REGULAR ARTICLE

JOURNAL OF **FISH** BIOLOGY

Whole-genome analysis reveals phylogenetic and demographic history of Eurasian perch

Vitalii Lichman[1] | Mikhail Ozerov[2,3,4] | María-Eugenia López[3] | Kristina Noreikiene[1,5] | Siim Kahar[1] | Lilian Pukk[1] | Oksana Burimski[1] | Riho Gross[1] | Anti Vasemägi[1,3]

[1]Institute of Veterinary Medicine and Animal Sciences, Chair of Aquaculture, Estonian University of Life Sciences, Tartu, Estonia

[2]Biodiversity Unit, University of Turku, Turku, Finland

[3]Department of Aquatic Resources, Institute of Freshwater Research, Swedish University of Agricultural Sciences, Drottningholm, Sweden

[4]Department of Biology, University of Turku, Turku, Finland

[5]Department of Botany and Genetics, Vilnius University, Vilnius, Lithuania

Correspondence
Vitalii Lichman, Institute of Veterinary Medicine and Animal Sciences, Chair of Aquaculture, Estonian University of Life Sciences, Tartu 51006, Estonia.
Email: vitalii.lichman@emu.ee

Anti Vasemägi, Department of Aquatic Resources, Institute of Freshwater Research, Swedish University of Agricultural Sciences, Drottningholm 17893, Sweden.
Email: anti.vasemagi@slu.se

Abstract

The contemporary diversity and distribution of species are shaped by their evolutionary and ecological history. This can be deciphered with the help of phylogenetic and demographic analysis methods, ideally combining and supplementing information from mitochondrial and nuclear genomes. In this study, we investigated the demographic history of Eurasian perch (*Perca fluviatilis*), a highly adaptable teleost with a distribution range across Eurasia. We combined whole-genome resequencing data with available genomic resources to analyse the phylogeny, phylogeography, and demographic history of *P. fluviatilis* populations from Europe and Siberia. We identified five highly diverged evolutionary mtDNA lineages, three of which show a strong signal of admixture in the Baltic Sea region. The estimated mean divergence time between these lineages ranged from 0.24 to 1.42 million years. Based on nuclear genomes, two distinct demographic trajectories were observed in European and Siberian samples reflecting contrasting demographic histories ca. 30,000–100,000 years before the present. A comparison of mtDNA and nuclear DNA evolutionary trees and AMOVA revealed concordances, as well as incongruences, between the two types of data, most likely reflecting recent postglacial colonization and hybridization events. Overall, our findings demonstrate the power and usefulness of genome-wide information for delineating historical processes that have shaped the genome of *P. fluviatilis*. We also highlight the added value of data-mining existing transcriptomic resources to complement novel sequence data, helping to shed light on putative glacial refugia and postglacial recolonization routes.

KEYWORDS
demography, mtDNA, phylogeny, phylogeography, SNPs, whole-genome sequencing

Mikhail Ozerov is the first author.

1 | INTRODUCTION

Understanding how geological, ecological, and demographic processes have shaped the genetic variation in extant species has been a long-standing question in evolutionary biology (Coyne, 1994; Wall, 2003; Yoder et al., 2010). Major breakthroughs in sequencing technology have allowed us to resolve the evolutionary relationships and reconstruct the demographic history of a species from the genomic data obtained from its present-day representatives (Beichman et al., 2018; Mather et al., 2020). As a result, the genome-wide perspective improves the discriminatory power and enhances our understanding of the speciation process (Arbogast & Kenagy, 2001; Hewitt, 2001), adaptive divergence (Ramette & Tiedje, 2007), and the history of migration, dispersal, and hybridization (Hewitt, 1999; Taberlet et al., 1998).

Eurasian perch (*Perca fluviatilis* L.), yellow perch (*Perca flavescens* [Mitchill 1814]), and Balkhash perch (*Perca schrenkii* [Kessler 1874]) are three species of the genus *Perca* from the family *Percidae* (Collette & Bănărescu, 1977). The native range of Eurasian perch encompasses the major part of the north temperate zone of Eurasia, stretching out from the Iberian Peninsula in the west to the Chukotka region in the Russian Far East, reaching the Caucasus on the south and partially covering Arctic regions in the north (Collette & Bănărescu, 1977; Thorpe, 1977). Besides their tolerance to a broad temperature range, Eurasian perch can be found in diverse freshwater habitats and even in brackish water environment (Christensen et al., 2019; Diehl, 1992; Nilsson et al., 2004; Ozerov et al., 2022). In many lakes, Eurasian perch (hereinafter perch) acts as a keystone species exerting a large impact on nutrient cycling and food webs through predation (Persson et al., 2003). Furthermore, recent findings have demonstrated that perch has a significant role in carbon cycling via methane efflux, through its effects on the zooplankton predation (Devlin et al., 2015). Perch is also a popular target for fisheries (Watson, 2008) and a promising candidate species for aquaculture (Policar et al., 2015).

The ubiquitous spatial distribution and wide environmental tolerance of perch are of particular interest from the phylogeographic and ecological perspectives. For example, the determination of relationships between contemporary populations and their approximate time of divergence from common ancestral groups can shed light on potential glacial refugia and the routes of postglacial recolonization. Earlier studies based on short mtDNA fragments have suggested the presence of several evolutionary diverged lineages of perch in Europe and Siberia (e.g., Nesbø et al., 1999; Ragauskas et al., 2020; Toomey et al., 2020). However, reliable delineation of phylogenetic relationships based on short DNA fragments is challenging (Vasemägi et al., 2023; Wilke et al., 2013), and frequently, short stretches of DNA provide insufficient resolution to accurately reconstruct the evolutionary history within species (Erpenbeck et al., 2006; Kivisild, 2015). Thus, genome-scale data fueled by the high-throughput sequencing technologies might fill this gap to advance the knowledge of phylogenetic relationships and demographic history of perch (Foster et al., 2009; Gronau et al., 2011).

Here, we aim to provide a whole-genome perspective of the phylogenetic and demographic history of *P. fluviatilis* by merging newly generated whole mitochondrial and nuclear genome data with publicly available genomic and transcriptomic resources. First, we characterized phylogenetic relationships based on near-complete mitochondrial genomes and identified several evolutionary lineages that likely reflect distinct glacial refugia and subsequent postglacial colonization patterns. Second, we estimated the divergence times between distinct lineages based on complete mtDNA genomes. Third, we tracked the demographic changes over approximately 300,000 years by using pair-wise sequentially Markovian coalescent (PSMC) analysis using >1 million nuclear single nucleotide polymorphism (SNPs). Finally, we examined how genetic structuring in perch is manifested in mitochondrial versus nuclear genomes by characterizing both similarities and inconsistencies between their evolutionary trees. Overall, our findings demonstrate the power of genome-wide information to refine existing and gain new knowledge on the historical processes that have shaped the genetic diversity and structuring of Eurasian perch.

2 | MATERIALS AND METHODS

2.1 | Data and sample collection

The entire dataset used in this study, covering 66 distinct sampling sites, comprises 114 individual and pooled mitochondrial genomes, along with 39 individual nuclear genomes. Individual samples were collected from 32 lakes in Europe (Ozerov et al., 2022), supplemented with seven newly sequenced specimens obtained from Western Siberia, Russia (Table 1; Table S1). Both the mitochondrial and nuclear genomes for these 39 samples were derived from the same individuals. Additionally, mitochondrial genomes from 55 pool-seq datasets from different watercourses (with 8–41 individuals in each pool) were incorporated (Table 1). Furthermore, we retrieved 19 individual and pooled RNA-seq-derived mtDNA sequences from the publicly available GenBank database, along with the mitochondrial reference genome (NCBI: KM410088.1). Locations for individual and pool-seq sampling partially overlapped, mainly within lakes across Estonia, Finland, Sweden, Lithuania, and Russia. Specimens representing the assembled RNA-seq-derived mtDNA subset were obtained from GenBank, and originated from Belarus, Ukraine, Sweden, France, China, and Hungary (accession numbers are available at Table S1).

Newly analysed fish were captured using gillnets, rods, or beach seine during the 2016–2020 fieldwork seasons. After capture, individuals were killed with a sharp blow to the head, and tissue samples were taken and stored in 96% ethanol. Samples were collected in accordance with national legislation based on permits issued by the Estonian Ministry of Environment (54/2016; 37/2017; 10-1/18/29; 10-1/19/23; 10-1/20/37) and the regional ethical review board in Uppsala, Sweden (Dnr 5.8.18-03449/2017). Sampling in Finland, Lithuania, and Russia was carried out using recreational fishing gear (rod and line), following the fishing rules set by the national

TABLE 1 Summary table of *Perca fluviatilis* samples and datasets used for different types of analyses (detailed information available in Table S1).

Type of analysis	Source of data					Number of genomes	
	Individual WGS	Pooled WGS	Individual RNA-seq	Pooled RNA-seq	mtDNA genome (NCBI)	mtDNA	nDNA
Phylogenetic relationships	39	55	13	6	1	114	
Haplotype network	39		1		1	41	
BEAST	12				1	13	
PSMC	39						39
AMOVA	39		1		1	41	39
Tanglegram	39					39	39
Admixture, f3, Ho	39						39

Abbreviations: mtDNA, mitochondrial DNA; nDNA, nuclear DNA; PSMC, pair-wise sequentially Markovian coalescent; WGS, whole genome sequencing.

legislations, and ethical issues related to recreational fishing and handling of caught fish recommended by the recreational fishing federations of the countries. Ethical aspects of sampling were conducted according to the requirements of Annex IV (Methods of killing animals) Section B point 11 of the "Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes."

2.2 | DNA extraction

Total genomic DNA (gDNA) from seven Western Siberian samples (pelvic fin) was extracted using NucleoSpin tissue kit (Macherey-Nagel) following the manufacturer's protocol. The concentration of total gDNA was measured using Qubit fluorometric quantification (Invitrogen), and the quality of gDNA was estimated by Fragment Analyzer (Advanced Analytical). For pooled samples, total gDNA was extracted from 1281 individuals using the same method, and the concentration was measured using a Nanodrop2000 spectrophotometer. The gDNA was then diluted to a concentration of 50 ng/µL, and 10 µL of gDNA from each sample was pooled equimolarly (Lopez et al., in preparation).

2.3 | Library preparation and sequencing

Sequencing libraries for individual Siberian samples were constructed following the Illumina TruSeq DNA PCR-Free Library Preparation Guide from 1 µg of total gDNA, with a mean insert size of 350 bp. For the preparation of pool-seq libraries, we used the Illumina DNA PCR-Free Library Preparation Kit according to manufacturer's protocol from 0.3 µg of total gDNA. The samples were then normalized and pooled for automated cluster preparation using an Illumina cBot station. Libraries were pooled and sequenced in four lanes on an Illumina NovaSeq 6000 using paired-end sequencing (2 × 150-bp read length with an 8-bp index).

2.4 | Extraction of mitochondrial data from the publicly available RNA-seq data

The RNA-seq data were retrieved from the GenBank, comprising perch samples collected from Sweden, France, Belarus, and Ukraine, as well as China (Table 1; Table S1). We compiled the sequencing read archives (SRA) from each BioProject and extracted sequence read archive runs (SRR) using SRA Toolkit ver. 3.0.0 (SRA Toolkit Development Team). The read quality of raw RNA-seq data was estimated using FASTQC ver. 0.11.8 (Andrews, 2010). Trimming of heads and tails containing low-quality regions, short (<60 bp), low-quality reads (mean Phred score <25), and Illumina adapters was performed using trimmomatic ver. 0.36 (Bolger et al., 2014).

2.5 | Quality control and generation of consensus mtDNA sequences

To assess the read quality of seven individual samples from Siberia and pool-seq samples, we used FASTQC. For trimming, fastp ver. 0.20 (Chen et al., 2018) was used to remove short (<60 bp) and low-quality reads (mean Phred score <25) along with Illumina adapters. Filtered paired-end sequence reads of individual genomic and transcriptomic data, as well as pool-seq genomic data, were mapped to the mitochondrial reference genome (NCBI: KM410088.1) of perch using bowtie2 ver. 2.3.5.1 (Langmead & Salzberg, 2012) with default parameters except for the modified score minimum threshold (−score-min L, −0.3, −0.3) and maximum fragment length for valid paired-end alignments criteria (−X 700). Aligned consensus mtDNA sequences were extracted using bcftools ver. 1.17 (Li, 2011) applying a minimum mapping quality of 20 (−q 20) and converted to fasta format using seqtk ver. 1.3 (Li, 2013). The final mitochondrial dataset comprised 114 complete or near-complete mtDNA genomes derived from 39 individual and 55 pooled WGS datasets, complemented with mtDNA reference genome sequence (KM410088.1) and mtDNA genomes obtained from 19 RNA-seq datasets from GenBank (Table 1; Table S1).

2.6 | SNP calling from individual sequences

SNPs for 39 individually sequenced samples were called following similar procedures as in Ozerov et al. (2022). Filtered sequence reads of each individual were mapped to the perch reference genome (NCBI: GCA_010015445.1) using bowtie2, applying the same parameters as for mtDNA consensus sequence generation. SNPs were called using two alternative pipelines. First, bcftools ver. 1.17 (Li, 2011) mpileup was applied on the locally realigned and sorted BAM files to generate genotype likelihoods with the following SNP calling. Second, GATK ver. 4.1.4.1 (McKenna et al., 2010) pipeline was applied to call variants using the same BAM files with the following import of single-sample GVCF files into GenomicsDB using GenomicsDBImport, and final calling of consensus genotypes with GenotypeGVCFs using the same parameters as in Ozerov et al. (2022). Further, genomic variants generated by the two pipelines were filtered using vcftools ver. 0.1.15 (Danecek et al., 2011) applying the following parameters: (i) minimum and maximum mean sequencing depth (d) was set to 10 and 66, respectively; (ii) the consensus quality was ≥30; (iii) only bi-allelic sites were included; (iv) a variant had at least two copies of an allele; (v) a variant did not occur in repetitive genomic regions; (vi) minor allele frequency (MAF) was 0.05; (vii) mitochondrial variants were excluded; and (viii) no missing data were allowed. After the quality filters were applied, bcftools and gatk pipelines resulted in the generation of 1,204,984 and 1,157,228 SNPs, respectively. Finally, to ensure high quality and reliability of the data, only the variants consistently called by both pipelines were retained (1,128,043 SNPs).

Mitochondrial variants were extracted from the vcf data generated by bcftools and GATK pipelines by retaining only mitochondrial genome as described earlier, except that maximum mean sequencing depth was unlimited due to high copy number of mtDNA and MAF filter was not applied. In total, 131 and 129 mtDNA SNPs were detected by bcftools and GATK pipelines, respectively. The final dataset included 129 mtDNA SNPs consistently called by the two pipelines.

2.7 | Haplotype network analysis

Haplotype network analysis was carried out for whole mtDNA genomes that did not contain any ambiguous nucleotides. This analysis involved examining 41 mtDNA genomes (39 individual mtDNA genomes from WGS complemented with one RNA-derived sequence from Noreikiene et al., 2020 and a reference genome). Construction of haplotype network was performed using PopART ver. 1.7 software (Leigh & Bryant, 2015). The median-joining network method (Bandelt et al., 1999) was employed to construct the haplotype network, which allows for the representation of genetic relationships between haplotypes.

2.8 | Analysis of phylogenetic relationships

Because the vast majority of the RNA-seq-derived mtDNA consensus sequences varied in length, we trimmed the whole mitochondrial

genome data (16,537 bp) to 16,152 bp by excluding a part of the D-loop sequence that was not present in all studied sequences. Overall, phylogenetic analysis was performed using 114 near-complete mtDNA genomes. After collating the data, we aligned the sequences using MUSCLE ver. 3.8 (Edgar, 2004) and ran jModelTest2 (Darriba et al., 2012; Guindon & Gascuel, 2003) to determine the optimal evolutionary substitution model for the multiple sequence alignment data obtained. According to AIC, the best-fit configuration was Transitional Incremental Model 2 (TIM2), whereas the Bayesian Information Criterion (BIC) estimate suggested Hasegawa-Kishino-Yano (HKY) as the best supported model of DNA sequence evolution. Therefore, to resolve the ambiguity, we ran a similar test in ModelFinder (included in IQ-TREE ver. 2.1.3, Kalyaanamoorthy et al., 2017), which involves calculating log-likelihoods for the starting parsimony tree across a range of DNA models of base substitution rates. ModelFinder incorporates corrected AIC (AICc), adjusted for correction in the bias due to small sample size, in addition to AIC and BIC.

Finally, we identified the HKY + F + I substitution model as the best fit for a given dataset. Here, the F parameter was applied by virtue of observed unequal base frequencies, and the I option was added due to the presence of invariable sites. Finally, we inferred the maximum likelihood (ML) tree with ultrafast bootstrap approximation in IQ-TREE ver. 2.1.3 (Minh et al., 2020) setting the minimum number of replicates to 1000 as recommended (Hoang et al., 2018). The resulting consensus trees were subsequently visualized in FigTree (Rambaut, 2009).

2.9 | Analysis of molecular variance

AMOVA was applied to evaluate the proportion of within- and among-group genetic variation between highly diverged mitochondrial lineages and countries. Mitochondrial sequence alignment data in FASTA format were transformed to Arlequin ver. 3.5.2.2 (Excoffier & Lischer, 2010) input files for later processing, whereas for nuclear SNP dataset we applied function poppr.amova from the R package poppr (Kamvar et al., 2014). Standard AMOVA tests (distance matrix computation, pair-wise differences) for haploid and diploid data were conducted, and an overall estimation of genetic differentiation across all populations (global F_{ST}) was performed. The populations were designated by the number of highly diverged evolutionary lineages ($n = 5$) and by country of origin ($n = 6$).

2.10 | Divergence time estimation based on mtDNA

By assessing the time since divergence between highly diverged lineages we aimed to investigate the possible association between the divergence time and geological events. Stepien and Haponski (2015) estimated the divergence time for genus *Perca* using two mtDNA regions: cytochrome b (cyt b, 1140 bp) and cytochrome c oxidase I (COI, 1550 bp), as well as nuclear recombination-activating gene

intron 1 (RAG1, 6879 bp). We partially integrated the framework from the given study to construct the mtDNA input dataset. Thus, we imported whole mtDNA genomes of the following species: saw-edged perch (*Niphon spinosus*, NC068731.1), as the closest living relative, was added as the out-group to root the tree and infer monophyly for family *Percidae* (Ghezelayagh et al., 2022); sauger (*Sander canadensis*, NC021444.1) and walleye (*Sander vitreus*, NC028285.1) sequences, as well as smallmouth bass (*Micropterus dolomieu*, NC011361.1) and largemouth bass (*Micropterus salmoides*, NC008106.1), were included to apply fossil priors for *Sander* and *Micropterus* genera (Haponski & Stepien, 2013). Balkhash perch (NC027745.1) and yellow perch (JX629448.1) were included to supplement the dataset with extra taxa, as we applied the estimate from the previous study (45 Mya for split between *N. spinosus* and *Percidae* family, Ghezelayagh et al., 2022) as the molecular prior for calibration within *Perca* genus. Twelve newly generated mtDNA genomes of *P. fluviatilis* and a published mtDNA genome from a Hungarian specimen (KM410088.1) were selected for estimation of divergence time based on the ML phylogenetic consensus tree (Table S1).

The presence of multiple species in a dataset resulted in a different length of genomic sequences; therefore, we (i) truncated each sequence by position 15,680, (ii) aligned partitions in Muscle (Edgar, 2004), and (iii) trimmed the sequences again to minimize the number of gaps. Final alignment, comprising 20 sequences of 15,704 bp length for each partition, was later processed in a statistical test of best model selection in terms of nucleotide substitution. TIM2 + F + I + G4 model was suggested as the best fit in ModelFinder (Kalyaanamoorthy et al., 2017) according to AIC, AICc, and BIC criteria. Extra verification in jModelTest2 (Darriba et al., 2012; Guindon & Gascuel, 2003) has shown that the given model is within the range of a credible set of selections. The input file in xml format was generated in BEAUTi ver. 2.6.7 and was processed later in BEAST ver. 2.6.7 (Bouckaert et al., 2019). We used a relaxed clock log-normal model as we assumed that the evolutionary rates could vary due to the presence of distinct species and taking into account their relatively old age. TIM2 model is not implemented in BEAUTi ver. 2.6.7; thus, we followed Lecocq et al. (2011) and replaced TIM2 with GTR, retaining the discrete Gamma site model (G4) for the whole dataset and applying the given configuration to each of the priors, following suggestions from Stepien and Haponski (2015) based on the output from ModelFinder.

We used the configuration of three priors (one molecular and two fossil) in our analysis. Priors within the *Percidae* family were calibrated according to the fossil records of *Micropterus* (ca. 12 Mya) and *Sander* (ca. 15 Mya, Stepien & Haponski, 2015), whereas setting minimum age constrain for the entire *Percidae* family was grounded on the molecular estimate (Ghezelayagh et al., 2022). We also reiterated the analysis by (i) increasing the chain length from 10 to 50 million generations, (ii) testing with sampled parameters every 100, 1000, and 10,000 generations, and (iii) running the programme with a specified setup configuration; however, the structure of relationships remained the same. The results presented in the study represent the simulation performed for 30,000,000 chain length. Most effective

sample size (ESS) values were >200 with no large-scale fluctuations observed, indicating that the MCMC chain has converged on a stable optimum. We discarded the first 20% of logged trees as burn-in, as the ESS values of all tests within 30,000,000 chain converged to optimum by reaching 6,000,000 steps.

2.11 | Admixture

We employed ADMIXTURE 1.3 (Alexander et al., 2015) to estimate individual ancestry proportions, performing 500 bootstraps (-B500) across a range of ancestral lineages (K) from 1 to 20 and a 10-fold cross-validation ($-cv = 10$). This analysis was carried out using 31,896 nuclear SNPs, meeting stringent criteria: a 100% call rate, and adherence to Hardy–Weinberg equilibrium (p-value <5e-06). Additionally, to minimize potential confounding effects of linkage disequilibrium (LD) on admixture patterns (Falush et al., 2003; Kaeuffer et al., 2007), we restricted LD values to a maximum of 0.2. LD pruning was performed in PLINK using—indep-pairwise 1000 100 0.2. Overall, 17,048 SNPs and 1,079,099 SNPs were removed by HWE deviation and LD filtering, respectively.

In addition, we used nuclear SNP data to assess observed heterozygosity (Ho) across the genomes of 39 individuals, as well as to identify putative signal of admixture in Estonian samples from mtDNA lineages 2, 3, and 4. To estimate the genetic diversity, we used vcftools applying–het flag. We used scikit-allele ver. 1.3.8 (Miles et al., 2024) to generate subsets of genotype arrays for population triplets (a;b;c), and estimate f3 scores. The f3-score is defined as (a-b)· (a-c), in which a, b, and c represent vectors with the allele frequencies in the putatively admixed population A and the two putative donor populations B and C respectively. Thus, negative f3-score is indicative of admixture. We calculated f3 for three different triplets: (i) admixture of Estonian mtDNA lineage 2 (ELOO1, EPAR1, EPAI1, ESAA1, EANT1), testing the donors of Swedish (lineage 2) and Estonian (non-lineage 2) populations; (ii) admixture of Estonian mtDNA lineage 3 (EHIN1, EKUU1, EMAT1, EMEE1, EUDR1, EVII1, EVIR1), testing the scenario where the donors of gene flow are samples from Lithuania (lineage 3) and Estonia (outside of lineage 3); and (iii) admixture of Estonian mtDNA lineage 4 (EHEI1, EPII1, EUIA1, EVER1), testing the donors of Finnish (lineage 4) and Estonian (non-lineage 4) samples.

2.12 | Reconstruction of demographic history

Pair-wise Sequentially Markovian Coalescent (PSMC; Li & Durbin, 2011) was used to infer the fluctuations in effective population size N_e over time using nuclear SNP data by estimating coalescence rates between haplotypes through time. The rates of coalescent events at the specific time, in turn, correlate with the changes in the local density of heterozygous sites across the genome and are inversely proportional to N_e (Li & Durbin, 2011; Nadachowska-Brzyska et al., 2016). However, the inferred N_e

dynamics is also strongly affected by population substructure. Whole-genome diploid consensus sequences were generated using samtools ver. 1.10 and bfctools ver. 1.17 using default settings (https://github.com/lh3/psmc), except that the minimum and maximum read depths were set to 10 and twice of the average coverage for each sample (Table S2), respectively. The input files for PSMC modeling were generated with the fq2psmcfa tool (−q 20) and processed in psmc applying the following parameters: −N25 −t15 −r5 −p "4 + 25*2 + 4 + 6," where N is the number of iterations, t stands for the number of time intervals per iteration, r denotes the ratio of heterozygosity to recombination rate, and p defines the N_e at each of the four main time intervals (Li & Durbin, 2011; Patil et al., 2021). A generation time of 3 years (Freyhof & Kottelat, 2008) and a mutation rate of 2.5e-08 (Liu et al., 2016) were applied for time calibration.

2.13 | Co-phylogeny of mitochondrial and nuclear DNA

To evaluate the congruence in genetic structuring between mtDNA and nDNA trees, we estimated evolutionary distances between 39 individuals of perch, separately for nDNA and mtDNA SNP datasets and applied a co-phylogenetic approach to compare them. First, we calculated p-distance matrices for both datasets, by running VCF2Dis ver. 1.50 software (https://github.com/BGI-shenzhen/VCF2Dis). Later, mtDNA and nDNA trees in Newick format were inferred from p-distance matrices using FastME ver. 2.0 (Lefort et al., 2015). The generated trees were visualized using phytools package (Revell, 2012) in R.

3 | RESULTS

3.1 | Haplotype network

The constructed haplotype network revealed the largest genetic distance between the single Hungarian specimen and the rest of the samples (97 mutations apart from the closest haplotype, out of 202 variable sites in total). The rest of the analysed mtDNA genomes were grouped into four distinct lineages (Figure 1b). Six specimens collected in Finland, as well as two from Lithuania and two from Russia, possessed identical haplotypes, respectively. However, only samples from Siberia constituted a clearly separated group in relation to geographical location, whereas multiple lineages were present in several countries around the Baltic Sea (Figure 1a).

3.2 | Phylogenetic relationships based on near-complete mtDNA genomes

Based on 114 near-complete mitochondrial genomes, phylogenetic ML tree revealed the presence of five highly diverged mitochondrial lineages, three of which were found in the Baltic Sea basin (Figure 1c). Similar to haplotype network analysis, the most divergent from all

other haplotypes was a single specimen from Lake Balaton, Hungary (KM410088.1). The consensus tree without Hungarian sample revealed divergence among four remaining lineages: a clearly separated group consisting of perch from Russia and China (denoted as lineage 1), as well as distinct clades named lineage 2 (Estonia, France, Ukraine, Sweden), lineage 3 (Estonia, Sweden, Lithuania, Belarus), and lineage 4 (Estonia, Finland, Ukraine, Sweden, Lithuania; Figure 1d). Within lineage 2, two haplotypes from Estonia were characterized by early branching with bootstrap values of 65%. A similar branching pattern was observed in lineage 3 for six Estonian haplotypes and a single Swedish haplotype, which exhibited multifurcating pattern.

3.3 | AMOVA

We examined molecular variance based on complete mitochondrial and nuclear genomes by quantifying the variance associated with the evolutionary lineages ($n = 5$) and the countries of origin ($n = 6$). For the mitochondrial dataset, the variance associated with differences between evolutionary lineages was considerably higher (90.78%) compared to the variance explained by the differences between countries of origin (65.61%, Table 2), reflecting the occurrence of multiple lineages in several countries, particularly around the Baltic Sea (Estonia, Lithuania, Sweden). As for nuclear dataset, the variance explained by country or lineage was much lower with more variation explained by country of origin (9.9%) than mtDNA lineage (6.74%).

3.4 | Divergence time estimation

The maximum credibility coalescent tree (Figure 2) suggested a split between P. fluviatilis, and the other Perca species occurred approximately 15.82 Mya (95% highest posterior density [HPD]: 10.24–22.05 Mya). The oldest intraspecific split was found between Hungarian and the remaining samples of P. fluviatilis on Early Pliocene 1.42 Mya (95% HPD: 0.73–2.27 Mya). The pattern of bifurcation was consistent with the branching of phylogenetic tree: the clade corresponding to lineage 1 is a sister group to lineage 2, whereas lineages 3 and 4 are grouped together. The estimated divergence between four remaining evolutionary lineages (lineages 1, 2 vs. 3, 4) occurred 0.64 Mya (95% HPD: 0.36–1.02 Mya). The timing of the split between lineages 1 and 2 was dated to 0.35 Mya (95% HPD: 0.16–0.58 Mya), whereas the estimated split between lineages 3 and 4 occurred 0.24 Mya (95% HPD: 0.01–0.41 Mya).

3.5 | Admixture

We performed the ADMIXTURE analysis based on nuclear SNPs to determine the ancestral lineage composition among 39 individuals (Figure 3). The cross-validation error indicated an optimal clustering at $K = 2$, revealing the largest genetic divergence between European and Siberian (Russia) samples, and potential contribution of Siberian ancestry to the Baltic region. However, at $K = 3$, no shared ancestry
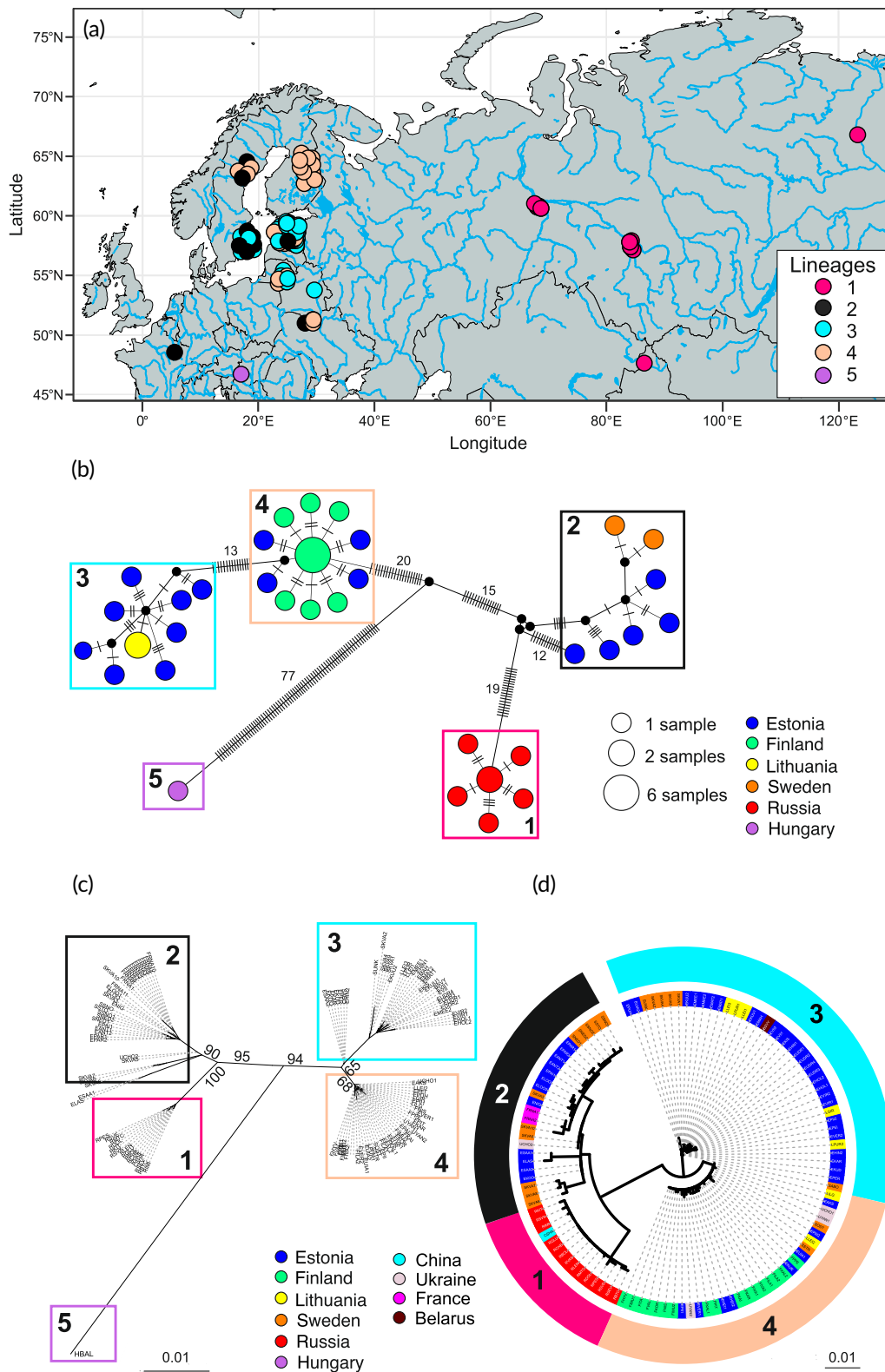
FIGURE 1 (a) Sampling locations of studied *Perca fluviatilis* populations. Different colors of sampling locations in (a), boxes in (b, c), and the outmost circle sectors in (d) correspond to the highly divergent evolutionary lineages 1–5. The size of the haplotype bubble reflects the number of sequences (individuals), its color indicates the country of origin, the numbers above the main branches correspond to bootstrap values, and the hatch marks on the branches indicate the number of nucleotide differences. ML representation of phylogenetic relationships between populations as unrooted (c) and circular (d) tree. Different colors in the inner circle in (d) indicate the country of origin of populations. A sample from Hungary (HBAL, lineage 5) is not present in (d) due to its considerable branch length.

TABLE 2 Results of AMOVA and global F_{ST} values based on 39 whole nuclear (nDNA) and mtDNA genomes grouped by country of origin ($n = 6$) or by evolutionary lineage ($n = 5$).

Source of variation	Sum of squares		Variance components		Variation (%)		F_{ST}	
	mtDNA	nDNA	mtDNA	nDNA	mtDNA	nDNA	mtDNA	nDNA
Among countries	419.2	133543.7	13.3	1358.7	65.6	9.9		
Within countries	243.0	889625.6	6.9	12357.5	34.4	90.1		
Total	662.2	1023169.3	20.2	13716.2			0.656*	0.099*
Among lineages	590.8	6.7	19.5	915.0	90.8	6.7		
Within lineages	71.3	925826.2	2.0	12662.5	9.2	93.3		
Total	662.2	1023169.3	21.5	13577.5			0.908*	0.067*

Note: Significance of variance explained set as *p*-value <0.05 (*).
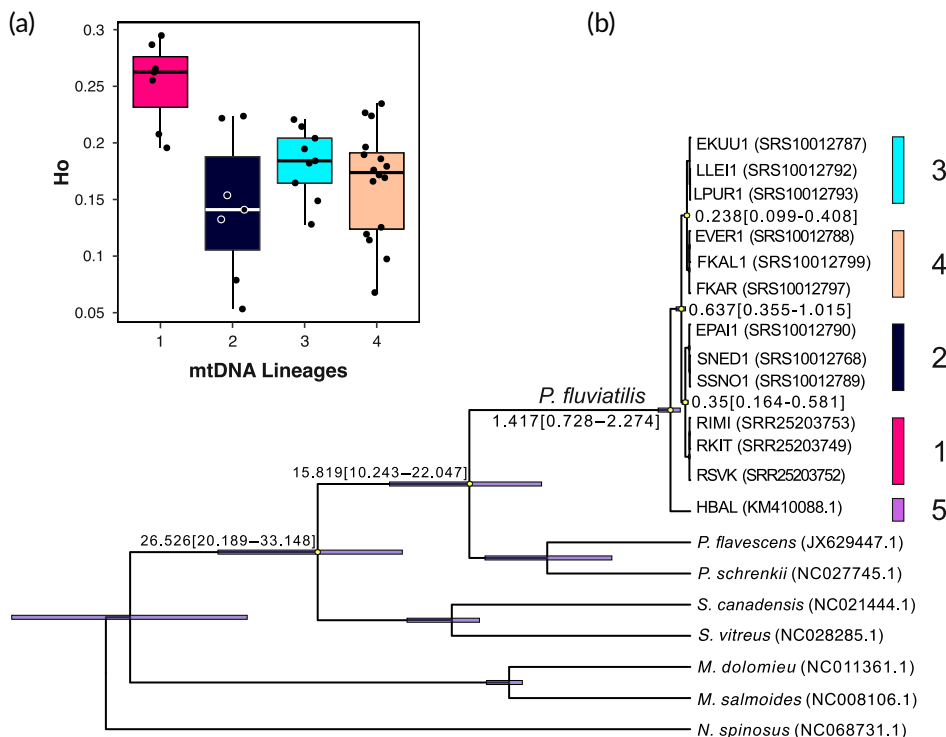

FIGURE 2 (a) Boxplots of observed heterozygosity (Ho) across nuclear genome in 39 individuals, grouped by mtDNA lineage. (b) Estimated divergence times (in millions of years ago, Mya) between species of the *Percidae* family and five highly diverged evolutionary lineages of *Perca fluviatilis*, based on nearly whole mtDNA genome sequences. The *P. fluviatilis* lineages 1 to 4 are represented by three individuals from each lineage, and lineage 5 is represented by a single individual HBAL (individual designations according to Table S1). The mean divergence time estimates are followed by a 95% HPD (highest posterior density) interval in brackets.

was observed between Siberian and European samples. Instead, $K = 3$ separated Finnish samples from the rest of European samples and also indicated potential contribution of Finnish ancestry to Estonian, Lithuanian, and Swedish samples. At higher K values, however, the Siberian and Finnish samples consistently emerged as the most distinct clusters, and we found no evidence of Finnish or Siberian contribution to the rest of the samples (Estonia, Lithuania, Sweden). Instead, individual samples started to diverge from the previous clusters most likely driven by genetic drift (Figure S1).

The results of the f3 test on nuclear SNPs provided no evidence for admixture in Estonian individuals from lineages 2, 3, and 4. In all cases, the average f3 score was positive (0.019, 0.024, and 0.024), and the z-scores were positive, indicating no support for admixture (Table S4). Finally, the Ho estimates based on nuclear SNP data (Figure 3, bottom graph; Table S3) indicated that nuclear diversity estimates in lineage 1 were higher compared to other lineages.

3.6 | Demographic history

The inferred demographic history reveals fluctuations in the population size dynamics for all studied samples over the past 300,000 years (Figure 4). However, there is a clear difference in the dynamics between perch populations from Europe and Western Siberia: during the early and middle stages of the Last Glacial Period (LGP, ca. 30,000–100,000 years before present, YBP), the population size of Siberian perch increased, whereas all European perch samples showed a consistent decline in population size. The growth in the population size of Siberian genomes overlaps with the Eemian interglacial (corresponding to Marine Isotope Stage 5e, MIS5e, ca. 120,000 years ago) and lasted throughout lower and middle Weichselian, until ca. 25,000 years ago. Interestingly, the population size fluctuations for all European samples were very similar to each other irrespective of their mtDNA lineage, indicating that the divergence of mitochondrial lineages to a large

FIGURE 3 Analysis of *Perca fluviatilis* ancestry based on 39 nuclear genomes assuming different number of clusters (K = 2–4) using ADMIXTURE, followed by observed heterozygosity (Ho), bottom graph. Sample names (bottom) are colored based on the designated mtDNA lineage.
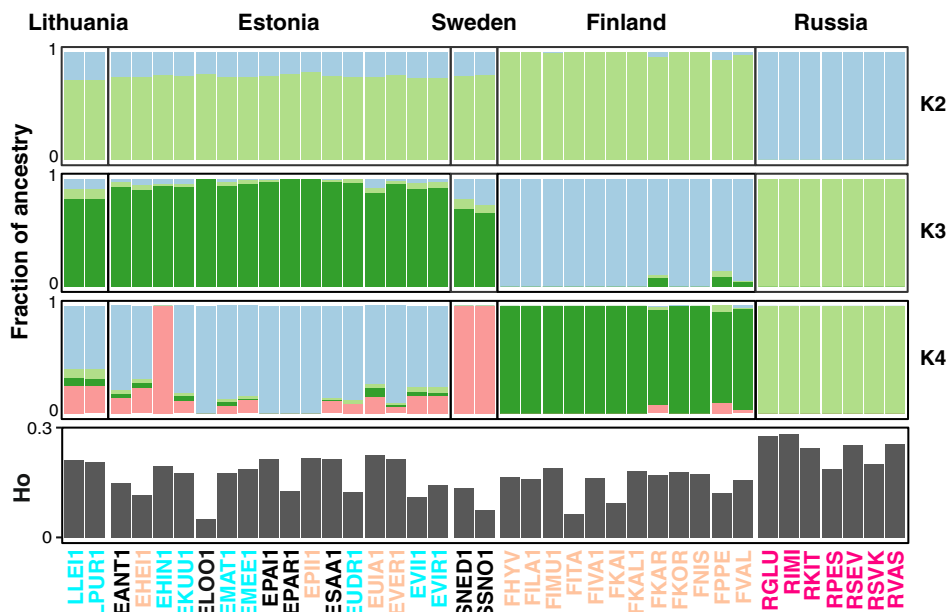

FIGURE 4 Population size dynamics based on 39 *Perca fluviatilis* nuclear genomes aligned to the oscillations of global average surface temperature (Snyder, 2016) during Early (E), Middle, and Late stages of the Pleistocene and Holocene epochs. Periods of glaciation are represented as the timeframes highlighted in blue. LGM, last glacial maximum; LGI, late glacial interstadial.
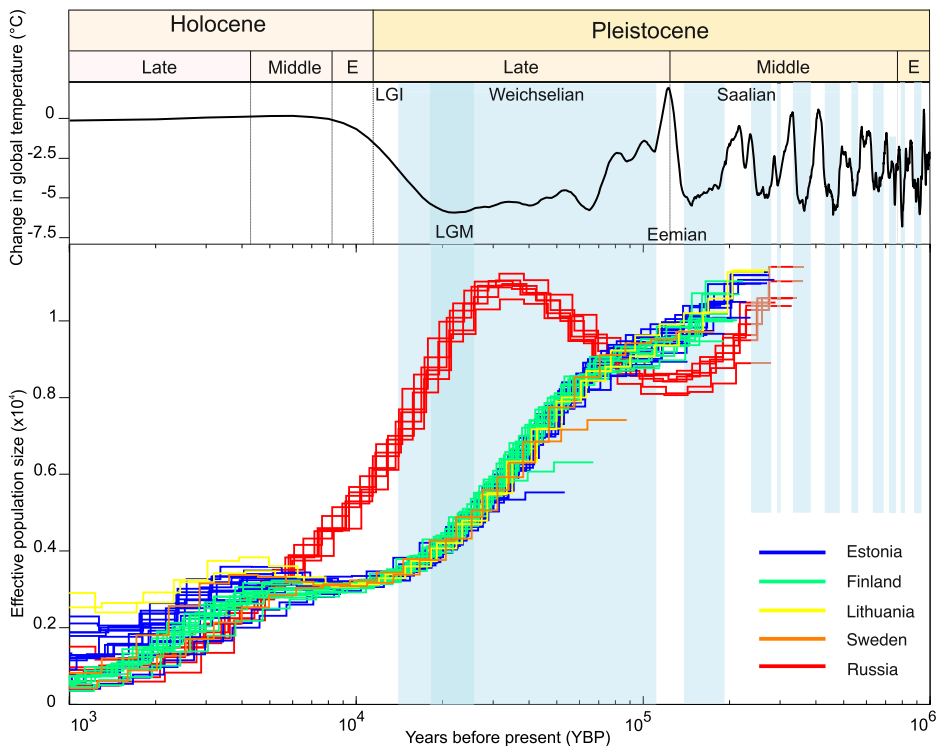
extent predates the demographic history accumulated in nuclear genome over the past 200,000 years.

3.7 | Co-phylogenies of mitochondrial and nuclear genomes

A comparison of mitochondrial and nuclear data-based evolutionary trees revealed both similarities and incongruences between evolutionary histories inferred from two different types of markers (Figure 5). For example, the Siberian samples clustered together based both on mtDNA and nuclear genome data, reflecting the consistency between maternally and bi-parentally inherited markers. On the contrary, although nDNA phylogram indicated that all samples from Finland form a monophyletic group, the mtDNA genome grouped some Finnish specimens together with Estonian samples. Similarly, although Swedish specimens clustered separately from other samples based on nuclear data, mitochondrial sequence data grouped them together with several Estonian samples within lineage 2.
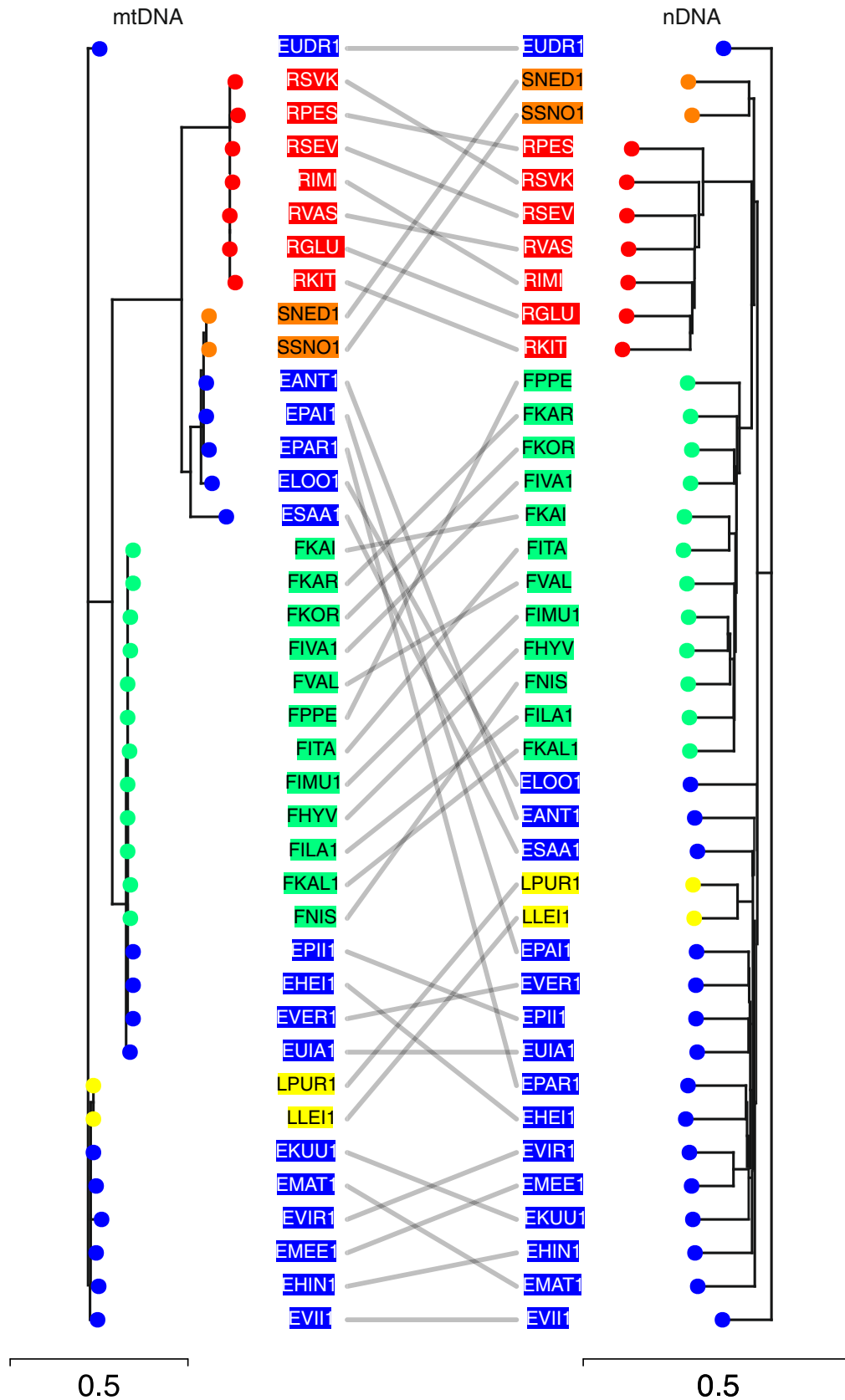
FIGURE 5 The tangle-gram, reflecting co-phylogeny of *Perca fluviatilis* evolutionary relationships measured based on mitochondrial (left) and nuclear (right) data. The scale depicts edge length, calculated from the p-distance matrix. The individual samples are designated according to Table S1; the colors of tips (bubbles) and labels correspond to the countries of origin. mtDNA, mitochondrial DNA; nDNA, nuclear DNA.

4 | DISCUSSION

The analysis of whole mitochondrial and nuclear genomes has enabled us, for the first time, to reconstruct the demographic history and estimate the divergence time for the major mtDNA evolutionary lineages of Eurasian perch. This was achieved by integrating newly obtained genomic information with the publicly available genomic and transcriptomic datasets, thereby expanding the geographic scope of our investigation. Mitochondrial haplotype network analysis and phylogenetic analysis of near-complete mitochondrial genomes revealed five highly diverged mtDNA lineages, which have diverged approximately 0.24–1.42 Mya. Notably, three of these lineages show overlapping distribution in the Baltic Sea region. The presence of multiple evolutionary lineages and their geographic distribution in *P. fluviatilis* aligns with known patterns of postglacial recolonization and population expansion observed in many other species, including teleost fish (Culling et al., 2006; Sommer & Zachos, 2009). Furthermore, the examination of nuclear genomes through ADMIXTURE analysis showed a consistent separation between European and Siberian perch. Similarly, PSMC analysis revealed distinct demographic trajectories in both groups, most likely reflecting contrasting evolutionary histories over a period ranging from 30,000 to 100,000 YBP. Together this suggests that in Europe *P. fluviatilis* likely underwent postglacial recolonization from several refugial zones, and experienced different population size fluctuations and/or isolation dynamics than Siberian perch.

Our findings partially reaffirm inferences from earlier works based on D-loop analysis (Nesbø et al., 1999), which suggested that present-day perch populations in Central and Northern Europe were colonized from three main refugia located in Southeastern, Northeastern, and Western Europe. Although the geographic distribution of lineages observed in our study does not fully match the areas observed in Nesbø et al. (1999) and Toomey et al. (2020), we did find a similar admixture pattern for perch in the Baltic region. Therefore, analyses of near-complete mitochondrial genomes performed in the current study support the hypothesis that after the last glaciation and ice retreat, Northern and Central Europe were colonized by perch originating from three refugia. Moreover, similar to earlier findings based on D-loop analysis, haplotype network and phylogenetic analyses presented here revealed that perch in Southeastern Europe belong to the most ancient mtDNA lineage (Nesbø et al., 1999; Toomey et al., 2020). Despite the limited number of samples and geographical coverage, our results also suggest that lineage 2 most likely corresponds to the putative East European refugium, previously designated as R3 by Nesbø et al. (1999). Similarly, the patterns of haplotype network and geographical distribution of lineages indicate the potential correspondence of our lineage 3 to the East European refugium R2 and our lineage 4 to the North European refugium R4 of Nesbø et al. (1999). However, we also detected a clear separation between Siberian and European haplotypes, which contradicts earlier reports that suggested close relatedness between the two lineages (Nesbø et al., 1999). In addition, Siberian and European perch exhibited drastic differences in N_e dynamics based on PSMC analysis of nuclear

data. This suggests the presence of the Siberian lineage, which forms a monophyletic group based on both mitochondrial and nuclear genome data. However, expanding the geographical coverage is clearly needed to provide a less ambiguous and more comprehensive understanding of the phylogeographic history, glacial refugia, and the dynamics of the recolonization of Eurasian perch.

This study marks the first instance of approximate divergence time estimation between evolutionary lineages of *P. fluviatilis* using nearly whole mitochondrial genome information. Overall, findings are broadly compatible with the inter- and intraspecific estimates of divergence time by Stepien and Haponski (2015), who suggested that the emergence of contemporary mtDNA lineages of perch took place around 2.9 Mya. We have included the Hungarian specimen from Lake Balaton to analyses, which indicated that the first split of present-day mtDNA haplotypes occurred more recently, ca. 1.42 Mya (95% HPD: 0.73–2.27). The upper HPD boundary approximately coincides with the extinction of Lake Pannon in Early Pliocene *ca* 2.4 Mya, formerly located in the Pannonian basin of Southeastern Europe (Kázmér, 1990). Notably, the long lifetime of Lake Pannon also contributed to the remarkable diversity of endemic molluscan fauna (Geary et al., 2000). The subsequent mitochondrial split occurred approximately 0.64 Mya during the Middle Pleistocene (MIS16), which was characterized by drastic cooling and intensification of glaciation in the Northern Hemisphere (Hughes & Gibbard, 2018). The subsequent glacial cycles likely gave rise to the current-day mitochondrial lineages, with estimated split times of 0.35 and 0.24 Mya. Thus, genomic data collected as part of this study suggests that the present perch haplotypes are older than suggested by Nesbø et al. (1999), and the observed lineages date back more than two glacial cycles for the most recent and more than five for the most ancient splits. Compared to other fish species in Eurasia, the origin of contemporary perch mitochondrial lineages is similar to the proposed emergence of two major clades within the European whitefish (*Coregonus lavaretus*), estimated to occur approximately 0.4–0.6 Mya (Jacobsen et al., 2012).

The analysis based on nuclear genomes provided additional, more recent insights into the demographic history of *P. fluviatilis* over the past 200,000 years. Notably, the PSMC analysis revealed distinct trajectories of N_e dynamics for Siberian and European perch. In Siberian perch, we observed a decline during the Saalian/Penultimate Glacial Period (MIS 6, 150,000–300,000 YBP), followed by a bottleneck during the Eemian interglacial (MIS 5e, 150,000 YBP), and subsequent fluctuations in N_e during the Last Glacial Maximum (LGM, equals MIS 2). These fluctuations are likely linked to the varying climatic conditions during the Middle and Late Pleistocene. Studies of West-Siberian lakes have shown that abrupt decreases in temperature after warm periods caused significant reductions in zooplankton and algae abundance (Astakhov, 2006; Bezrukova et al., 2010; Mangerud et al., 2001; Nazarov et al., 2022). In contrast, the Penultimate interglacial in Siberia led to the formation of numerous lakes after ice core melting (Mangerud et al., 2004), possibly serving as sanctuaries for perch. The N_e changes in European populations, on the contrary, showed a continuous decline from 200,000 YBP through the Late Glacial Interstadial period (ca. 13,000–14,000 YBP), followed by a more

subtle decline during the past 5000 years. These demographic patterns bear resemblance to those observed in other fish species like big-eye mandarin fish (*Siniperca knerii*) and Wels catfish (*Silurus glanis*) (Lu et al., 2020; Ozerov et al., 2020). However, it is important to approach the inferences of N_e dynamics from PSMC analysis with caution, as population structuring can significantly influence inferred coalescence rates used for N_e estimation (Mazet et al., 2016). For example, a lack of gene flow may lead to the overestimation of N_e due to limited allele coalescence in different isolated populations (Li & Durbin, 2011; Nei & Takahata, 1993). Therefore, increases in N_e observed in PSMC plots may not always indicate larger effective population sizes but rather periods of long-term population structuring. Thus, PSMC outputs can be viewed as N_e trajectories if the populations are panmictic or, alternatively, reflect changes in connectivity when the population relationships are more accurately characterized using a structured model. The observed consistent differences between European and Siberian nuclear and mitochondrial genomes, however, suggest distinct evolutionary histories between these two groups before the last Glacial maximum irrespective of changes in genetic structuring over time.

Deep divergence between European and Siberian samples was also supported by ADMIXTURE analysis, revealing a potential contribution of Siberian ancestry to the Baltic region. However, no consistent evidence for admixture between European and Siberian samples was observed at higher *K* values. Similarly, the lack of Siberian mtDNA haplotypes in the Baltic region supports deep evolutionary divergence between European and Siberian samples. Interestingly, Siberian samples also showed higher nuclear diversity than other European samples. Furthermore, at higher *K* values, ADMIXTURE successfully clustered samples based on the country of origin but failed to reveal consistent signals of admixture between different mitochondrial lineages. Similarly, f3 statistics calculated for specific population triplets did not provide support for consistent admixture signals. Therefore, to accurately characterize putative introgression patterns in Eurasian perch, analyses of samples collected from the whole distribution range, including southern putative glacial refugium areas, are needed.

By comparing the mtDNA and nuclear genomes obtained from the same specimens, we identified major phylogenetic incongruences and differences in variance (AMOVA) between the two types of data. As such, our findings corroborate an increasing number of studies that show incongruences between gene trees generated using mitochondrial and nuclear data (Toews & Brelsford, 2012). Yet, differentiating between the possible causes of incongruence is not straightforward, and mito-nuclear discordance can be driven by various factors, including incomplete lineage sorting (McGuire et al., 2007; Pollard et al., 2006), introgression or hybridization between lineages (Bernal et al., 2017; Ozerov et al., 2016), and retention of ancestral polymorphisms (Moran & Kornfield, 1993), which are shaped by glacial oscillations and subsequent colonization and hybridization events. Intriguingly, the genetic clustering of perch based on nuclear data closely reflected the geographical origin of the fish, whereas the mtDNA lineages showed a more mixed pattern. This incongruence between mitochondrial and nuclear trees in perch likely results from

two main factors. First, it could be attributed to the effect of repeated glacial cycles, which led to the establishment of multiple refugia and the emergence of distinct mitochondrial lineages. Second, subsequent colonization and postglacial dispersal events likely facilitated hybridization between lineages and the shuffling of nuclear genomes. As nuclear genomes undergo diploid inheritance and recombination, they tend to reflect more recent processes of colonization and admixture. In contrast, the divergent mitochondrial lineages reveal more ancient evolutionary splits.

5 | CONCLUSIONS

This study provided new whole-genome insights into the phylogenetic and demographic history of *P. fluviatilis*. By integrating novel mitochondrial and nuclear genome data with the publicly available transcriptomic resources, we shed light on the demographic history of contemporary mitochondrial lineages, potential glacial refugia, and dispersal postglacial routes. Notably, five distinct mtDNA lineages, including three with extensive overlap in the Baltic Sea area, were identified, corroborating prior short mtDNA fragment-based inferences. The estimated divergence times between mtDNA lineages were older (0.24, 0.35, 0.64, and 1.42 Mya) than proposed by Nesbø et al. (1999) suggesting their emergence at the onset of Saalian glaciation (ca. 150,000–300,000 YBP) and earlier. Thus, dramatic climate changes characterized by long glacial periods causing large ice sheets to expand and short warmer interglacial periods have likely shaped the current-day mtDNA distribution in Eurasian perch. Furthermore, the analysis of nuclear genomes provided more detailed insights into the demographic history over the past 300,000 years, reflecting the distinct demographic fluctuations of perch in Europe and Siberia. Finally, the incongruences of mtDNA and nDNA evolutionary trees likely reflect the differences in inheritance and recombination between markers, as nuclear genomes are predominantly shaped by recent colonization and admixture processes, whereas divergent mitochondrial lineages manifest more ancient evolutionary splits being spread and redistributed during the recolonization process. Overall, this study underscores the power of genome-wide approaches to enhance our understanding of the historical processes shaping the genetic diversity of widespread freshwater fish species.

phylogenetic analyses. Tissue samples from Western Siberia were collected in collaboration with Yugra State University (Elena Lapshina, Mikhail Zolotov, Evgeni Zarov, and Alexei Dmitrichenko) and Tomsk State University (Sergei Kirpotin, Viktor Popkov, and Viktor Drozdov). Tissue samples from the River Lena were collected by Nadezhda Popova (Arctic State Agrotechnological University), and Ruslan Popov (Yakut Research Institute of Agriculture). Tissue samples from Lithuania were collected in collaboration with Linas Ložys (Nature Research Centre, Lithuania).

ORCID

Vitalii Lichman https://orcid.org/0009-0007-5955-0479

REFERENCES

Alexander, D. H., Shringarpure, S. S., Novembre, J., & Lange, K. (2015). *Admixture 1.3 software manual*. UCLA Human Genetics Software Distribution.

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. www.bioinformatics.babraham.ac.uk/projects/fastqc/

Arbogast, B. S., & Kenagy, G. J. (2001). Comparative phylogeography as an integrative approach to historical biogeography. *Journal of Biogeography*, 28(7), 819–825.

Astakhov, V. I. (2006). Evidence of late Pleistocene ice dammed lakes in West Siberia. *Boreas*, 35(4), 607–621.

Bandelt, H., Forster, P., & Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, 16(1), 37–48.

Beichman, A. C., Huerta-Sanchez, E., & Lohmueller, K. E. (2018). Using genomic data to infer historic population dynamics of nonmodel organisms. *Annual Review of Ecology, Evolution, and Systematics*, 49, 433–456.

Bernal, M. A., Gaither, M. R., Simison, W. B., & Rocha, L. A. (2017). Introgression and selection shaped the evolutionary history of sympatric sister species of coral reef fishes (genus: *Haemulon*). *Molecular Ecology*, 26(2), 639–652.

Bezrukova, E. V., Tarasov, P. E., Solovieva, N., Krivonogov, S. K., & Riedel, F. (2010). Last glacial–Interglacial vegetation and environmental dynamics in southern Siberia: Chronology, forcing and feedbacks. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 296(1–2), 185–198.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120.

Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., de Maio, N., Matschiner, M., Mendes, F. K., Müller, N. F., Ogilvie, H. A., du Plessis, L., Popinga, A., Rambaut, A., Rasmussen, D., Siveroni, I., ... Drummond, A. J. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, 15(4), e1006650.

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, 34(17), i884–i890.

Christensen, E. A., Grosell, M., & Steffensen, J. F. (2019). Maximum salinity tolerance and osmoregulatory capabilities of European perch *Perca fluviatilis* populations originating from different salinity habitats. *Conservation Physiology*, 7(1), coz004.

Collette, B. B., & Bănărescu, P. (1977). Systematics and zoogeography of the fishes of the family *Percidae*. *Journal of the Fisheries Board of Canada*, 34(10), 1450–1463.

Coyne, J. A. (1994). Ernst Mayr and the origin of species. *Evolution*, 48(1), 19–30.

Culling, M. A., Janko, K., Boron, A., Vasil'ev, V. P., Cote, I. M., & Hewitt, G. M. (2006). European colonization by the spined loach (*Cobitistaenia*) from Ponto-Caspian refugia based on mitochondrial DNA variation. *Molecular Ecology*, 15(1), 173–190.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156–2158.

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: More models, new heuristics and parallel computing. *Nature Methods*, 9(8), 772.

Devlin, S. P., Saarenheimo, J., Syväranta, J., & Jones, R. I. (2015). Top consumer abundance influences lake methane efflux. *Nature Communications*, 6(1), 8787.

Diehl, S. (1992). Fish predation and benthic community structure: The role of omnivory and habitat complexity. *Ecology*, 73(5), 1646–1661.

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32, 1792–1797.

Erpenbeck, D., Hooper, J. N. A., & Wörheide, G. (2006). CO1 phylogenies in diploblasts and the 'barcoding of life'—Are we sequencing a suboptimal partition? *Molecular Ecology Notes*, 6(2), 550–553.

Excoffier, L., & Lischer, H. E. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and windows. *Molecular Ecology Resources*, 10(3), 564–567.

Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*, 164(4), 1567–1587.

Foster, J. T., Beckstrom-Sternberg, S. M., Pearson, T., Beckstrom-Sternberg, J. S., Chain, P. S., Roberto, F. F., Hnath, J., Brettin, T., & Keim, P. (2009). Whole-genome-based phylogeny and divergence of the genus *Brucella*. *Journal of Bacteriology*, 191(8), 2864–2870.

Freyhof, J., & Kottelat, M. (2008). *Perca fluviatilis. The IUCN red list of threatened species 2008:e.T16580A6135168*. Accessed July 25, 2023. https://doi.org/10.2305/IUCN.UK.2008.RLTS.T16580A6135168.en

Geary, D. H., Magyar, I., & Müller, P. (2000). Ancient Lake Pannon and its endemic molluscan fauna (Central Europe; mio-pliocene). *Advances in Ecological Research*, 31, 463–482.

Ghezelayagh, A., Harrington, R. C., Burress, E. D., Campbell, M. A., Buckner, J. C., Chakrabarty, P., Glass, J. R., McCraney, W. T., Unmack, P. J., Thacker, C. E., Alfaro, M. E., Friedman, S. T., Ludt, W. B., Cowman, P. F., Friedman, M., Price, S. A., Dornburg, A., Faircloth, B. C., Wainwright, P. C., & Near, T. J. (2022). Prolonged morphological expansion of spiny-rayed fishes following the end-cretaceous. *Nature Ecology & Evolution*, 6(8), 1211–1220.

Gronau, I., Hubisz, M. J., Gulko, B., Danko, C. G., & Siepel, A. (2011). Bayesian inference of ancient human demography from individual genome sequences. *Nature Genetics*, 43(10), 1031–1034.

Guindon, S., & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, 52(5), 696–704.

Haponski, A. E., & Stepien, C. A. (2013). Phylogenetic and biogeographical relationships of the Sander pikeperches (*Percidae: Perciformes*): Patterns across North America and Eurasia. *Biological Journal of the Linnean Society*, 110(1), 156–179.

Hewitt, G. M. (1999). Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society*, 68(1–2), 87–112.

Hewitt, G. M. (2001). Speciation, hybrid zones and phylogeography—Or seeing genes in space and time. *Molecular Ecology*, 10(3), 537–549.

Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, 35(2), 518–522.

Hughes, P. D., & Gibbard, P. L. (2018). Global glacier dynamics during 100 ka Pleistocene glacial cycles. *Quaternary Research*, 90(1), 222–243.

Jacobsen, M. W., Hansen, M. M., Orlando, L., Bekkevold, D., Bernatchez, L., Willerslev, E., & Gilbert, M. T. P. (2012). Mitogenome sequencing reveals shallow evolutionary histories and recent divergence time between morphologically and ecologically distinct European whitefish (*Coregonus* spp.). *Molecular Ecology*, 21(11), 2727–2742.

Kaeuffer, R., Réale, D., Coltman, D. W., & Pontier, D. (2007). Detecting population STRUCTURE using STRUCTURE software: Effect of background linkage disequilibrium. *Heredity*, 99(4), 374–380.

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., Von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, 14(6), 587–589.

Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, 2, e281.

Kázmér, M. (1990). Birth, life and death of the Pannonian Lake. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 79(1–2), 171–188.

Kivisild, T. (2015). Maternal ancestry and population history from whole mitochondrial genomes. *Investigative Genetics*, 6, 1–10.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with bowtie 2. *Nature Methods*, 9(4), 357–359.

Lecocq, T., Lhomme, P., Michez, D., Dellicour, S., Valterova, I., & Rasmont, P. (2011). Molecular and chemical characters to evaluate species status of two cuckoo bumblebees: *Bombus barbutellus* and *Bombus maxillosus* (hymenoptera, Apidae, Bombini). *Systematic Entomology*, 36(3), 453–469.

Lefort, V., Desper, R., & Gascuel, O. (2015). FastME 2.0: A comprehensive, accurate, and fast distance-based phylogeny inference program. *Molecular Biology and Evolution*, 32(10), 2798–2800.

Leigh, J. W., & Bryant, D. (2015). POPART: Full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, 6(9), 1110–1116.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, 27, 2987–2993.

Li, H. (2013). Seqtk: A fast and lightweight tool for processing sequences in the FASTA or FASTQ format. Accessed June 21, 2023. https://github.com/lh3/seqtk

Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, 475(7357), 493–496.

Liu, Z., Liu, S., Yao, J., Bao, L., Zhang, J., & Li, Y. (2016). The channel catfish genome sequence provides insights into the evolution of scale formation in teleosts. *Nature Communications*, 7, 11757.

Lu, L., Zhao, J., & Li, C. (2020). High-quality genome assembly and annotation of the big-eye mandarin fish (*Sinipercaknerii*). *G3: Genes, Genomes, Genetics*, 10(3), 877–880.

Mangerud, J., Astakhov, V., Jakobsson, M., & Svendsen, J. I. (2001). Huge Iceage lakes in Russia. *Journal of Quaternary Science*, 16(8), 773–777.

Mangerud, J., Jakobsson, M., Alexanderson, H., Astakhov, V., Clarke, G. K., Henriksen, M., Hjort, C., Krinner, G., Lunkka, J. P., Möller, P., Murray, A., Nikolskaya, O., Saarnisto, M., & Svendsen, J. I. (2004). Ice-dammed lakes and rerouting of the drainage of northern Eurasia during the last glaciation. *Quaternary Science Reviews*, 23(11–13), 1313–1332.

Mather, N., Traves, S. M., & Ho, S. Y. (2020). A practical introduction to sequentially Markovian coalescent methods for estimating demographic history from genomic data. *Ecology and Evolution*, 10(1), 579–589.

Mazet, O., Rodríguez, W., Grusea, S., Boitard, S., & Chikhi, L. (2016). On the importance of being structured: Instantaneous coalescence rates and human evolution—Lessons for ancestral population size inference? *Heredity*, 116(4), 362–371.

McGuire, J. A., Linkem, C. W., Koo, M. S., Hutchison, D. W., Lappin, A. K., Orange, D. I., Lemos-Espinal, J., Riddle, B. R., & Jaeger, J. R. (2007). Mitochondrial introgression and incomplete lineage sorting through space and time: Phylogenetics of crotaphytid lizards. *Evolution*, 61(12), 2879–2897.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), 1297–1303.

Miles, A., pyup.io bot, Rodrigues, M. F., Ralph, P., Kelleher, J., Schelker, M., Pisupati, R., Rae, S., & Millar, T. (2024). cggh/scikit-allel: v1.3.8. DOI: 105281:/zenodo.10876220.

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, 37(5), 1530–1534.

Moran, P., & Kornfield, I. (1993). Retention of an ancestral polymorphism in the Mbuna species flock (*Teleostei: Cichlidae*) of Lake Malawi. *Molecular Biology and Evolution*, 10, 1015.

Nadachowska-Brzyska, K., Burri, R., Smeds, L., & Ellegren, H. (2016). PSMC analysis of effective population sizes in molecular ecology and its application to black-and-white *Ficedula* flycatchers. *Molecular Ecology*, 25(5), 1058–1072.

Nazarov, D. V., Nikolskaia, O. A., Zhigmanovskiy, I. V., Ruchkin, M. V., & Cherezova, A. A. (2022). Lake Yamal, an ice-dammed megalake in the west Siberian Arctic during the late Pleistocene, ~ 60–35 ka. *Quaternary Science Reviews*, 289, 107614.

Nei, M., & Takahata, N. (1993). Effective population size, genetic diversity, and coalescence time in subdivided populations. *Journal of Molecular Evolution*, 37, 240–244. https://doi.org/10.1007/BF00175500

Nesbø, C. L., Fossheim, T., Vøllestad, L. A., & Jakobsen, K. S. (1999). Genetic divergence and phylogeographic relationships among European perch (*Perca fluviatilis*) populations reflect glacial refugia and postglacial colonization. *Molecular Ecology*, 8(9), 1387–1404.

Nilsson, J., Andersson, J., Karås, P., & Sandström, O. (2004). Recruitment failure and decreasing catches of perch (*Perca fluviatilis* L.) and pike (*Esox lucius* L.) in the coastal waters of southeast Sweden. *Boreal Environment Research*, 9(4), 295.

Noreikiene, K., Ozerov, M., Ahmad, F., Kõiv, T., Kahar, S., Gross, R., Sepp, M., Pellizzone, A., Vesterinen, E. J., Kisand, V., & Vasemägi, A. (2020). Humic-acid-driven escape from eye parasites revealed by RNA-seq and target-specific metabarcoding. *Parasites & Vectors*, 13(1), 1–11.

Ozerov, M., Noreikiene, K., Kahar, S., Huss, M., Huusko, A., Kõiv, T., Sepp, M., López, M. E., Gårdmark, A., Gross, R., & Vasemägi, A. (2022). Whole-genome sequencing illuminates multifaceted targets of selection to humic substances in Eurasian perch. *Molecular Ecology*, 31(8), 2367–2383.

Ozerov, M. Y., Flajšhans, M., Noreikiene, K., Vasemägi, A., & Gross, R. (2020). Draft genome assembly of the freshwater apex predator Wels catfish (*Silurus glanis*) using linked-read sequencing. *G3: Genes, Genomes Genetics*, 10(11), 3897–3906.

Ozerov, M. Y., Gross, R., Bruneaux, M., Vähä, J. P., Burimski, O., Pukk, L., & Vasemägi, A. (2016). Genomewideintrogressive hybridization patterns in wild Atlantic salmon influenced by inadvertent gene flow from hatchery releases. *Molecular Ecology*, 25(6), 1275–1293.

Patil, A. B., Shinde, S. S., Raghavendra, S., Satish, B. N., Kushalappa, C. G., & Vijay, N. (2021). The genome sequence of Mesua ferrea and comparative demographic histories of forest trees. *Gene*, 769, 145214.

Persson, L., de Roos, A. M., Claessen, D., Byström, P., Lövgren, J., Sjögren, S., Svanbäck, R., Wahlström, E., & Westman, E. (2003). Gigantic cannibals driving a whole-lake trophic cascade. *Proceedings of the National Academy of Sciences*, 100(7), 4035–4039.

Policar, T., Samarin, A. M., & Mélard, C. (2015). Culture methods of Eurasian perch during ongrowing. In P. Kestemont, K. Dabrowski, & R. Summerfelt (Eds.), *Biology and culture of percid fishes* (pp. 417–435). Principles and practices, Springer.

Pollard, D. A., Iyer, V. N., Moses, A. M., & Eisen, M. B. (2006). Widespread discordance of gene trees with species tree in *Drosophila*: Evidence for incomplete lineage sorting. *PLoS Genetics*, 2(10), e173.

Ragauskas, A., Butkauskas, D., Prakas, P., Gadliauskienė, K., Gajduchenko, H., & Grauda, D. (2020). Complex phylogeographic relationships among the Eurasian perch (*Perca fluviatilis*) populations in the eastern part of the Baltic Sea region. *Hydrobiologia*, 847(3), 925–938.

Rambaut, A. (2009). FigTree. Tree figure drawing tool. http://tree.bio.ed.ac.uk/software/figtree/

Ramette, A., & Tiedje, J. M. (2007). Biogeography: An emerging cornerstone for understanding prokaryotic diversity, ecology, and evolution. *Microbial Ecology*, 53, 197–207.

Revell, L. J. (2012). Phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, 3, 217–223.

Snyder, C. W. (2016). Evolution of global temperature over the past two million years. *Nature*, 538(7624), 226–228.

Sommer, R. S., & Zachos, F. E. (2009). Fossil evidence and phylogeography of temperate species: 'Glacial refugia' and post-glacial recolonization. *Journal of Biogeography*, 36(11), 2013–2020.

Stepien, C. A., & Haponski, A. E. (2015). Taxonomy, distribution, and evolution of the Percidae. In *Biology and culture of percid fishes: Principles and practices* (pp. 3–60). Springer Science+Business Media.

Taberlet, P., Fumagalli, L., Wust-Saucy, A. G., & Cosson, J. F. (1998). Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, 7(4), 453–464.

Thorpe, J. E. (1977). Morphology, physiology, behavior, and ecology of *Perca fluviatilis* L. and *P. Flavescens* Mitchill. *Journal of the Fisheries Board of Canada*, 34(10), 1504–1514.

Toews, D. P., & Brelsford, A. (2012). The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*, 21(16), 3907–3930.

Toomey, L., Dellicour, S., Vanina, T., Pegg, J., Kaczkowski, Z., Kouřil, J., Teletchea, F., Bláha, M., Fontaine, P., & Lecocq, T. (2020). Getting off on the right foot: Integration of spatial distribution of genetic variability for aquaculture development and regulations, the European perch case. *Aquaculture*, 521, 734981.

Vasemägi, A., Ozerov, M., Noreikiene, K., López, M.-E., & Gårdmark, A. (2023). Unlocking the genome of perch – From genes to ecology and back again. *Ecology of Freshwater Fish*, 32, 677–702.

Wall, J. D. (2003). Estimating ancestral population sizes and divergence times. *Genetics*, 163(1), 395–404.

Watson, L. (2008). The European market for perch (*Perca fluviatilis*). In *Percid Fish Culture* (pp. 10–16). Presses universitaires de Namur.

Wilke, T., Haase, M., Hershler, R., Liu, H. P., Misof, B., & Ponder, W. (2013). Pushing short DNA fragments to the limit: Phylogenetic relationships of 'hydrobioid' gastropods (*Caenogastropoda: Rissooidea*). *Molecular Phylogenetics and Evolution*, 66(3), 715–736.

Yoder, J. B., Clancey, E., Des Roches, S., Eastman, J. M., Gentry, L., Godsoe, W., Hagey, T. J., Jochimsen, D., Oswald, B. P., Robertson, J., Sarver, B. A. J., Schenk, J. J., Spear, S. F., & Harmon, L. J. (2010). Ecological opportunity and the origin of adaptive radiations. *Journal of Evolutionary Biology*, 23(8), 1581–1596.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Lichman, V., Ozerov, M., López, M.-E., Noreikiene, K., Kahar, S., Pukk, L., Burimski, O., Gross, R., & Vasemägi, A. (2024). Whole-genome analysis reveals phylogenetic and demographic history of Eurasian perch. *Journal of Fish Biology*, 105(3), 871–885. https://doi.org/10.1111/jfb.15821