano J. Accumulation of deleterious mutations in landlocked threespine stickleback populations. *Genome Biol Evol* 2020;12(4):479–492.
42. Fang B, Merilä J, Matschiner M, Momigliano P. Estimating uncertainty in divergence times among three-spined stickleback clades using the multispecies coalescent. *Mol Phylogenet Evol* 2020;142:106646.

43. Fang B, Merilä J, Ribeiro F, Alexandre CM, Momigliano P. Worldwide phylogeny of three-spined sticklebacks. *Mol Phylogenet Evol* 2018;127:613–625.
44. Hu A, Meehl GA, Otto-Bliesner BL, Waelbroeck C, Han W, Loutre M-F, Lambeck K, Mitrovica JX, Rosenbloom N. Influence of Bering Strait flow and North Atlantic circulation on glacial sea-level changes. *Nat Geosci* 2010;3(2):118–121.
45. Ikeda K. Togeuo no bumpu to sono hen'i [The distribution and morphological variation of sticklebacks]. *Zool Mag* 1933;45:141–173 (in Japanese).
46. Yamamoto M. Equatorward shift of the subarctic boundary in the northwestern Pacific during the last deglaciation. *Geophys Res Lett* 2005;32(5):L05609.
47. Itaki T, Komatsu N, Motoyama I. Orbital- and millennial-scale changes of radiolarian assemblages during the last 220 kyrs in the Japan Sea. *Palaeogeogr Palaeoclimatol Palaeoecol* 2007;247(1-2):115–130.
48. Takei T, Minoura K, Tsukawaki S, Nakamura T. Intrusion of a branch of the Oyashio Current into the Japan Sea during the Holocene. *Paleoceanography* 2002;17(3):11-11–11-10.
49. Bell MA. Reduction and loss of the pelvic girdle in *Gasterosteus* (Pisces): a case of parallel evolution. *Nat Hist Mus Los Angel Cty Contrib Sci* 1974;257:1–36.
50. Bell MA. A late Miocene marine threespine stickleback, *Gasterosteus aculeatus aculeatus*, and its zoogeographic and evolutionary significance. *Copeia* 1977;1977(2):277–282.
51. Mäkinen HS, Merilä J. Mitochondrial DNA phylogeography of the three-spined stickleback (*Gasterosteus aculeatus*) in Europe—evidence for multiple glacial refugia. *Mol Phylogenet Evol* 2008;46(1):167–182.
52. Mäkinen HS, Cano JM, Merilä J. Genetic relationships among marine and freshwater populations of the European three-spined stickleback (*Gasterosteus aculeatus*) revealed by microsatellites. *Mol Ecol* 2006;15(6):1519–1534.
53. DeFaveri J, Zanella LN, Zanella D, Mrakovcic M, Merila J. Phylogeography of isolated freshwater three-spined stickleback *Gasterosteus aculeatus* populations in the Adriatic Sea basin. *J Fish Biol* 2012;80(1):61–85.
54. Lucek K, Seehausen O. Distinctive insular forms of threespine stickleback (*Gasterosteus aculeatus*) from western Mediterranean islands. *Conserv Genet* 2015;16(6):1319–1333.
55. Sanz N, Araguas RM, Vidal O, Viñas J. Glacial refuges for three-spined stickleback in the Iberian Peninsula: mitochondrial DNA phylogeography. *Freshw Biol* 2015;60(9):1794–1809.
56. Ohe F, Koike H. Threespined stickleback from the Miocene Bessho Formation, at Nakatani, Toyoshina, Tazawa in Azumino City, Nagano Prefecture. *Bull Nagano City Mus Div Nat Sci* 2015;16:16–29 (in Japanese).
57. Bell MA: Paleobiology and evolution of threespine stickleback. In: *The Evolutionary Biology of the Threespine Stickleback*. Edited by Bell MA, Foster SA. Oxford: Oxford University Press; 1994: 439–471.

58. Bell MA, Stewart JD, Park PJ. The world's oldest fossil threespine stickleback fish. *Copeia* 2009;2009(2):256–265.
59. Nelson TC, Cresko WA. Ancient genomic variation underlies repeated ecological adaptation in young stickleback populations. *Evol Lett* 2018;2(1):9–21.
60. Rhymer JM, Simberloff D. Extinction by hybridization and introgression. *Annu Rev Ecol Syst* 1996;27(1):83–109.
61. Hosoki T, Mori S, Nishida S, Sumi T, Kitano J. Diversity of gill raker number and diets among stickleback populations in novel habitats created by the 2011 Tōhoku earthquake and tsunami. *Evol Ecol Res* 2019;20:213–230.
62. Kume M, Mori S, Kitano J, Sumi T, Nishida S. Impact of the huge 2011 Tohoku-oki tsunami on the phenotypes and genotypes of Japanese coastal threespine stickleback populations. *Sci Rep* 2018;8(1):1684.
63. Bell MA, Aguirre WE. Contemporary evolution, allelic recycling, and adaptive radiation of the threespine stickleback. *Evol Ecol Res* 2013;15:377–411.
64. Ravinet M, Faria R, Butlin RK, Galindo J, Bierne N, Rafajlovic M, Noor MAF, Mehlig B, Westram AM. Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *J Evol Biol* 2017;30(8):1450–1477.
65. Nosil P: *Ecological Speciation*. Oxford: Oxford University Press; 2012.
66. Ravinet M, Takeuchi N, Kume M, Mori S, Kitano J. Comparative analysis of Japanese three-spined stickleback clades reveals the Pacific Ocean lineage has adapted to freshwater environments while the Japan Sea has not. *PLoS ONE* 2014;9(12):e112404.
67. Kitano J, Bolnick DI, Beauchamp DA, Mazur MM, Mori S, Nakano T, Peichel CL. Reverse evolution of armor plates in the threespine stickleback. *Curr Biol* 2008;18(10):769–774.
68. Sakaguchi S, Sugino T, Tsumura Y, Ito M, Crisp MD, Bowman DMJS, Nagano AJ, Honjo MN, Yasugi M, Kudoh H, Matsuki Y, Suyama Y, Isagi Y. High-throughput linkage mapping of Australian white cypress pine (*Callitris glaucophylla*) and map transferability to related species. *Tree Genet Genomes* 2015;11(6):121.
69. Yoshida K, Makino T, Yamaguchi K, Shigenobu S, Hasebe M, Kawata M, Kume M, Mori S, Peichel CL, Toyoda A, Fujiyama A, Kitano J. Sex chromosome turnover contributes to genomic divergence between incipient stickleback species. *PLoS Genet* 2014;10(3):e1004223.
70. Feulner PG, Chain FJ, Panchal M, Eizaguirre C, Kalbe M, Lenz TL, Mundry M, Samonte IE, Stoll M, Milinski M, Reusch TB, Bornberg-Bauer E. Genome-wide patterns of standing genetic variation in a marine population of three-spined sticklebacks. *Mol Ecol* 2013;22(3):635–649.
71. Feulner PG, Chain FJ, Panchal M, Huang Y, Eizaguirre C, Kalbe M, Lenz TL, Samonte IE, Stoll M, Bornberg-Bauer E, Reusch TB, Milinski M. Genomics of divergence along a continuum of parapatric population differentiation. *PLoS Genet* 2015;11(2):e1004966.
72. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;30(15):2114–2120.

73. Sedlazeck FJ, Rescheneder P, von Haeseler A. NextGenMap: fast and accurate read mapping in highly polymorphic genomes. *Bioinformatics* 2013;29(21):2790–2791.
74. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. arXiv 2012:1207.3907 [q-bio.GN].
75. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 2011;27(21):2987–2993.
76. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010;26(6):841–842.
77. Tan A, Abecasis GR, Kang HM. Unified representation of genetic variants. *Bioinformatics* 2015;31(13):2202–2204.
78. Zhang J, Kobert K, Flouri T, Stamatakis A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* 2014;30(5):614–620.
79. Edwards JA, Edwards RA. Fastq-pair: efficient synchronization of paired-end fastq files. bioRxiv. 2019. doi:10.1101/552885.
80. Shen W, Le S, Li Y, Hu F. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS ONE* 2016;11(10):e0163962.
81. Broad Institute: Picard Tools ver. 2.21.8. <http://broadinstitute.github.io/picard/> (2020) Accessed 10 Feb 2020.
82. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 2009;19(9):1655–1664.
83. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, Genomes Project Analysis G. The variant call format and VCFtools. *Bioinformatics* 2011;27(15):2156–2158.
84. 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;4:7.
85. Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I. CLUMPAK: a program for identifying clustering modes and packaging population structure inferences across *K*. *Mol Ecol Resour* 2015;15(5):1179–1191.
86. Jombart T. *adeigenet*: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 2008;24(11):1403–1405.
87. Jombart T, Ahmed I. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* 2011;27(21):3070–3071.
88. R Core Team: R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing; 2020.
89. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAXML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* 2019;35(21):4453–4455.

90. Leaché AD, Banbury BL, Felsenstein J, de Oca AN, Stamatakis A. Short tree, long tree, right tree, wrong tree: new acquisition bias corrections for inferring SNP phylogenies. *Syst Biol* 2015;64(6):1032–1047.
91. Lewis PO. A likelihood approach to estimating phylogeny from discrete morphological character data. *Syst Biol* 2001;50(6):913–925.
92. Rambaut A: FigTree ver. 1.4.4. <http://tree.bio.ed.ac.uk/software/figtree/> (2018) Accessed 9 Mar 2020.
93. Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA, RoyChoudhury A. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Mol Biol Evol* 2012;29(8):1917–1932.
94. Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, Heled J, Jones G, Kühnert D, De Maio N, Matschiner M, Mendes FK, Müller NF, Ogilvie HA, du Plessis L, Poppinga A, Rambaut A, Rasmussen D, Siveroni I, Suchard MA, Wu CH, Xie D, Zhang C, Stadler T, Drummond AJ. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput Biol* 2019;15(4):e1006650.
95. Stange M, Sanchez-Villagra MR, Salzburger W, Matschiner M. Bayesian divergence-time estimation with genome-wide single-nucleotide polymorphism data of sea catfishes (Ariidae) supports Miocene closure of the Panamanian Isthmus. *Syst Biol* 2018;67(4):681–699.
96. Matschiner M: snapp_prep.rb. https://github.com/mmatschiner/snapp_prep (2018). Accessed 10 Apr 2020.
97. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst Biol* 2018;67(5):901–904.
98. Bouckaert R, Heled J. DensiTree 2: seeing trees through the forest. *bioRxiv*. 2014. doi:10.1101/012401.
99. Rambaut A, Drummond AJ: TreeAnnotator ver. 2.6.2. <https://www.beast2.org/treeannotator/> (2020) Accessed 16 Mar 2020.
100. Higuchi M: Nippon-rettō shūhen no itoyo-zoku gyorui no identeki tayōsei to bunka [Genetic diversity and divergence of *Gasterosteus* fishes around the Japanese Archipelago]. In: Togeuo no Shizenshi [Natural History of Sticklebacks]. Edited by Goto A, Mori S. Sapporo: Hokkaido University Press; 2003: 49–60 (in Japanese).
101. Yoshigou H. The inland-water fishes collected from the Dogo Island, Oki Islands, in the Sea of Japan. *Misc Rep Hiwa Mus Nat Hist* 2001;40:1–15 (in Japanese).
102. Batchelor CL, Margold M, Krapp M, Murton DK, Dalton AS, Gibbard PL, Stokes CR, Murton JB, Manica A. The configuration of Northern Hemisphere ice sheets through the Quaternary. *Nat Commun* 2019;10(1):3713.

Figures

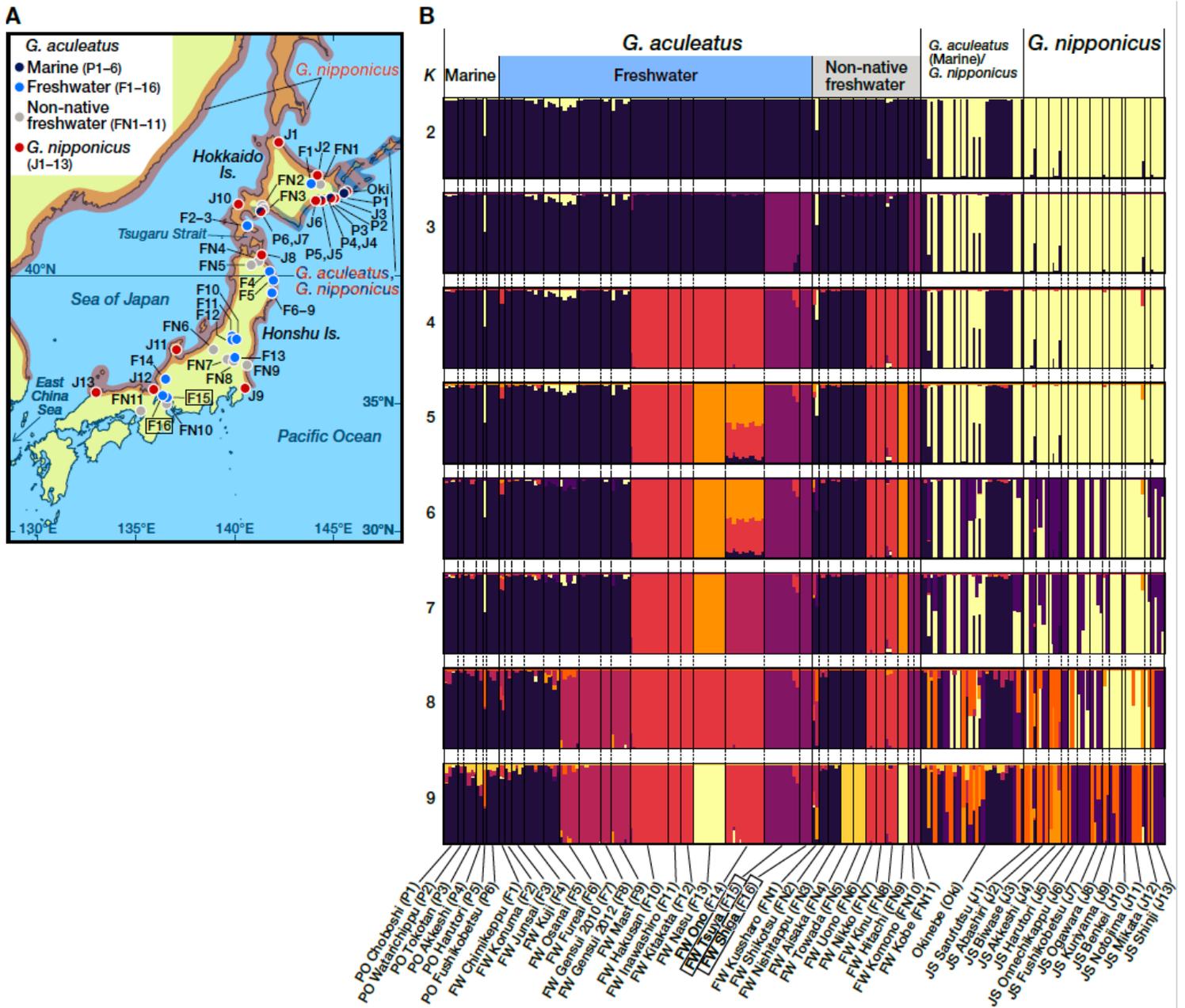


Figure 1

(A) Sampling sites in the Japanese Archipelago and distribution ranges of marine *Gasterosteus aculeatus* and *G. nipponicus*. See Fig. 1B for population codes. Distribution range of *G. nipponicus* is shaded in red. The distribution range of marine *G. aculeatus*, which overlaps with that of *G. nipponicus* in the Japanese Archipelago, is hatched in dark blue. Population codes enclosed in a box (F15 and F16) indicate the native habitats of the hariyo sticklebacks. The distribution ranges followed those of Higuchi [100], Higuchi et al. [33], Kitano and Mori [28] and Yoshigou [101]. The map was created with rnaturalearth ver. 0.1.0 (<https://github.com/ropensci/rnaturalearth>) and sf ver. 0.9-0 (<https://r-spatial.github.io/sf/>). (B) Bar plots showing the results of the population structure analyses of Japanese samples based on 2,735 SNPs with ADMIXTURE (K = 2-9). Individuals are represented as vertical bars

proportional to the genotypes belonging to each of the genetic clusters. Population codes in brackets follow population names. Population names enclosed in a box (FW Tsuya [F15] and FW Shiga [F16]) indicate the hariyo sticklebacks.

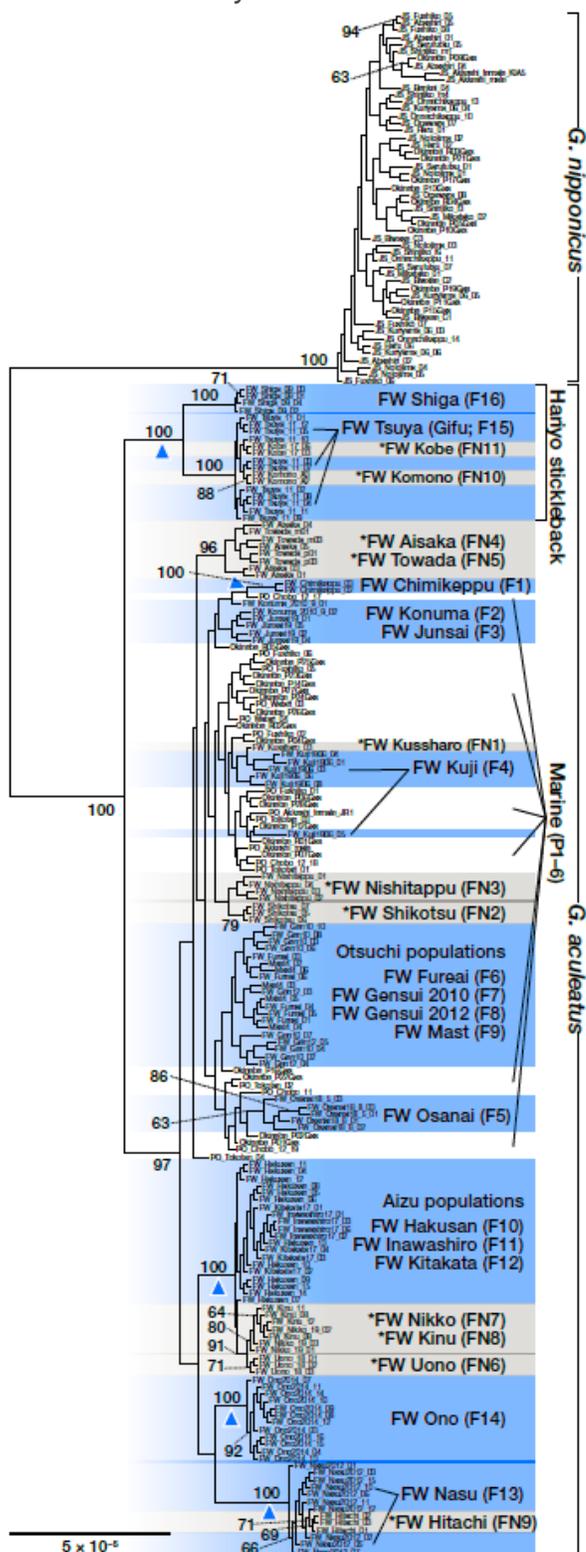


Figure 2

Maximum likelihood phylogenetic tree of non-hybrid individuals from Japan based on 1,919 concatenated SNPs. Bootstrap values (> 60%) are shown. Individuals from native freshwater populations

of *Gasterosteus aculeatus* are highlighted in blue and those from non-native populations of *G. aculeatus* are highlighted in grey. Non-native population names are marked with asterisks. Blue arrowheads indicate monophyletic freshwater populations and a group of freshwater populations in vicinity.

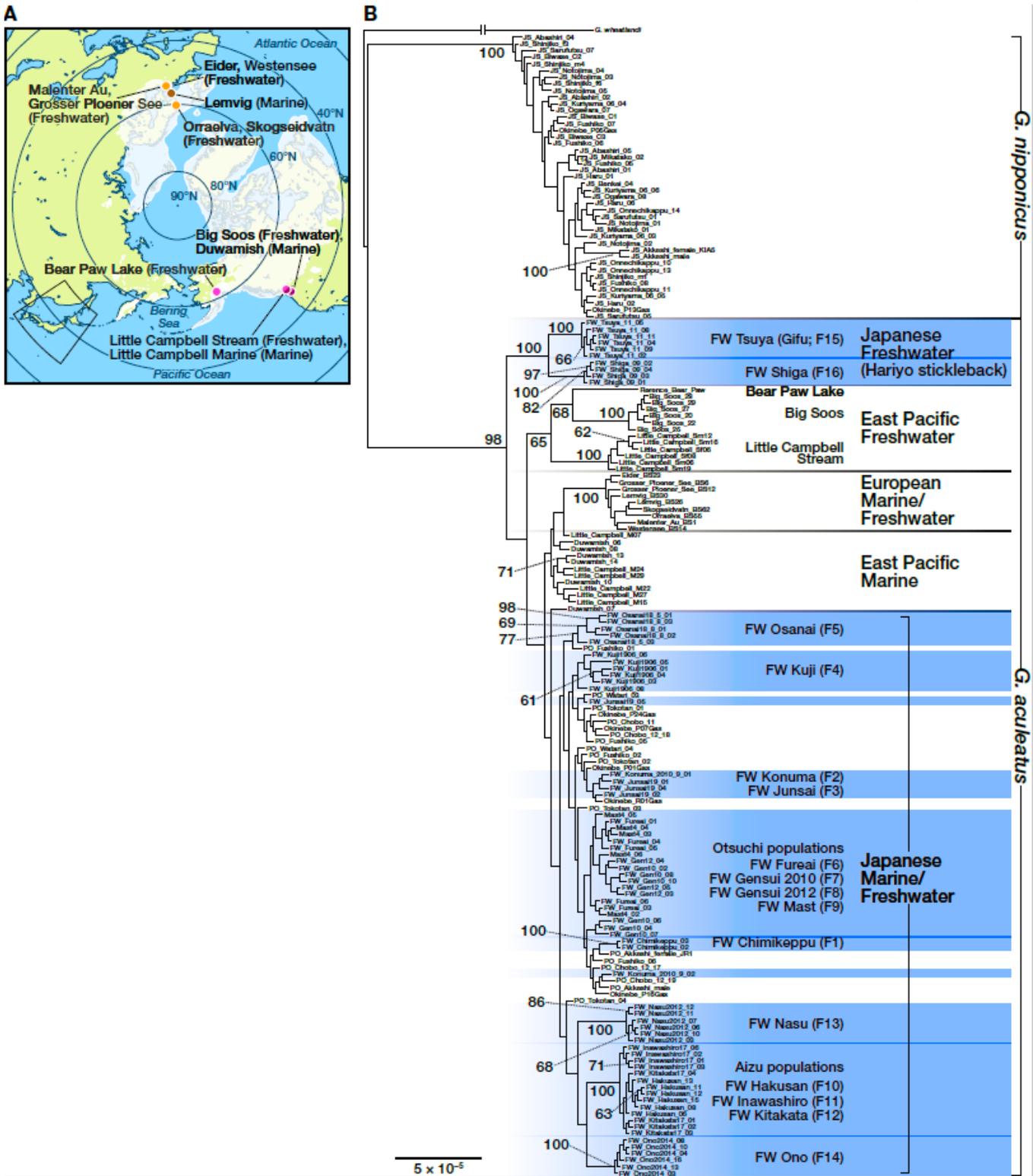


Figure 3

(A) Sampling sites of *Gasterosteus aculeatus* in the eastern Pacific basin and northern Europe. The extent of Fig. 1A is bounded with black lines. The configuration of ice sheets at the last glacial maximum [102]

is shown with white shading. The map was created with *rnatuarearth* ver. 0.1.0 (<https://github.com/ropensci/rnatuarearth>) and *sf* ver. 0.9-0 (<https://r-spatial.github.io/sf/>). (B) Maximum likelihood phylogenetic tree of native non-hybrid individuals from western and eastern basins of the Pacific and northern Europe based on 3,717 concatenated SNPs. Individuals from Japanese freshwater populations of *G. aculeatus* are highlighted in blue. Bootstrap values (> 60%) are shown.

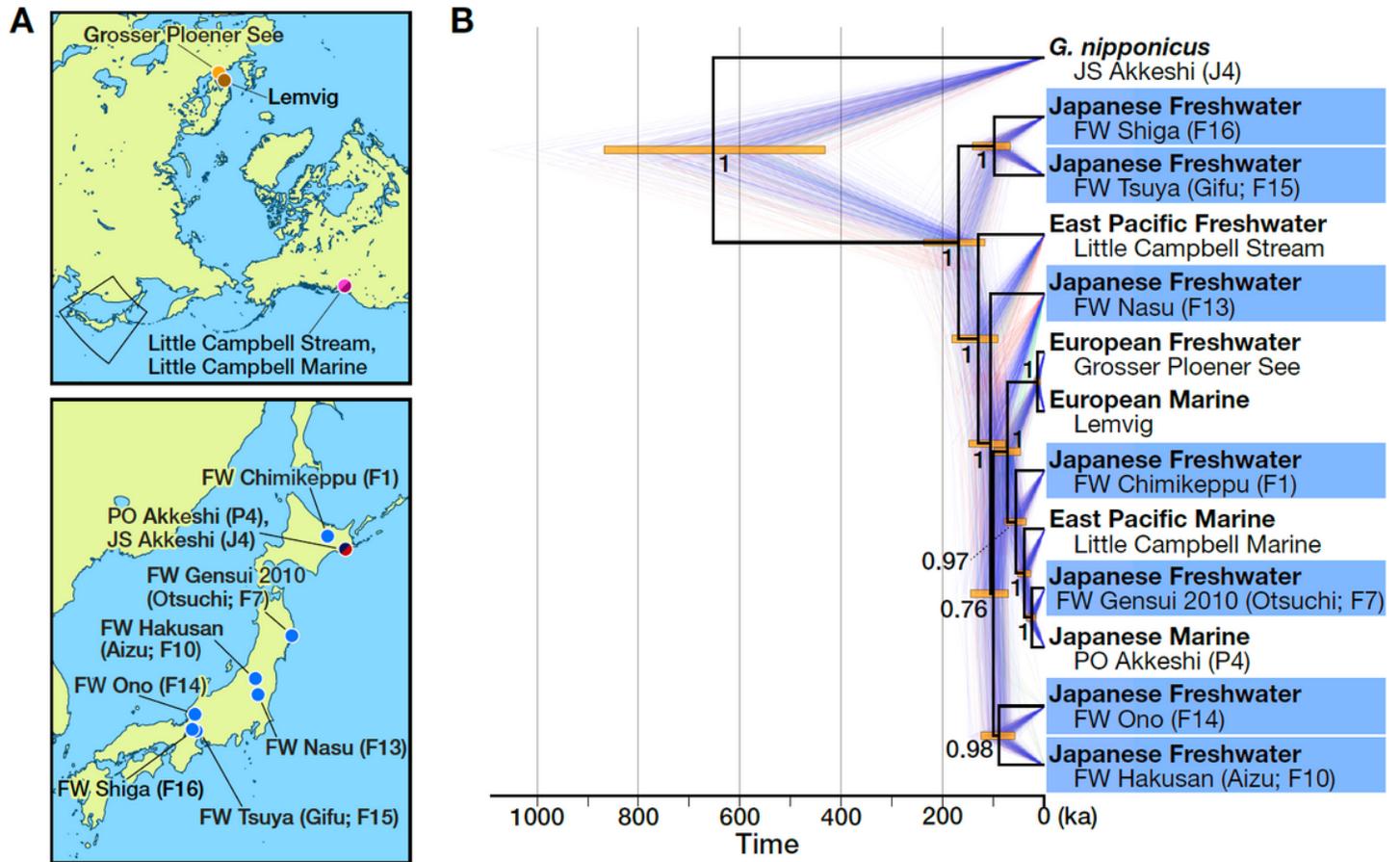


Figure 4

(A) Sampling sites of *Gasterosteus aculeatus* used for a time-calibrated species tree analysis. (B) A time-calibrated species tree of representative populations of *Gasterosteus aculeatus* and *G. nipponicus* inferred with SNAPP based on 2,022 SNPs with the root calibration at 680 ka. The trees recorded in a run are overlaid by the maximum clade credibility tree. Posterior probabilities of each node are shown. Each bar plot indicates the 95% highest posterior density interval of the node height. Individuals from Japanese freshwater populations of *G. aculeatus* are highlighted in blue. The results of other runs are provided in Supplementary Fig. S5.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [FigS1.v4a.pdf](#)

- [FigS2.v3.pdf](#)
- [FigS3.v1a.pdf](#)
- [FigS4.v1a.pdf](#)
- [FigS5.v2.pdf](#)
- [TableS1.v1.0.1.xlsx](#)
- [TableS2.v1.0.0.xlsx](#)