

Temporal Dynamics of Teleost Populations During the Pleistocene: A Report from Publicly Available Genome Data

Jia Li

Guangdong Provincial Key Lab of Molecular Breeding in Marine Economic Animals, BGI Academy of Marine Sciences, BGI Marine, BGI, Shenzhen <https://orcid.org/0000-0003-1500-869X>

Chao Bian

BGI Academy of Marine Sciences

Yunhai Yi

BGI Marine

Hui Yu

BGI Marine

Xinxin You

BGI Marine

Qiong Shi (✉ shiqiong@genomics.cn)

BGI Marine

Research article

Keywords: global climate change, temporal dynamics, effective population size, high-quality genome assembly, teleost

Posted Date: January 13th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-142802/v1>

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

[Read Full License](#)

Abstract

Global climate oscillation, as a selection dynamic, is an ecologically important element resulting in global biodiversity. During the glacial geological periods, most organisms suffered detrimental selection pressures (such as food shortage and habitat loss) and went through population declines. However, during the mild interglacial periods, many species re-flourished. These temporal dynamics of effective population sizes (N_e) provide essential information in understanding and predicting evolutionary outcomes during historical and ongoing global climate changes. Using high-quality genome assemblies and corresponding sequencing data, we applied the Pairwise Sequentially Markovian Coalescent (PSMC) method to quantify N_e changes of fifteen representative teleost species from approximately 10 million years ago (mya) to 10 thousand years ago (kya). The results revealed multiple rounds of population contraction and expansion in most of the examined teleost species during the Neogene and the Quaternary periods. It was found that 67% (10/15) of the examined teleosts had experienced a drastic decline in N_e before the last glacial period (LGP, 110–12 kya), slightly earlier than the reported pattern of N_e changes in 38 avian species. In comparison with the peaks, almost all of the examined teleosts maintained long-term lower N_e values during the last few million years. This is consistent with increasingly dramatic glaciation observed during this period. In summary, these findings provide a more comprehensive understanding of the historical N_e changes in teleosts. Results presented here could lead to the development of appropriate strategies to protect species in light of ongoing global climate changes.

Background

Global climate oscillation is an ecologically important element leading to the global biodiversity. Global climate fluctuations can exert a wide range of selection pressures influencing the distribution and proliferation of organisms [1, 2]. In earth's history there have been five mass extinctions identified [3], several of which were related to climate change. For example, possible causes for the Ordovician-Silurian extinction events (443.8 mya) include glaciation and volcanism [4]. Moreover, detrimental environmental changes associated with glaciation might have also contributed to this extinction. First, the cooling global climate may have elicited the species that were living to an intense greenhouse extinction [5, 6]. Secondly, a reduction in the sea level drained vast epicontinental seaways, most likely destroying the habitat of many endemic communities [7–9]. Volcanism released harmful gases and ashes into the atmosphere, thereby causing the greenhouse effect, atmosphere darkening, reduction in photosynthesis and atmospheric oxygen, as well as destruction of critical food webs [10]. Nearly all major taxonomic groups on earth were affected during the Ordovician-Silurian extinction event, eliminating 49–60% of marine genera and nearly 85% of marine species [11, 12].

Almost all species are influenced by climate changes and many have developed various responses, such as transference of habitat and alterations in phenotype and genotype [13, 14]. Climatic fluctuations in earth's past dramatically impacted the demography and distribution of species. During the glaciation

periods, native environmental conditions were destroyed and replaced with ice and permafrost. Consequently, the cooling global climate produced either species extinction or forced species to migrate into new habitats. During the interglacial periods, however, ice and permafrost melted, which provided more suitable habitats for species to recolonize [15, 16].

How to discern changes in effective population size (N_e) through time has long plagued researchers. Several methods have been developed to reconstruct demographic history, such as the Skyline-Plot Methods [17]. However, most of these methods only employ a restricted number of marker loci [18, 19], or due to heavy computation requirements they cannot be used for large-scale datasets [20, 21]. With the rapid development of high-throughput sequencing technologies, scientists have generated massive genomic data and marker loci for many species. This advantageous development, and several other new analytical approaches, have been enabled the use of large numbers of loci to infer the changes of N_e . For instance, the pairwise sequential Markovian Coalescent Method (PSMC) establishes a model within a single diploid genome inferring the N_e as well as the projected time of the most recent common ancestor (TMRCA; [22]. Compared with other methods, PSMC works on resequencing data too. PSMC method produces less potential ascertainment bias and is easier to implement. Unlike other methods with simulations of predefined models, such as parametric structure of times, divergences, and size changes, the PSMC method inferring the N_e for each species does not require a predefined model.

Many genome projects have sought to investigate the relationship between sea levels and the population changes of amphibious fishes [23]. In addition, it has been a focus of research to detect bird species population decline or expansion during the Quaternary period [24, 25]. These studies have attempted to analyze the interaction of population size with geological events [26–28]. Nadachowska-Brzyska et al. [24] explored historic population changes in many avian species. For example, it was found that the population size of endangered birds experienced long-term decline. During the Quaternary period, the fluctuation in population size presented consistent correlative characteristics for many bird species. Both extinction and speciation, in fact, coexisted in this period.

However, cold-blooded fish are more vulnerable than warm-blooded birds to climate changes [29]. Fish metabolic rates are closely linked to body temperature, which are dependent on external temperature conditions [30]. Changes in fish population size are therefore more strongly correlated with climate fluctuations than that observed in birds.

In this paper, the PSMC method was used to infer population size changes in fifteen representative fish species, starting from 10 kya to 10 mya. It was found that multiple rounds of N_e increase and decline occurred in most of the examined fish species. This was predicted to be directly generated by the drastic oscillations in global climate during the Neogene and the Quaternary periods. Ten of the fifteen studied fish species experienced drastic reductions in N_e since the beginning of the last glacial period (LGP, 110–12 kya) [31], which was slightly earlier than the pattern observed in birds [24].

Results

Effective population sizes of representative fish during the Quaternary period

In the present study, genome-wide heterozygosity varied among the fifteen representative fish species with high-quality genome assemblies (Fig. 2a). Interestingly, the mean heterozygous sites per nucleotide in the genome sequence of zebrafish was 1.1×10^{-2} . This finding is much higher than in other fish. The zebrafish variant was directly obtained from the Ensembl database, which integrated comprehensive results based on data from different individuals. Hence, this guaranteed the higher sequencing coverage and efficient variant data in zebrafish, possibly contributing to the higher heterozygosity.

The heterozygous rate per nucleotide in Mexican tetra was about 2×10^{-4} , representing the lowest among these examined fish species. The mean heterozygous rates observed among the three varieties of arowana were 4.5×10^{-3} , 2.7×10^{-3} and 2.0×10^{-3} (golden, green and red arowana), respectively. The rate was 2.0×10^{-3} for Japanese flounder, which was greatly reduced compared with the heterozygosity in the half-smooth tongue sole (5.3×10^{-3}).

The historical N_e values of the fifteen representative fish species were inferred using the PSMC approach over a range from 10 kya up to 20 mya (Fig. 2b). The results allow us to track population changes during the Neogene period (2.58–23.03 mya) and the Quaternary period (following the Neogene Period spanning from 2.58 mya to present). Over these periods, the maximal N_e values among the fifteen examined fish species had experienced remarkable variations, ranging from 2,000,000 to 6,000,000 in blue tilapia, Nile tilapia and half-smooth tongue sole; the minimal N_e ranged between 1,000–6,000 (large yellow croaker and Mexican tetra), 50,000–10,000 (Atlantic herring, channel catfish, and spotted green pufferfish), and 100,000–200,000 (half-smooth tongue sole, green arowana, and golden arowana). Thus, the N_e values ranged in at least four orders of magnitude among various teleost species during the examined periods.

Multiple rounds of N_e increase and decline were observed in most of the examined fish species. These observations may potentially be related to the drastic variations in global climate that occurred during the Neogene and the Quaternary periods. Over several million years, climatic glaciation movements appeared in a relatively fixed order and the cycle repeated every 100,000 years or less [32]. Consequently, fish species were forced to adapt to new environmental circumstances during interglacial periods. Climatic glaciation movements may have contributed to the severe decline in population sizes, or even in some cases of extinction. During the mild interglacial periods, however, the population sizes recovered rapidly (see more details in Fig. 2b).

Ten of the fifteen examined fish species demonstrated drastic reductions in N_e at the beginning of the LGP. This either occurred just before the LGP or at the beginning of this period (Fig. 2b). For example, the green arowana reduced from approximately 180,000 individuals at the beginning of the LGP to 9,600 by the end. The stickleback experienced a dramatic five-fold reduction from approximately 1,540,000 to 31,000 individuals during the same period.

Fluctuations of Effective Population Expansions and Contractions

Eleven of the fifteen examined fish species in our analysis showed that N_e had changed dramatically over the observation time. Fluctuations of effective population expansions and contractions had considered to be a general feature of most of fishes.

More specifically, the N_e among three varieties of Asian arowana (*Scleropages formosus*) varied remarkably (Fig. 3). Before 2 mya, three varieties presented relatively similar N_e values. Since then, however, the green arowana had higher N_e values (100,000–240,000) than the other two varieties, reaching the first peak (240,000) at 600–800 kya; after that, it experienced a reduction to 80,000 until approximately 60 kya. The red and the golden arowana demonstrated slight variations with N_e values of 50,000–100,000 until approximately the beginning of the LGP, when the red arowana experienced an undulation of rising steadily first and then falling to 50,000 until 50 kya; however, the golden arowana experienced an opposite trend simultaneously from 70,000 decreasing to around 50,000 and then rising again to 130,000 until 50 kya. At the end of the examined period (until 10 kya), the N_e of golden arowana was approximately three-fold greater than the red arowana and approximately two-fold greater than the green arowana. In our previous study [33], it was predicted that Asian arowana and spotted gar diverged 384 mya. Since then, three varieties diverged approximately 1–4 mya to evolve as an independent lineage. Green arowana diverged from the other two varieties approximately 3.8 mya, and the latter two varieties diverged approximately 1.7 mya. The N_e variations could briefly reflect the evolutionary trajectories of this fish. Prior to 2 mya, three varieties of arowana had similar population sizes, with a slightly lower value in the red arowana. Subsequently, the green arowana diverged from the other two varieties and experienced a different evolutionary process, with a remarkable expansion in population (the first and second peaks). Although the red arowana diverged from the golden arowana approximately 1.7 mya, both varieties experienced similar changes in N_e from 1 mya.

In 2016, Liu et al. [34] completed the genome sequencing project of channel catfish (from an American population), and generated a draft reference genome assembly with a total length of 783 Mb and a scaffold N50 of 7.7 Mb. Subsequently, we reported another high-quality genome assembly of channel catfish (from a Chinese population; Chen et al., 2016), with approximately 845 Mb in total length and a scaffold N50 of 7.7 Mb (see Table 1). Although the Chinese population has been recognized as the introduction from America, genetic characteristics of the Chinese population seem to exhibit significant differences from the American counterpart after 30 years of artificial breeding in China [35]. These differences were also witnessed from the present prediction of the temporal dynamics in the N_e . The similar trend of N_e was observed from 10 to 100 kya; prior to that, the tendencies of N_e for both populations experienced dramatic differences (see Fig. 3). The heterozygosity in the Chinese population was 0.2%, which was higher than the American population (0.155%). These results confirmed the observed differences at a genomic level between the two populations.

Table 1
Summary of the examined teleost species and their genomes

Common name	Scientific name	Assembly version	Genome size (bp)	Gene number
Zebrafish	<i>Danio rerio</i>	GRCz11	1,373,471,384	25,709
Mexican tetra	<i>Astyanax mexicanus</i>	GCA_000372685.1	1,191,242,572	23,041
Atlantic herring	<i>Clupea harengus</i>	GCA_000966335.1	807,711,962	23,095
Japanese flounder	<i>Paralichthys olivaceus</i>	GCA_001904815.2	545,775,252	21,787
Tongue sole	<i>Cynoglossus semilaevis</i>	GCA_000523025.1	470,199,494	21,381
Three-spined stickleback	<i>Gasterosteus aculeatus</i>	BROAD S1	461,533,448	20,772
Channel catfish (BG) ¹	<i>Ictalurus punctatus</i>	PRJNA319455 ⁴³	845,391,728	21,937
Large yellow croaker	<i>Larimichthys crocea</i>	GCA_000972845.1	678,938,134	24,418
Asian seabass	<i>Lates calcarifer</i>	GCA_001640805.1	668,481,366	24,348
Nile tilapia	<i>Oreochromis niloticus</i>	GCA_000188235.1	927,383,394	21,437
Asia arowana ²	<i>Scleropages formosus</i>	GCA_001624265.1(Golden)	777,359,276	22,016
		GCA_001624245.1(Green)	746,544,453	21,524
		GCA_001624255.1 (Red)	738,407,480	21,256
Spotted green pufferfish	<i>Tetraodon nigroviridis</i>	TETRAODON 8.0	358,618,246	19,583
Fugu	<i>Takifugu rubripes</i>	FUGU5	358,618,246	18,505

¹ The high-quality genome assembly was available in GigaScience Database at <http://dx.doi.org/10.5524/100212>.

² The genome information of golden variety of Asian arowana was used for the divergence time evaluation.

³ The genome assembly of grass carp was downloaded from the Grass Carp Genome project at <http://www.ncgr.ac.cn/grasscarp/>.

⁴ The outgroup for construction of the divergence time tree (see Fig. 1).

Common name	Scientific name	Assembly version	Genome size (bp)	Gene number
Grass carp ³	<i>Ctenopharyngodon idella</i>	—	1,076,149,922	32,785
Atlantic salmon	<i>Salmo salar</i>	GCA_000233375.4	2,966,890,203	43,899
Atlantic cod	<i>Gadus morhua</i>	gadMor1	832,114,588	20,083
Medaka	<i>Oryzias latipes</i>	ASM223467v1	734,057,086	19,669
Amazon molly	<i>Poecilia formosa</i>	GCA_000485575.1	748,923,461	23,613
Southern platyfish	<i>Xiphophorus maculatus</i>	GCA_000241075.1	729,662,853	20,367
Spotted gar	<i>Lepisosteus oculatus</i>	GCA_000242695.1	945,878,036	18,328
Coelacanth	<i>Latimeria chalumnae</i>	GCA_000225785.1	2,860,591,921	19,555
Tropical clawed frog ⁴	<i>Xenopus tropicalis</i>	GCA_000004195.1	1,511,735,326	18,442
¹ The high-quality genome assembly was available in GigaScience Database at http://dx.doi.org/10.5524/100212 .				
² The genome information of golden variety of Asian arowana was used for the divergence time evaluation.				
³ The genome assembly of grass carp was downloaded from the Grass Carp Genome project at http://www.ncgr.ac.cn/grasscarp/ .				
⁴ The outgroup for construction of the divergence time tree (see Fig. 1).				

Nile tilapia had a stable N_e (around 1,000,000) until approximately 1 mya, when it increased to a peak of 2,500,000. Subsequently, tilapia remained stable for 600,000 years until approximately 220 kya, when it experienced a drastic reduction in N_e to 200,000 within the LGP (see Fig. 3). Recently, a high-quality genome assembly of another tilapia species, blue tilapia (*Oreochromis aureus*), was reported by us [36], which had a significant variation in population size compared to the Nile tilapia (Fig. 4). More specifically, from 6 to 2 mya the N_e of the blue tilapia presented an approximately 6-fold increase (at the peak) around 2–3 mya, when it reached 5,700,000 at 2 mya (see Fig. 4). During the following 1 million years, the N_e of blue tilapia declined rapidly. Since then, the population size of blue tilapia has never recovered, but it was steadily declining to less than 200,000. The two sequenced tilapia species diverged approximately 23.2 mya, suggesting that they experienced different evolutionary processes during the examined period. The population of blue tilapia flourished until the junction of the Neogene and Quaternary periods. However,

blue tilapia experienced one cycle of population expansion and contraction (see Fig. 4). Possibly due to the decline in temperature within the LGP and the low tolerance of tilapias to low temperature, these two species had low N_e values during the LGP.

The effective population size of two sequenced flatfishes, half-smooth tongue sole and Japanese flounder, also were fluctuating. Before the LGP, the two flatfish species had similar trends of the N_e with a steady increase over time. Both flatfishes experienced a sharp decline since the beginning of LGP. The N_e of Japanese flounder reached its lowest level around 100 kya, which was 1 kya earlier than the half-smooth tongue sole. The most significant variance between the two flatfishes occurred 60 kya, when the population of the half-smooth tongue sole was maintained at the lowest level while the population of the Japanese flounder rapidly increased to high levels (Fig. 3).

The population sizes varied significantly in Atlantic herring and spotted green pufferfish (Fig. 3) over time, especially within the LGP. The N_e values of these two fish reached peaks at 0.2–1 mya. Interestingly, the increase in populations was observed in these species during the LGP. Medaka (Fig. 3) demonstrated N_e changes within a shorter period than the other species. It is proposed that the low coverage of the SNP data set (Supplementary Table S1) may contribute to this phenomenon.

Other teleost species

Four fishes with a shorter generation time, including stickleback, zebrafish, Mexican tetra and blue tilapia mentioned above, presented similar population trends (Fig. 4). Each species showed a population rise at first and then a decline with differing times to reach peak size. The peak population of stickleback reached 1,550,000 at the beginning of the LGP; in contrast, zebrafish and Mexican tetra trended to reach peaks approximately 0.3 mya.

However, fugu, comparable with medaka, also shown N_e changes within a shorter period than the other species. Its N_e presented a rise during the LGP (Fig. 4).

Discussion

Global climatic oscillations have been shown to play a vital role in the proliferation and colonization of life on our planet [37]. Five mass extinctions have been identified previously in earth's history, primarily related to dramatic climate changes [3]. Uncovering temporal dynamics of effective population size can help better understand evolution in light of global climate variations [38]. Predictions of historical population sizes can also help to evaluate the impact of human activities on biodiversity [39].

The effective use of public genome datasets in evaluating fish population sizes

The PSMC method, which relies on whole-genome variation information, has been developed as an efficient approach to calculate the N_e value of any organism [22]. For example, Tollis et al. [40] compared

the demographic difference between two species of North Atlantic Humpback whales based on the PSMC calculation. It was concluded that the difference between PSMC trajectories may be caused by recent admixture, intraspecific variation, and population structure. Conversely, errors generated by variations in the sequence coverage may also contribute in part to the difference. Barth et al. [41] revealed similar demographic histories of Atlantic and Indo-Pacific yellowfin tuna, suggesting that the two populations diverged only recently. We and our collaborators also applied PSMC to analyze the N_e of the Kanglang white minnow [26], and concluded that the population demography was highly consistent within the three development periods of local habitat in the Fuxian Lake. In addition, our research group reported the first genome assemblies of amphibious fish, mudskipper [23]. In this study, a tight correlation between demographic expansion, or bottleneck events, was well established between two representative mudskippers and sea-level fluctuations.

Here, with public available of high-quality genome assemblies and sequencing data, PSMC was used to predict the temporal dynamics of several species of fish. The results presented here depicted multiple rounds of population contraction and expansion in most representative fish species during the Neogene and the Quaternary periods. In fact, 67% (10/15) of the fish examined showed a drastic reduction in the N_e values before the beginning of the LGP (110–12 kya). Nearly all of the examined species had a lower N_e during very recent periods, in comparison to the peak(s), which can be explained by increasingly dramatic glaciation over the past few million years. These results provide a comprehensive summary of the historical N_e changes in fish, and can also help us to protect endangered species amidst future global climate changes.

Ambiguous correlations between heterozygosity and effective population size

Whole-genome heterozygosity has been widely accepted as a measure of genetic variation in many species [42]. Heterozygosity also provides useful information of long-term N_e changes [43]. A high-level of heterozygosity usually represents a large N_e value, as only sufficiently large populations can harbor extensive genetic diversity [44]. In contrast, a low-level of heterozygosity reflects a small N_e , as small populations may easily lose genetic diversity [42]. Based on a probabilistic model of coalescence with recombination and variations in heterozygosity over time, the PSMC algorithm can predict N_e values in an individual diploid species.

In the present study, a clear relationship was drawn between heterozygosity and N_e . For example, zebrafish showed the highest level of heterozygous sites among the species examined. However, the N_e values, whether the peak of N_e or the amplitude of variation, in zebrafish were not the most remarkable (Fig. 2b). In fact, the peak of N_e value in zebrafish was 1,150,000 at approximately 300 kya, which was far below the 1,550,000 of stickleback, the 2,500,000 of Nile tilapia, and the highest N_e among these fish species at 5,700,000 for the blue tilapia. Zebrafish experienced only one round of N_e increasing and decreasing, in comparison with multiple rounds of observed fluctuations in most other species.

In the practice of PSMC analysis, mutation rate estimation can be a predictor of sequence biodiversity. Purifying selection, like inbreeding and genetic drift, can lead to reduction of heterozygosity and thereby loss of heterozygous sites [45]. Practically speaking, mutations in the PSMC analysis are independent of one another. The influence of reduced heterozygosity due to selection is correlated to the results of a reduced mutation rate.

Low coverage of genomic regions can also lead to an underestimation of the N_e [46]. For those diploid genomes with low coverage of sequencing, heterozygotes will be randomly lost due to the lack of coverage of both alleles. This also has a similar effect as the lower mutation rate [47]. In the present study, Mexican tetra and medaka with the lowest genomic coverage (86.2% and 79.3%, respectively; Supplementary Table S1) exhibited a shorter period of N_e (Fig. 3, 4).

Drawbacks of the PSMC method

Similar to any other population genetics method for analyzing high-throughput sequencing data, the PSMC is influenced by sequencing errors and missing data [46]. Genotyping sites with low coverage can lead to misidentification of heterozygous sites, which will be classified as false positive homozygous sites [48]. By comparing different degrees of genome-wide coverage and missing data rate in black-and-white *Ficedula* flycatchers, Nadachowska-Brzyska et al. [25] proposed that the appropriate genome-wide coverage should be higher than 18-fold with less than 25% of missing data for a reasonable PSMC analysis.

With the exception of the quality of sequencing data, the parameter estimation for PSMC may also play a critical role in demographic analysis, such as using the scaling parameters for an explanation of the results. Proper mutation rate (u) and generation time (g) are the prerequisites for scaling the outputs of PSMC analyses. Thus, if these two parameters are misjudged, any PSMC analysis would become tendentious. Fortunately, however, both mutation rate and generation time affect the PSMC results in a fixed manner. The principal pattern of a PSMC curve can be corrected by sliding the target curve along the axes [24]. In the present study, the divergence time estimation was accurate, which is consistent with the analyses of multiple nuclear gene sequences [49] and Fish T1K transcriptomes [50]. With the literature supported generation time, our PSMC analysis provided trustful outputs.

Despite its wide usage, the PSMC method is restricted to a limited number of samples. This may lead to an impossible inference of recent population size history. PopSizeABC, a new Bayesian computational approach, uses a large set of samples and allows an inference to the evolution of the N_e over recent time [51]. In contrast to the PSMC with only reconstruction of the scenarios before 10 kya, this novel method can infer the most recent (even the last 100 generations) population size of an examined species. This method uses a small number of statistics related to allele frequencies and linkage disequilibrium. Recently, the PopSizeABC method has been well applied in four cattle breeds and in an endangered bird, the crested ibis [52].

Methods

Genome Data Collection

For these public teleost sequencing projects, relevant genome data were downloaded through several public databases, including NCBI, Ensembl, and several professional databases (such as the Grass Carp Genome Project; see more details in Table 1). The genome assembly and raw sequencing data of each fish species were obtained individually.

Heterozygous SNP Calling and Heterozygosity Estimation

For each species, the ALN module of Burrows-Wheeler Alignment (BWA) v0.7.12 [53] was used with default parameters to align raw sequencing data with the corresponding genome assembly (Supplementary Table S1). Subsequently, Picard-tools v1.117 (<http://broadinstitute.github.io/picard/>) “SortSam” and “MarkDuplicates” were used to sort and delete duplicates in the alignment output. Then, Samtools v1.5 [54] “mpileup” module with the parameter “-C50” and the Bcftools v1.5 [54] “call” module with the parameter “-vm0 v -V indels” were used to generate the raw SNP dataset. Based on the average mapping depth of reads, the minimal read depth was set to one third of the average mapping depth and the maximal depth was set to twice the average mapping depth in order to discard low-quality SNP markers. Using a local Perl script, SNPs with the distance to the next marker less than 10 bp was discarded. The final dataset of refined heterozygous SNPs was obtained for each species examined.

The heterozygosity rate of each species was calculated based on the number of heterozygous SNPs and the genome size of each species.

Mutation Rate Estimation

To obtain the mutation rate of each species, a divergence time tree of related species (Fig. 1) was constructed. For this analysis, the protein coding sequences of teleosts were collected (see Table 1). In total, a 628 one-to-one orthologues geneset was generated from 22 species by using the Blastp (Altschul et al., 1990) and Hcluster_sg [55] with a parameter of “-w 10 -s 0.34”. The Bayesian Inference (BI) method was applied to perform phylogenomic analyses. MrMTgui [56] program was employed to complete the best-fit nucleotide substitution model test. The general time reversible (GTR) model with optimization of substitution rates and Gamma model of rate heterogeneity (GTR + I + G) were the best-fit. Finally, a phylogenetic tree was constructed using MrBayes v3.1.2 [57] with the parameter setting as “ngen = 100000”. Branch support confidence was assessed by Bayesian posterior probabilities. Based on the topology generated from the phylogenomic analysis, the divergence time of these species was inferred by applying the Mcmctree module of PAML package [58] with the parameter “clock = 2, model = 4”. Six branch nodes were calibrated using fossil records.

Almost all of our divergence age estimations were accurate. We estimated that the *P. formosa* diverged with *X. maculatus* at very recent year (9.4–27.9 mya), which agrees with the analyses of multiple nuclear gene sequences [49] and Fish T1K transcriptomes [50]. We estimated the divergence time between the Ostariophysi (*D. rerio* and *I. punctatus*) and the Clupeiformes at 242.3 mya (95% interval: 211.5–272.0 mya), which is also consistent with the above mentioned analyses. The divergence time between *I.*

punctatus and *A. mexicanus* was 118 mya (95% HPD interval: 104.6-143.4 mya), which is also validated in the above mentioned analyses. The accurate divergence time estimations indicated our mutation rate calculation more convincingly.

For the second step, the zebrafish genome was used as the reference to obtain whole-genome level pairwise alignments between zebrafish and other fish species using the LASTZ program v1.02.00 [59] with the parameter setting as “-step = 19 -hsptresh = 2200 -inner = 2000 -ydrop = 3400 -gappedthresh = 10000 -format = axt”. The length of alignment block without any gaps was ascertained and Ns was counted. The number of variation sites in each alignment block was also counted. The following formula was applied to calculate the mutation rate of each fish: $u = Nv/Len/2t$, where u is the mutation rate, Nv is the number of variation sites, Len is the length of the alignment block, and t is the divergence time between zebrafish and each examined species based on the phylogenomic analysis (Supplementary Table S2).

Demographic History

To predict the demographic history of each fish species, the pairwise sequentially Markovian Coalescent (PSMC) model [22] was applied with the parameter “-N 30 -t 15 -r 5 -b -p "4 + 10*1 + 20*2 + 4 + 6””. Only those heterozygous SNP loci with minor allele frequency (MAF) ≥ 0.2 were used. Subsequently, 100 bootstraps were operated for each species to examine the variance in N_e estimates. Finally, the results were scaled by using the generation time and mutation rate of each species to construct the demographic history from 10 kya to 20 mya.

Declarations

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

Q.S. and J.L. conceived the project. J.L. analyzed the data. C.B., Y.Y., H.Y., and X.Y. participated in data analysis and discussion. J.L. wrote the manuscript. Q.S. revised the manuscript. All authors approved submission of the final manuscript.

Funding

This research was supported by Shenzhen Dapeng Special Program for Industrial Development (no. PT201901-08) and Shenzhen Special Program for Development of Emerging Strategic Industries (no. JSGG20170412153411369).

References

1. Hewitt G: **The genetic legacy of the Quaternary ice ages.** *Nature* 2000, **405**(6789):907.
2. Hewitt G: **Genetic consequences of climatic oscillations in the Quaternary.** *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences* 2004, **359**(1442):183-195.
3. Alroy J: **Colloquium paper: dynamics of origination and extinction in the marine fossil record.** *Proc Natl Acad Sci U S A* 2008, **105 Suppl 1**(Suppl 1):11536-11542.
4. McGhee GR, Sheehan PM, Bottjer DJ, Droser ML: **Ecological ranking of Phanerozoic biodiversity crises: the Serpukhovian (early Carboniferous) crisis had a greater ecological impact than the end-Ordovician.** *Geology* 2012, **40**(2):147-150.
5. Trotter JA, Williams IS, Barnes CR, Lécuyer C, Nicoll RS: **Did Cooling Oceans Trigger Ordovician Biodiversification? Evidence from Conodont Thermometry.** *Science* 2008, **321**(5888):550-554.
6. Finnegan S, Heim NA, Peters SE, Fischer WW: **Climate change and the selective signature of the Late Ordovician mass extinction.** *Proceedings of the National Academy of Sciences* 2012, **109**(18):6829-6834.
7. Johnson ME: **Relationship of Silurian sea-level fluctuations to oceanic episodes and events.** *GFF* 2006, **128**(2):115-121.
8. Finney SC, Berry WB, Cooper JD, Ripperdan RL, Sweet WC, Jacobson SR, Soufiane A, Achab A, Noble PJ: **Late Ordovician mass extinction: a new perspective from stratigraphic sections in central Nevada.** *Geology* 1999, **27**(3):215-218.
9. Munnecke A, Calner M, Harper DAT, Servais T: **Ordovician and Silurian sea–water chemistry, sea level, and climate: A synopsis.** *Palaeogeography, Palaeoclimatology, Palaeoecology* 2010, **296**(3):389-413.
10. Young SA, Saltzman MR, Foland KA, Linder JS, Kump LR: **A major drop in seawater $87\text{Sr}/86\text{Sr}$ during the Middle Ordovician (Darriwilian): Links to volcanism and climate?** *Geology* 2009, **37**(10):951-954.
11. Erwin DH: **The end and the beginning: recoveries from mass extinctions.** *Trends in Ecology & Evolution* 1998, **13**(9):344-349.
12. Sheehan PM: **The late Ordovician mass extinction.** *Annual Review of Earth and Planetary Sciences* 2001, **29**(1):331-364.
13. Peñuelas J, Sardans J, Estiarte M, Ogaya R, Carnicer J, Coll M, Barbeta A, Rivas-Ubach A, Llusà J, Garbulsky M: **Evidence of current impact of climate change on life: a walk from genes to the biosphere.** *Global change biology* 2013, **19**(8):2303-2338.
14. Rivas-Ubach A, Sardans J, Pérez-Trujillo M, Estiarte M, Peñuelas J: **Strong relationship between elemental stoichiometry and metabolome in plants.** *Proceedings of the National Academy of Sciences* 2012, **109**(11):4181-4186.
15. Gitay H, Suárez A, Watson RT, Dokken DJ: **Climate change and biodiversity.** 2002.
16. Lorenzen ED, Nogués-Bravo D, Orlando L, Weinstock J, Binladen J, Marske KA, Ugan A, Borregaard MK, Gilbert MTP, Nielsen R: **Species-specific responses of Late Quaternary megafauna to climate and**

- humans. *Nature* 2011, **479**(7373):359.
17. Ho SY, Shapiro B: **Skyline-plot methods for estimating demographic history from nucleotide sequences.** *Molecular ecology resources* 2011, **11**(3):423-434.
 18. Drummond AJ, Rambaut A, Shapiro B, Pybus OG: **Bayesian coalescent inference of past population dynamics from molecular sequences.** *Molecular biology and evolution* 2005, **22**(5):1185-1192.
 19. Pybus OG, Rambaut A, Harvey PH: **An integrated framework for the inference of viral population history from reconstructed genealogies.** *Genetics* 2000, **155**(3):1429-1437.
 20. Beaumont MA: **Approximate Bayesian computation in evolution and ecology.** *Annual review of ecology, evolution, and systematics* 2010, **41**:379-406.
 21. Csilléry K, Blum MG, Gaggiotti OE, François O: **Approximate Bayesian computation (ABC) in practice.** *Trends in ecology & evolution* 2010, **25**(7):410-418.
 22. Li H, Durbin R: **Inference of human population history from individual whole-genome sequences.** *Nature* 2011, **475**(7357):493.
 23. You X, Bian C, Zan Q, Xu X, Liu X, Chen J, Wang J, Qiu Y, Li W, Zhang X *et al*: **Mudskipper genomes provide insights into the terrestrial adaptation of amphibious fishes.** *Nature communications* 2014, **5**:5594.
 24. Nadachowska-Brzyska K, Li C, Smeds L, Zhang G, Ellegren H: **Temporal dynamics of avian populations during Pleistocene revealed by whole-genome sequences.** *Current Biology* 2015, **25**(10):1375-1380.
 25. Nadachowska-Brzyska K, Burri R, Smeds L, Ellegren H: **PSMC analysis of effective population sizes in molecular ecology and its application to black-and-white Ficedula flycatchers.** *Molecular Ecology* 2016, **25**(5):1058-1072.
 26. Jiang W, Qiu Y, Pan X, Zhang Y, Wang X, Lv Y, Bian C, Li J, You X, Chen J *et al*: **Genome Assembly for a Yunnan-Guizhou Plateau "3E" Fish, Anabarrilius grahami (Regan), and Its Evolutionary and Genetic Applications.** *Frontiers in Genetics* 2018, **9**(614).
 27. Yang J, Chen X, Bai J, Fang D, Qiu Y, Jiang W, Yuan H, Bian C, Lu J, He S *et al*: **The Sinocyclocheilus cavefish genome provides insights into cave adaptation.** *BMC Biology* 2016, **14**:1.
 28. Kang J, Ma X, He S: **Population genetics analysis of the Nujiang catfish *Creteuchiloglanis macropterus* through a genome-wide single nucleotide polymorphisms resource generated by RAD-seq.** *Scientific reports* 2017, **7**(1):2813.
 29. Bradley MJ, Kutz SJ, Jenkins E, O'hara TM: **The potential impact of climate change on infectious diseases of Arctic fauna.** *International Journal of Circumpolar Health* 2005, **64**(5):468-477.
 30. Ohlberger J, Mehner T, Staaks G, Hölker F: **Intraspecific temperature dependence of the scaling of metabolic rate with body mass in fishes and its ecological implications.** *Oikos* 2012, **121**(2):245-251.
 31. Andersen KK, Azuma N, Barnola J-M, Bigler M, Biscaye P, Caillon N, Chappellaz J, Clausen HB, Dahl-Jensen D, Fischer H: **High-resolution record of Northern Hemisphere climate extending into the last interglacial period.** *Nature* 2004, **431**(7005):147.

32. Rahmstorf S: **Ocean circulation and climate during the past 120,000 years.** *Nature* 2002, **419**(6903):207-214.
33. Bian C, Hu Y, Ravi V, Kuznetsova IS, Shen X, Mu X, Sun Y, You X, Li J, Li X *et al*: **The Asian arowana (*Scleropages formosus*) genome provides new insights into the evolution of an early lineage of teleosts.** *Scientific Reports* 2016, **6**:24501.
34. Liu Z, Liu S, Yao J, Bao L, Zhang J, Li Y, Jiang C, Sun L, Wang R, Zhang Y: **The channel catfish genome sequence provides insights into the evolution of scale formation in teleosts.** *Nature communications* 2016, **7**:11757.
35. Zhong L, Song C, Chen X, Deng W, Xiao Y, Wang M, Qin Q, Luan S, Kong J, Bian W: **Channel catfish in China: Historical aspects, current status, and problems.** *Aquaculture* 2016, **465**:367-373.
36. Bian C, Li J, Lin X, Chen X, Yi Y, You X, Zhang Y, Lv Y, Shi Q: **Whole Genome Sequencing of the Blue Tilapia (*Oreochromis aureus*) Provides a Valuable Genetic Resource for Biomedical Research on Tilapias.** *Marine drugs* 2019, **17**(7).
37. Bellard C, Bertelsmeier C, Leadley P, Thuiller W, Courchamp F: **Impacts of climate change on the future of biodiversity.** 2012, **15**(4):365-377.
38. Bellard C, Bertelsmeier C, Leadley P, Thuiller W, Courchamp F: **Impacts of climate change on the future of biodiversity.** *Ecology letters* 2012, **15**(4):365-377.
39. Pauls SU, Nowak C, Bálint M, Pfenninger M: **The impact of global climate change on genetic diversity within populations and species.** *Molecular ecology* 2013, **22**(4):925-946.
40. Tollis M, Robbins J, Webb AE, Kuderna LFK, Caulin AF, Garcia JD, Bèrubè M, Pourmand N, Marques-Bonet T, O'Connell MJ *et al*: **Return to the Sea, Get Huge, Beat Cancer: An Analysis of Cetacean Genomes Including an Assembly for the Humpback Whale (*Megaptera novaeangliae*).** *Molecular Biology and Evolution* 2019, **36**(8):1746-1763.
41. Barth JMI, Damerau M, Matschiner M, Jentoft S, Hanel R: **Genomic differentiation and demographic histories of Atlantic and Indo-Pacific yellowfin tuna (*Thunnus albacares*) populations.** *Genome biology and evolution* 2017.
42. Allendorf FW: **Genetic drift and the loss of alleles versus heterozygosity.** *Zoo biology* 1986, **5**(2):181-190.
43. Motro U, Thomson G: **On heterozygosity and the effective size of populations subject to size changes.** *Evolution* 1982, **36**(5):1059-1066.
44. Caballero A: **Developments in the prediction of effective population size.** *Heredity* 1994, **73**(6):657.
45. Palkopoulou E, Mallick S, Skoglund P, Enk J, Rohland N, Li H, Omrak A, Vartanyan S, Poinar H, Götherström A: **Complete genomes reveal signatures of demographic and genetic declines in the woolly mammoth.** *Current Biology* 2015, **25**(10):1395-1400.
46. Han E, Sinsheimer JS, Novembre J: **Characterizing bias in population genetic inferences from low-coverage sequencing data.** *Molecular biology and evolution* 2013, **31**(3):723-735.

47. Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, He W, Chen Y-J, Makhijani V, Roth GT: **The complete genome of an individual by massively parallel DNA sequencing.** *nature* 2008, **452**(7189):872.
48. Alex Buerkle C, Gompert Z: **Population genomics based on low coverage sequencing: how low should we go?** *Molecular ecology* 2013, **22**(11):3028-3035.
49. Near TJ, Eytan RI, Dornburg A, Kuhn KL, Moore JA, Davis MP, Wainwright PC, Friedman M, Smith WL: **Resolution of ray-finned fish phylogeny and timing of diversification.** 2012, **109**(34):13698-13703.
50. Hughes LC, Ortí G, Huang Y, Sun Y, Baldwin CC, Thompson AW, Arcila D, Betancur-R. R, Li C, Becker L *et al*: **Comprehensive phylogeny of ray-finned fishes (Actinopterygii) based on transcriptomic and genomic data.** 2018, **115**(24):6249-6254.
51. Boitard S, Rodriguez 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 genetics* 2016, **12**(3):e1005877.
52. Feng S, Fang Q, Barnett R, Li C, Han S, Kuhlwilm M, Zhou L, Pan H, Deng Y, Chen G: **The genomic footprints of the fall and recovery of the crested ibis.** *Current Biology* 2019, **29**(2):340-349. e347.
53. Li H, Durbin R: **Fast and accurate short read alignment with Burrows–Wheeler transform.** *bioinformatics* 2009, **25**(14):1754-1760.
54. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R: **The sequence alignment/map format and SAMtools.** *Bioinformatics* 2009, **25**(16):2078-2079.
55. Birney E, Clamp M, Durbin R: **GeneWise and Genomewise.** *Genome Research* 2004, **14**(5):988-995.
56. Nuin P: **MrMTgui. v 1.0. MrModelTest/ModelTest Graphical interface for Windows/Linux.** 2007.
57. Ronquist F, Teslenko M, Van dMP, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP: **MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space.** *Systematic Biology* 2012, **61**(3):: 539– 542.
58. Yang Z: **PAML 4: Phylogenetic Analysis by Maximum Likelihood.** *Molecular Biology and Evolution* 2007, **24**(8):1586-1591.
59. Schwartz S, Kent WJ, Smit A, Zhang Z, Baertsch R, Hardison RC, Haussler D, Miller W: **Human–mouse alignments with BLASTZ.** *Genome research* 2003, **13**(1):103-107.

Figures

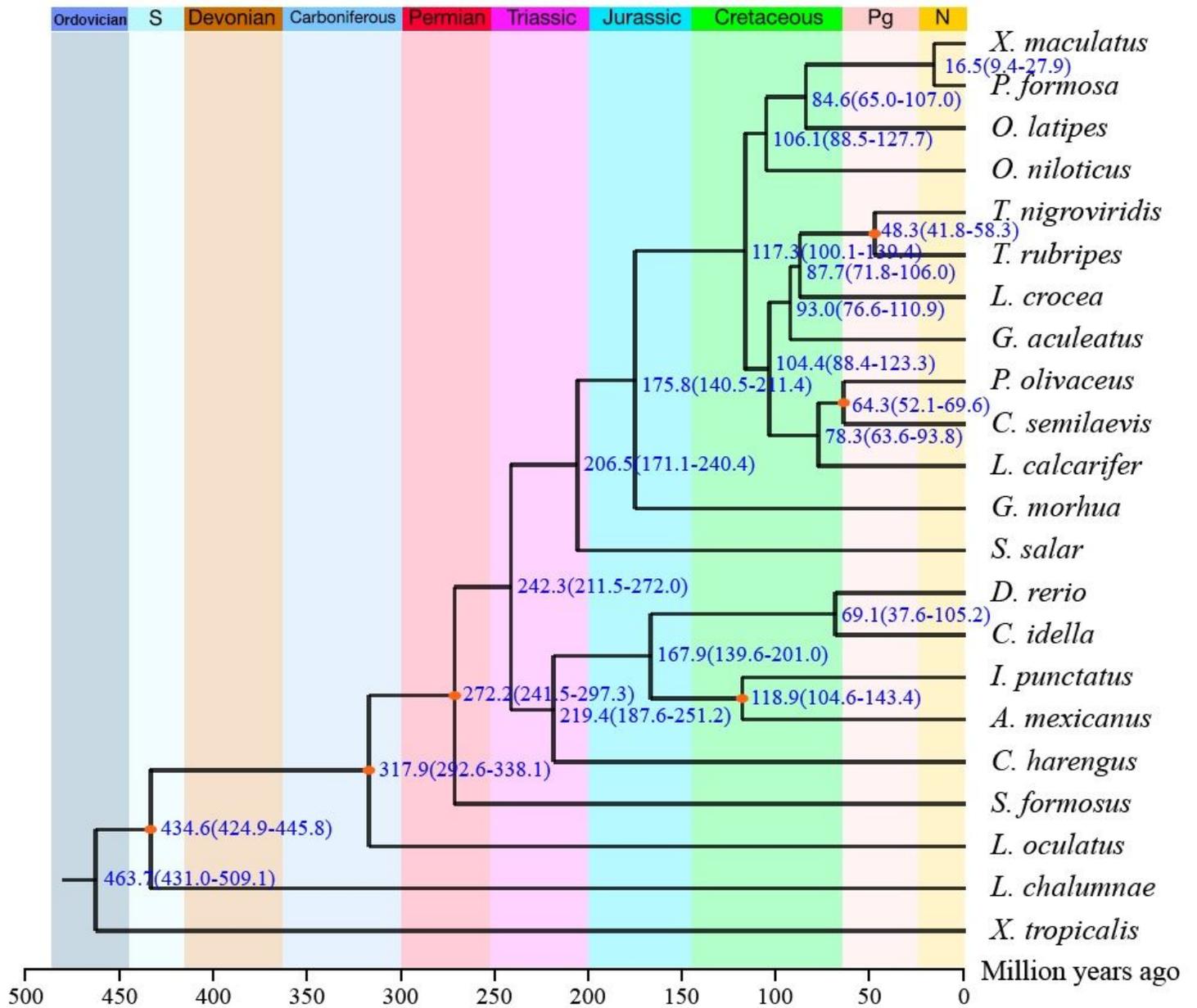


Figure 1

The divergence time tree of 22 representative species. Tropical clawed frog (*X. tropicalis*) was used as the outgroup. The branch nodes with crimson dots were calibrated by using reported fossil records.

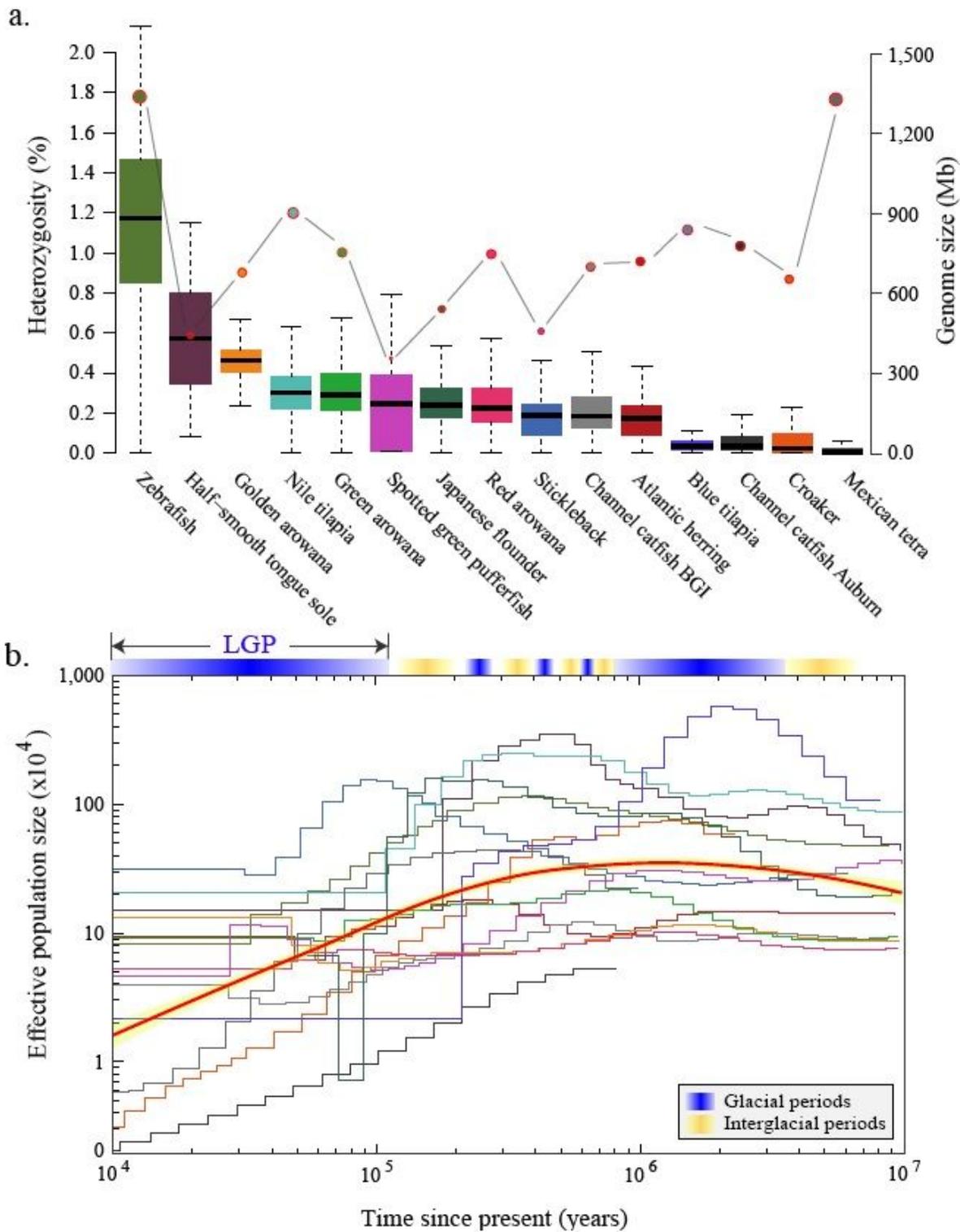


Figure 2

Demographic history and genome-wide heterozygosity of 15 representative fish species with high-quality genome assemblies. (a) The boxplot represents the genome-wide heterozygosity that was estimated from genomic 500-kb windows. Genome sizes of the examined fish species were presented by dots. (b) Historical N_e of the 15 examined fish genomes. The x axis depicts the historical time, and the y axis represents N_e . These plots were scaled using species-specific mutation rates (μ) and generation times (g)

(Table S1). The color of each species is consistent with that of boxplot in (a). Gradient blue bars above the plots represent the glacial periods and the gradient yellow bars represent the interglacial periods. The red smooth line shows the best-fit model with 95% confidence interval. LGP, the last glacial period (110–12 kya).

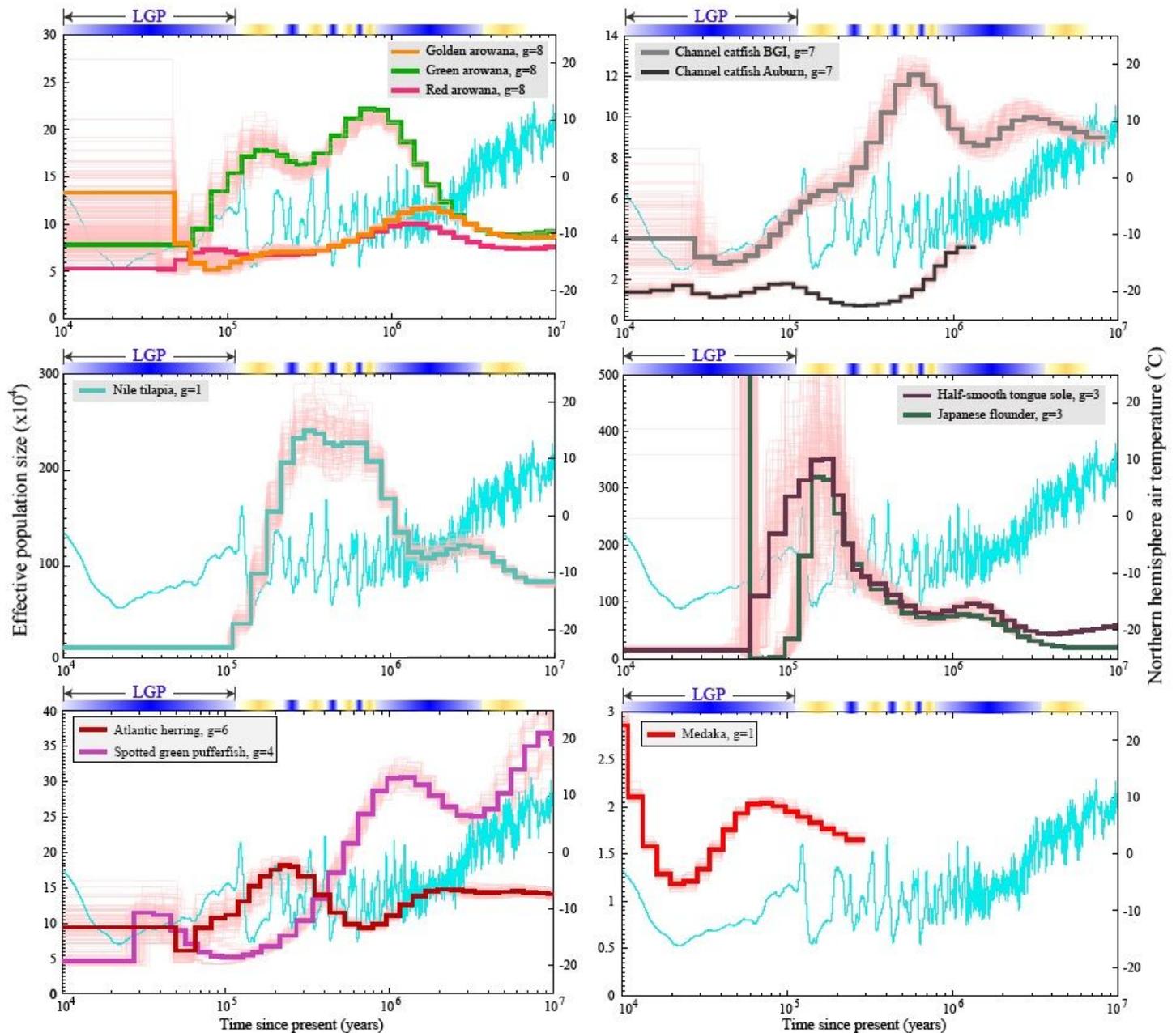


Figure 3

Fluctuations of Effective Population Expansions and Contractions. The pink curves represent PSMC estimates for 100 bootstrapped sequences. The light blue line indicates changes of the Northern hemisphere air temperature [32]. The gradient blue bars above the plots represent the glacial periods and the gradient yellow bars represent the interglacial periods.

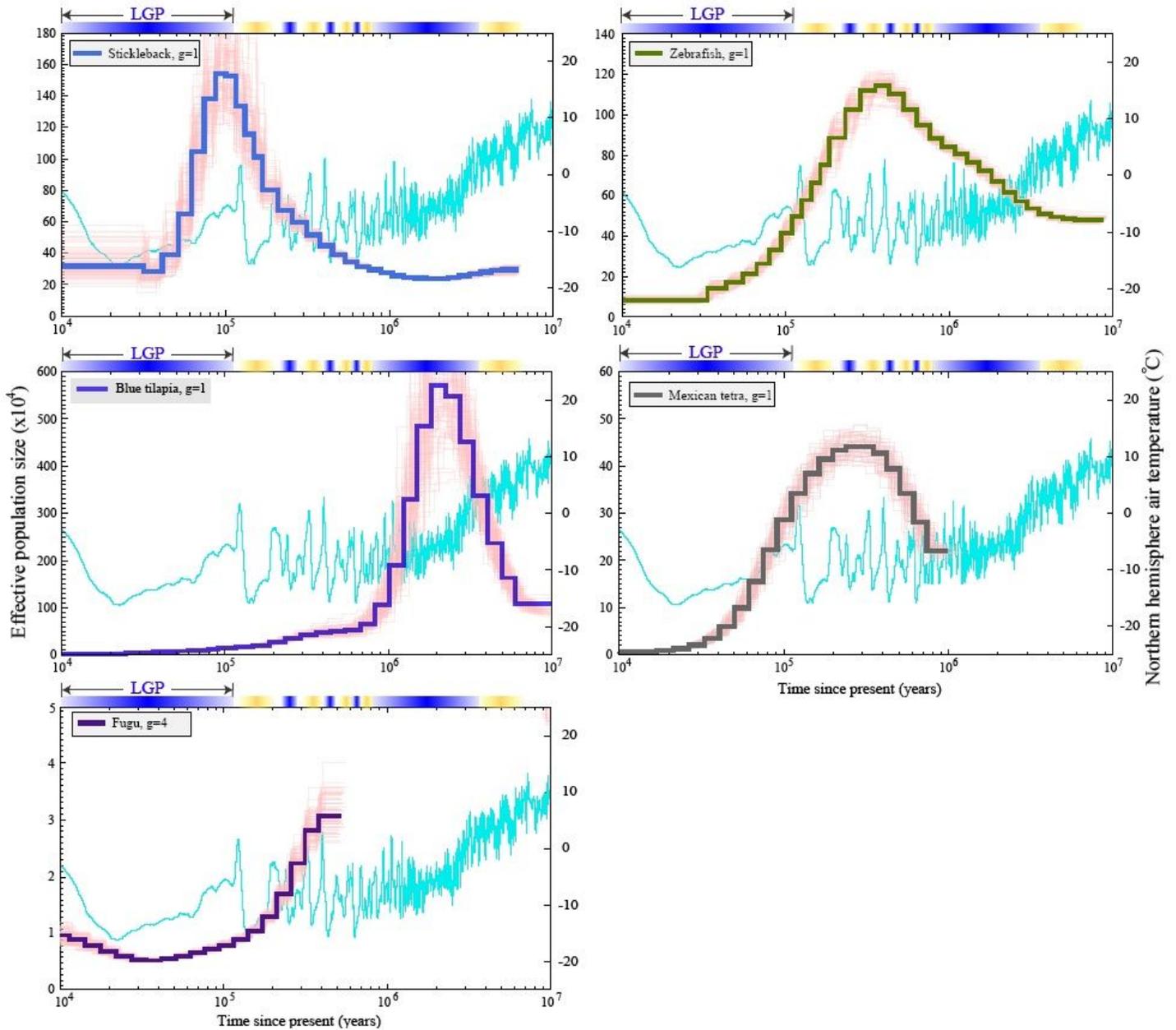


Figure 4

Temporal dynamics of effective population size for other teleost species. The pink curves in each subplot represent PSMC estimates for 100 bootstrapped sequences. The light blue line indicates changes of the Northern hemisphere air temperature [32]. The gradient blue bars above the plots represent the glacial periods and the gradient yellow bars represent the interglacial periods.

Supplementary Files

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

- [SupplementaryTables.docx](#)