



OPEN

Hierarchical genetic structure and implications for conservation of the world's largest salmonid, *Hucho taimen*

Lanie M. Galland^{1,2,✉}, James B. Simmons^{1,2}, Joshua P. Jahner^{1,2},
Agusto R. Luzuriaga-Neira^{1,2}, Matthew R. Sloat^{1,3}, Sudeep Chandra^{1,2,4}, Zeb Hogan^{1,2,4},
Olaf P. Jensen^{1,5} & Thomas L. Parchman^{1,2}

Population genetic analyses can evaluate how evolutionary processes shape diversity and inform conservation and management of imperiled species. Taimen (*Hucho taimen*), the world's largest freshwater salmonid, is threatened, endangered, or extirpated across much of its range due to anthropogenic activity including overfishing and habitat degradation. We generated genetic data using high throughput sequencing of reduced representation libraries for taimen from multiple drainages in Mongolia and Russia. Nucleotide diversity estimates were within the range documented in other salmonids, suggesting moderate diversity despite widespread population declines. Similar to other recent studies, our analyses revealed pronounced differentiation among the Arctic (Selenge) and Pacific (Amur and Tugur) drainages, suggesting historical isolation among these systems. However, we found evidence for finer-scale structure within the Pacific drainages, including unexpected differentiation between tributaries and the mainstem of the Tugur River. Differentiation across the Amur and Tugur basins together with coalescent-based demographic modeling suggests the ancestors of Tugur tributary taimen likely diverged in the eastern Amur basin, prior to eventual colonization of the Tugur basin. Our results suggest the potential for differentiation of taimen at different geographic scales, and suggest more thorough geographic and genomic sampling may be needed to inform conservation and management of this iconic salmonid.

Population genetic data is relevant for shaping conservation, restoration, and management activities, and for understanding the response of populations to environmental change. Modern high throughput sequencing technologies have enabled genome-wide perspectives and improved our ability to quantify genetic variation across populations, including those of conservation concern^{1,2}. Reduced representation approaches such as restriction site-associated DNA sequencing (RADseq) and genotyping-by-sequencing (GBS)³ have facilitated genome-wide population genetic analyses in organisms without genomic resources, and have often recovered patterns of fine-scale genetic structure and resolved patterns of recent diversification that were not evident with traditional molecular marker systems^{4,5}. Such approaches have improved our understanding of patterns of population structure and connectivity^{6,7}, the frequency and dynamics of hybridization^{8,9}, and the potential for species response to environmental change^{10,11}, and have guided the identification of conservation or management units^{12,13}.

Globally, many salmonid fishes have experienced sharp declines, especially in recent decades¹⁴, due to anthropogenic factors including aquaculture¹⁵, introduced species, habitat degradation, overfishing, and climate change^{16–18}. Population genetic data have been central to understanding evolutionary history of sensitive salmonid populations^{19,20} and for guiding their conservation and management^{19,21}. High throughput sequencing in salmonids has improved the delineation of evolutionarily significant units¹² and the detection of introgression between introduced and native populations^{22,23}, as well as identifying the genetic basis of adaptive phenotypes^{24–26}. These perspectives, however, have mostly been limited to certain regions where substantial

¹Graduate Program in Ecology, Evolution, and Conservation Biology, University of Nevada, Reno, Reno, NV, USA. ²Department of Biology, University of Nevada, Reno, Reno, NV, USA. ³Wild Salmon Center, Portland, OR, USA. ⁴Global Water Center, University of Nevada, Reno, Reno, NV, USA. ⁵Center for Limnology, University of Wisconsin - Madison, Madison, WI, USA. ✉email: lgalland@unr.edu

funding has been allocated to bolster fisheries conservation and management, such as the North American Pacific Northwest (e.g., see Ref.²⁷).

In contrast, salmonids occurring in sparsely populated regions of northern Asia have received comparatively little research attention. Despite hosting some of the world's most isolated aquatic ecosystems, and being among the least densely populated regions in the world, salmonid populations in this region are rapidly declining as a result of anthropogenic influences with cascading effects on riverine ecosystems²⁸. Still, headwater regions across Siberia and northern Mongolia host some of the world's most pristine rivers and wetlands, receiving some of the highest ecological and chemical status ratings from the European Water Framework Directive standards (e.g., regions including the Selenge River basin headwaters in northern Mongolia²⁹). Here, in the remote headwaters of river basins, species including Siberian taimen (*Hucho taimen*), lenok (*Brachymystax lenok* and *B. tumensis*), grayling (*Thymallus spp.*), and pike (*Esox lucius* and *E. reichertii*) are thought to be more abundant compared to other parts of northern Asia^{30,31}, as there are fewer disturbances from hatcheries, large habitat modifications to the landscape (e.g., deforestation), and dam development.

The taimen is the world's largest salmonid, and one of the most ancient extant members of subfamily Salmoninae^{32,33}. Reaching up to 2 m in length and 100 kg in weight³⁴, taimen reside in home ranges of large and variable size (mean 23 km, maximum up to nearly 100 km³⁵) with movements over 200 km recorded for tagged individuals in northern Mongolia (O.P. Jensen, *unpublished data*), a characteristic that could limit the potential for spatial genetic differentiation. However, taimen form pair bonds from weeks to months before moving to spawning areas (unlike other salmonids³⁶), and upon reaching an acceptable spawning area, a male will aggressively attack other males within a 20 m radius of the female³⁷. Thus, the apparent monogamous nature of taimen spawning behavior, as well as the timing of pair bond formation, could potentially promote spatial genetic differentiation at some scales despite large movements.

Taimen were historically found from the west slope of the Ural Mountains in Eastern Europe to the Pacific Ocean in the east, to the Arctic Circle in the north and the Gobi Desert in southern Mongolia³⁴. Similar to most of the world's largest freshwater fish species³⁸, taimen have been negatively affected by human activity and are listed as “Vulnerable” on the IUCN Red List³⁹. Anthropogenic disturbances (e.g., overfishing, pollution, mining contamination, energy development) have substantially decreased its native range, including putative extirpations of populations in the Volga, Ural, and Pechora River basins⁴⁰. Further, large stretches of rivers in northern Mongolia have seen local extirpations associated with rapid increases in industrialization^{28,40}. Consequently, taimen are listed as threatened or endangered in Mongolia, portions of Russia, Kazakhstan, and China²⁸. Though experiencing population declines across much of its historic range, several of the remaining population strongholds exist in the rivers of northern Mongolia and Siberia³¹.

Understanding patterns of genetic diversity and differentiation among river systems and basins at different geographic scales will be critical for informing taimen conservation strategies. Previous research utilized mtDNA and microsatellite markers to illustrate broad phylogeographic relationships in taimen across its native range^{41–45}. Variation across three mtDNA regions has shown substantial divergence between populations in different basins, with one distinct clade consisting of the Amur (Pacific) and Lena (Arctic) drainages and the other consisting of the Yenesei (Arctic, specifically referring to the drainage downstream of Lake Baikal) and Khatanga (Arctic) drainages⁴¹. More recently, Marić et al.⁴⁵ found two distinct haplogroups representing individuals from the (1) Lena and Amur basins and the (2) Volga (Caspian Sea outflow), Ob (Arctic), Yenesei (Arctic), and Khatanga (Arctic) basins. Kaus et al.⁴⁶ used both mitochondrial and nuclear markers and found pronounced population differentiation between populations from the Amur basin (Pacific) and the Upper Yenesei (Arctic) and Selenge (Arctic, specifically referring to the drainage upstream of Lake Baikal) basins with analyses identifying two ancestral genetic clusters that the authors suggested should be considered separate evolutionarily significant units (ESUs). Importantly, none of the aforementioned studies detected any evidence for genetic differentiation or isolation by distance within these large basins, perhaps suggesting that management plans should be implemented at the basin scale. While the lack of evidence for genetic structure across finer geographic scales within basins is consistent with the potential for large movement and population connectivity among river systems, it could alternatively reflect gaps in our understanding and sampling of taimen genetic variation.

Here, we used high throughput sequencing of reduced representation libraries (ddRADseq⁴⁷) to characterize population genetic structure and diversity of taimen within and among several major river drainages in Mongolia and Russia. Specifically, we used single nucleotide polymorphism (SNP) data from multiple sampling sites within one Arctic and two Pacific drainages to (1) characterize levels of genetic differentiation among drainages, (2) evaluate the potential for fine-scale differentiation within drainages, and (3) quantify levels of genetic diversity across sampling sites. We additionally used coalescent-based demographic modeling to infer the demographic and historical context of divergence between two groups of taimen within one Pacific drainage for which we detected unanticipated genetic differentiation at smaller spatial scales. Our results shed further light on the evolutionary history of taimen and suggest the need for more thorough geographic and genomic sampling to facilitate the development of effective strategies for conservation and management.

Methods

Sample collection, DNA sequencing, and quality filtering. We used catch and release fly fishing to obtain samples from 174 taimen using single, barbless hooks, from five sites across northern Mongolia and four sites in southeastern Russia (Fig. 1, Table 1). These sites are distributed across river systems that drain to the Arctic (Eg, Uur, Delgermörön) and those that drain to the Pacific (Amur basin: Upper and Lower Onon; Tugur basin: Konin, Munikan, Konin/Assyni Junction, and Tugur mainstem). Upon capture, small pelvic fin clip samples were taken and the fish were released. Fin clips were stored either in ethanol or dried in coin envelopes for transport. All methods were carried out in accordance with local and national regulations and guidelines

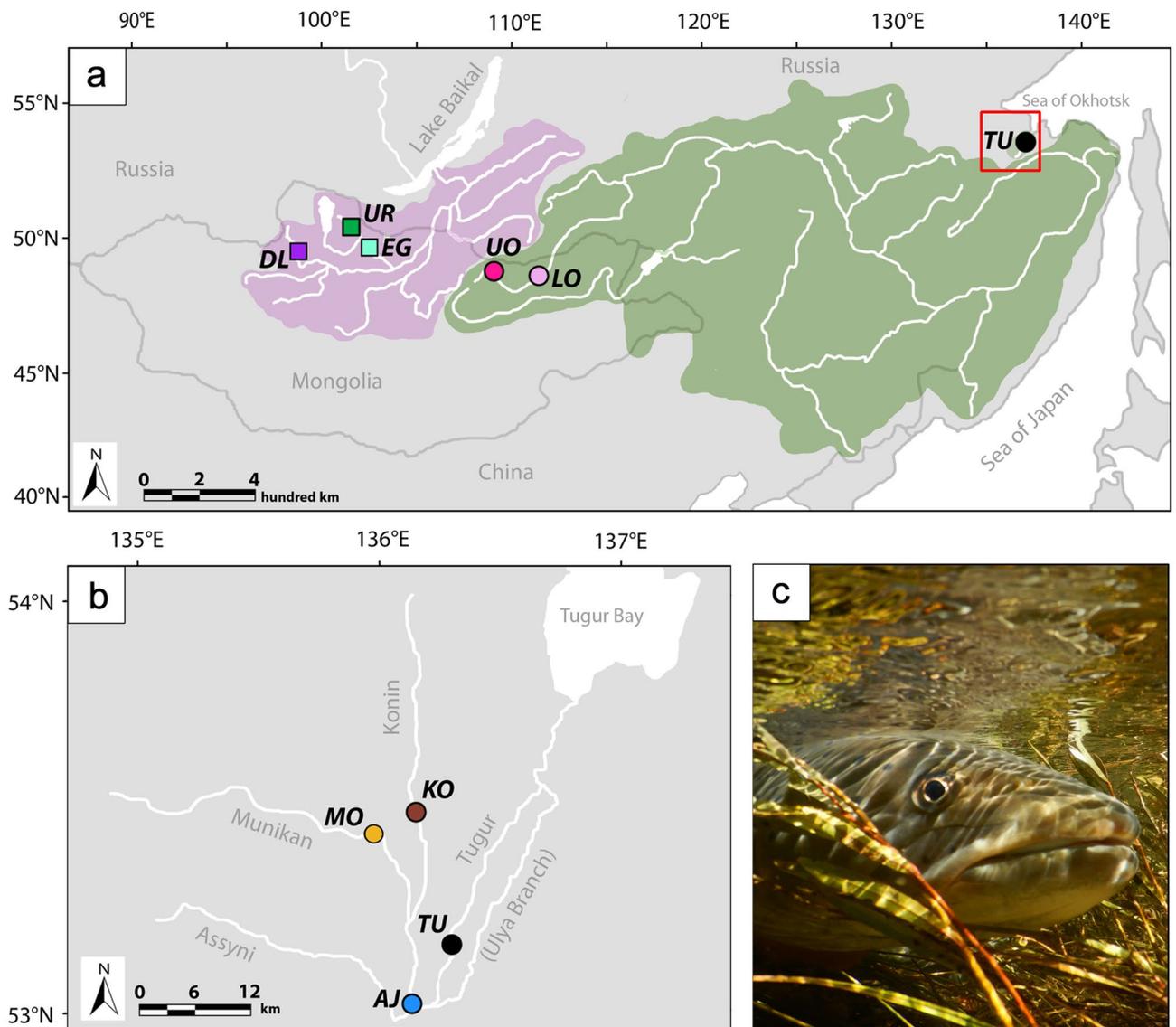


Figure 1. Map illustrating the main drainages from which taimen were sampled from Mongolia and Russia. One Arctic (purple) and two Pacific (green) drainage sampling sites (a). The red rectangle denotes the Tugur basin sampling area, shown in greater detail (b). Sampling sites correspond to those in Table 1 and include the Delgermörön (DL), Eg (EG), Uur (UR), Lower Onon (LO), Upper Onon (UO), Tugur (TU), Konin/Assyni Junction (AJ), Konin (KO), and Munikan (MU) sampling sites. A recently released taimen rests in the flooded grassland (c). Panels (a) and (b) were modified using Adobe Creative Suite.

Outflow	Basin	Location name	Latitude, longitude	N
Arctic	Selenge	Eg (EG)	49.84, 102.57	13
		Uur (UR)	50.34, 101.87	62
		Delgermörön (DL)	50.11, 98.85	18
Pacific	Amur	Upper Onon (UO)	49.24, 112.02	6
		Lower Onon (LO)	49.12, 111.96	5
	Tugur	Konin (KO)	53.28, 136.26	5
		Munikan (MO)	53.27, 136.12	3
		Konin/Assyni Junction (AJ)	53.07, 136.03	19
		Tugur (TU)	53.11, 136.25	43

Table 1. Geographic locations, sample sizes, and capture information for taimen analyzed in this study. Site abbreviations correspond to those in Figs. 1 and 2.

(including fishing permits obtained from the local governments), and all experimental protocols were approved by the University of Nevada, Reno, International Animal Care and Use Committee (IACUC protocol ID 20-10-1098). Genomic DNA was isolated from fin clips using Qiagen DNeasy Blood and Tissue kits (Qiagen Inc., Valencia CA), and DNA concentrations were quantified using a Qiagen QIAxpert microfluidics analyzer. Due to variability in DNA yield from fin clip extractions, DNA samples were standardized to within the range of 20–40 ng/uL to ensure similar template concentrations for sequencing library preparation.

We used a reduced representation approach based on two restriction endonucleases to generate ddRADseq libraries. DNA was digested with restriction endonucleases *MseI* (4-base recognition site) and *EcoRI* (6-base recognition site). To the *EcoRI* cut-sites, we ligated Illumina adaptors embedded with unique 8–10 bp barcode sequences which were used to tag DNA from each individual. An Illumina sequencing adaptor was ligated to the *MseI* cut-sites. We then pooled the uniquely barcoded samples and amplified fragments using Illumina PCR primers. To reduce the portion of the genome sampled for sequencing, libraries were size selected for fragments ranging from 350 to 450 bp using a Pippin Prep unit (Sage Science, Beverly, MA) at the University of Texas Genome and Sequencing Analysis Center (Austin, TX). Full details on the laboratory methods used for library preparation can be found in Ref.⁴⁸ Size selected libraries were then sequenced on two lanes of an Illumina HiSeq 2500 at the University of Wisconsin-Madison Biotechnology Center's Genome Center (Madison, WI).

Raw sequencing data was filtered for contaminant sequences including *E. coli* and *PhiX*, and for Illumina sequencing adaptors, using `bowtie2_db`⁴⁹ and a pipeline of Perl and bash scripts (<http://github.com/ncgr/tapioca>). Importantly, barcode sequences all differed by a minimum of three bases, allowing us to detect one or two base sequencing errors within them. A custom Perl script (available at DRYAD doi: <https://doi.org/10.5061/dryad.wstqjq2kd>) was used to correct 1 or 2 bp sequencing errors in barcode sequences, remove barcodes and cut-site associated bases, and match sequences with individual sample names. Reads were then split into fastq files specific to each individual. We then removed individuals represented by volumes of sequencing data lower than the 1st quartile of the distribution spanning all samples to reduce the fraction of missing data and to increase the number of loci retained for analyses of a sufficient number of samples per sampled locality.

Alignment, variant calling, and filtering. Since we quantified substantial genetic differentiation among samples from the two major outflows (Pacific and Arctic, see “Results” section below), we conducted separate analyses based on: (1) all individuals (full dataset); and (2) a subset of individuals sampled from populations in the Pacific drainages (Pacific subset). We did not analyze data separately for the Arctic subset, as analysis of the full dataset (in “Results” section) suggested limited genetic structure among populations within this drainage. All methods for alignment, variant calling, and filtering were identical for both datasets, except for the de novo reference generation, which was implemented separately with the different sets of samples. As there were no reference genomes available for any closely related taxa, we used a de novo clustering approach to build a reference of genomic regions sampled with our sequencing approach as a basis for aligning all of the reads. We used CD-HIT⁵⁰ to generate contig consensus sequences (partial reference hereafter) built from clustering the unique sequences in our entire dataset with a minimum match percentage of 90%. This de novo clustering algorithm, also utilized by the commonly used RADseq pipeline dDocent⁵¹ was used to generate a partial reference to serve as a target for subsequent read mapping. We used `bwa v0.7.5`⁵² to map sequences generated for each individual to the partial reference based on an edit distance of three. Sequence alignment map (.sam) files were converted to binary alignment map (.bam) files with `samtools v1.3`⁵³, before `samtools v1.3` and `bcftools v1.3`⁵³ were used to identify and call variants across the alignments of all individuals. We calculated genotype likelihoods for SNPs at sites with a minimum base quality of 20, maximum coverage depth of 100, minimum map quality of 20, minimum site quality of 20, and minimum genotype quality of 10. We used `vcftools v0.1.14`⁵⁴ to further filter called variants. We retained only bi-allelic SNPs with minor allele frequencies (MAF) greater than 0.04, and those where at least 60% of individuals had at least one read. We randomly sampled one SNP per 100 bp contig and discarded individuals with missing data at more than 30% of loci.

As mis-assembly of reads representing paralogous regions can lead to genotyping error in high throughput sequencing data⁵⁵, we took several steps to mitigate the potential influence of such loci. First, we used `vcftools v0.1.14` to remove loci with exceptionally high coverage depth per individual, greater than or equal to 50. We then additionally identified and removed potentially paralogous loci using the `HDplot` approach described in Ref.⁵⁵. This method identifies potential paralogs in sequence data from populations based on deviations from the expected frequency of heterozygotes and from the expected 1:1 ratio of read counts for alternate alleles in heterozygotes. Here, we retained loci with heterozygosity (H) levels between 0 and 0.6, and read ratio deviance (D) between -18 and 18. We took these steps to exclude potentially misassembled genomic regions representing duplicate or diverged duplicate loci.

Population genetic analyses. We quantified population structure without a priori sample information using the Bayesian ancestry-based model `entropy v1.2`^{56,57} which is based on the correlated allele frequency model of `structure`⁵⁸. We used `entropy` to infer the number of k ancestral populations represented by the data and to estimate admixture proportions (q) for individuals. Importantly, this model accounts for statistical uncertainty arising from sequencing and alignment error and stochastic variation in coverage depth inherent in low to medium coverage sequencing data^{59,60}. Because `entropy` provides posterior estimates of genotype probabilities at each locus for each individual, it allows for the incorporation of genotype uncertainty into downstream population genetic analyses. To seed and speed the convergence of Markov chain Monte Carlo runs, we first generated starting values for the q parameter. We conducted a principal component analysis (PCA) on the covariance matrix of genotype likelihoods calculated above using the `prcomp` function in R version 4.1⁶¹ and then used k -means clustering and linear discriminant analyses (LDA) to estimate ancestry proportions

for each individual for models representing $k=2$ through $k=9$ (or through $k=6$ for the Pacific dataset). We ran `entropy` models for $k=2$ through $k=9$ (or $k=6$) ancestral groups, with 5 replicate chains per k . We ran 100,000 MCMC iterations, retaining every tenth step after a burn-in of 30,000 steps. Model fit was assessed using the deviance information criterion (DIC), where lower DIC values represent better model fit. We conducted `entropy` runs separately for (1) the set of SNP genotype likelihoods for all individuals and (2) the subset of samples representing locations within the Pacific drainages. We used genotype probabilities from `entropy` for the majority of the population genetic analyses described below.

We additionally characterized genetic variation with PCA using the `prcomp` function in R. As metrics of pairwise genetic differentiation among populations, we calculated Hudson's F_{ST} ⁶² and Nei's D ⁶³ based on population allele frequencies. As metrics of genetic diversity for each sampling location, we calculated nucleotide diversity (θ_n , or the average number of pairwise differences between sequences⁶⁴), Watterson's theta (θ_w , or the number of segregating sites⁶⁵), and the scaled difference between the two (Tajima's D ⁶⁶) using methods that incorporate genotype uncertainty implemented in `ANGSD`^{67,68}. We used the *de novo* artificial reference genome and individual `.bam` files to estimate site allele frequency likelihoods using the “doSaf 1” tool. We then used site allele frequency likelihoods as input for `REALSFS`⁶⁸ to generate folded site frequency spectrum likelihoods. Using “doSaf 1,” we calculated posterior allele frequency probabilities. Lastly, we used the `thetastat` utility from `ANGSD` to estimate per-locus measures of each diversity metric and generated the per-population average of these values over all contigs and nucleotides. For comparison with other studies, we also calculated expected heterozygosity based on allele frequencies.

As we found unexpected evidence for divergence between taimen in the Tugur mainstem and its tributaries, we used a coalescent-based demographic modeling approach to explore parameters characterizing the divergence and demographic history of these two populations. The site frequency spectrum (SFS) contains the signatures of divergence and demographic processes (e.g., time, migration, changes in effective population size), and high throughput sequencing data has substantially improved our ability to estimate such parameters from population genetic data^{69,70}. Before generating the SFS, we further filtered the `vcf` file generated above. First, we removed variants with $MAF < 0.1$ to guard against rare variants that could represent sequencing errors and trimmed outlier loci with $F_{ST} > 0.15$ (the 0.95 quantile of the F_{ST} distribution) between the two populations. We generated the unfolded SFS for each population using `easySFS` (<https://github.com/isaacovercast/easySFS#easysfs>) on the stringently filtered `vcf` file, down sampling populations to sizes of 10 and 10 (`-proj 10, 10`).

Using the unfolded SFS, we estimated demographic parameters for eight different models using coalescent simulation and a maximum likelihood framework in `fastsimcoal2`⁶⁹. These models represented two-population divergence (Tugur mainstem and Tugur tributaries), with and without migration, and with and without population expansion or contraction. We ran 50,000 coalescent simulations per replicate and a total of 50 replicates, with minimum (`-n`) of 100,000 simulations for a total (`-L`) of 40 cycles. We used a mutation rate for salmonids (Salmoninae) of $8e-9$ bp per generation for model estimation (coho salmon⁷¹), and to estimate coalescent effective population size. For each model, the replicate with the smallest difference between the maximum expected likelihood (MEL) and the maximum observed likelihood (MOL) represented the best-fit run⁶⁷. To account for differences in the number of parameters included in each model, we calculated AIC scores for each model's best-fit run. We then calculated ΔAIC to compare models.

Following parameter estimation for the best fit model, we calculated 95% confidence intervals for each parameter. Using the maximum likelihood parameters from the `*_maxL.par` file, we generated 100 bootstrap replicates of the SFS. Next, we estimated parameters of these 100 SFS using the same 50-replicate analyses described above. The best-fit model parameter estimates for each of the new 100 simulated SFS were used to calculate mean parameter estimates and subsequently to infer 95% confidence intervals.

Results

Full dataset. After filtering for contaminants and removing individuals lacking sufficient sequencing data, we retained 174 individuals with a mean number of 1,898,867 reads per individual. `bwa` mapped reads from all individuals onto the *de novo* partial reference consisting of 221,178 genomic regions. After variant calling and filtering based on sequencing coverage and quality, we retained 7597 loci with $MAF > 0.04$. We additionally discarded 1551 SNPs that potentially represented paralogous regions, leaving a final set of 6046 SNPs from 174 individuals (mean coverage = 10.1X per locus per individual). Sequence data is available on NCBI's Short Read Archive (accession PRJNA745962; <https://dataview.ncbi.nlm.nih.gov/object/PRJNA745962>). Both the sequence data and the `vcf` file are available at DRYAD (doi: <https://doi.org/10.5061/dryad.wstqjq2kd>).

Pronounced genetic differentiation was evident between taimen sampled from the Arctic and Pacific drainages across all analyses (Figs. 2a, 3). DIC values from `entropy` indicated that the $k=2$ model best fit the data, though the $k=3$ model had similar support and illustrated additional population differentiation (Supplementary Table S1). Individuals were assigned nearly 100% to one of two ancestries (Fig. 2a): the Arctic with the single Selenge drainage (Eg, Uur, and Delgermörön Rivers), and the Pacific with the Amur (Upper and Lower Onon River sites) and the Tugur (Konin, Munikan, Konin/Assyni Junction, and Tugur Rivers) drainages (Fig. 1). The $k=3$ model reflected the same pattern for the Arctic versus Pacific drainages, but individuals were assigned with variable ancestry across several Pacific sites exhibiting differentiation (Fig. 2b).

PCA of the genotype probabilities revealed genetic structure similar to ancestry estimates from `entropy` (Fig. 3a). PC 1 explained 59.99% of variation in the data, separating individuals and populations from the Arctic and Pacific drainages. PC 2 explained only 0.98% of variation in the data, but suggested finer-scale structure within the Pacific drainages. Pairwise measures of genetic divergence among sites from the two outflows (Arctic and Pacific) were high (F_{ST} range = 0.209–0.258, Nei's D range = 0.1405–0.1823; Supplementary Table S2) consistent with independent evolutionary histories and substantial isolation of drainages (Table 2). A neighbor joining

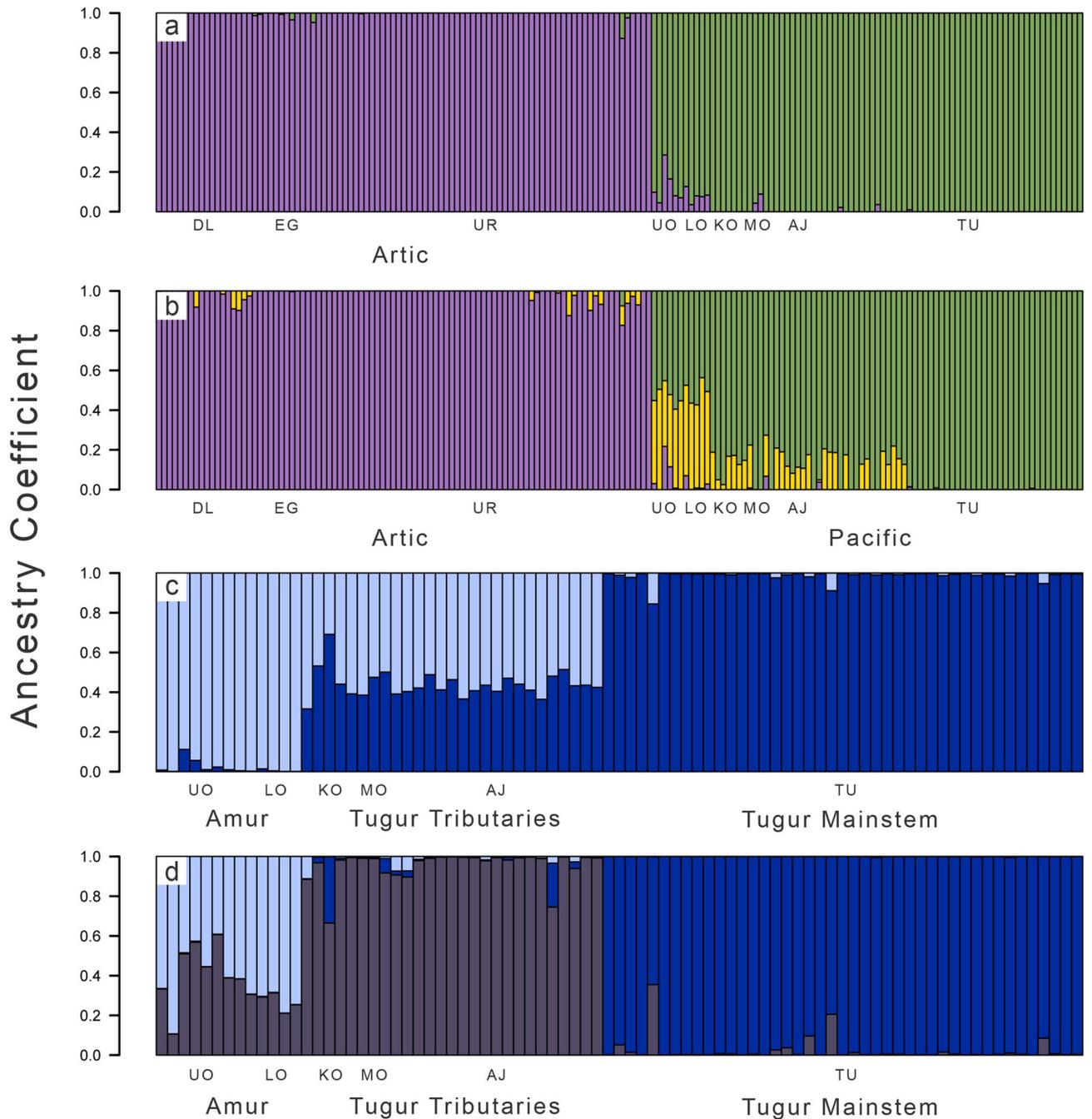


Figure 2. Ancestry coefficient estimates (q) generated with `entropy` for analyses based on all sampled individuals (**a, b**) and for separate analyses based on the subset of individuals from the two Pacific drainages (**c, d**). Vertical bars represent individuals, and colors correspond to the admixture proportions for each of k clusters. For both sets of analyses, the $k=2$ models (**a, c**) fit the data best. Models for $k=3$ (**b, d**) are additionally shown for each set of analyses as the revealed patterns of clustering further illustrate population genetic structure within river systems. As in Fig. 1, sampling sites correspond to those in Table 1 and include the Delgermörön (DL), Eg (EG), Uur (UR), Lower Onon (LO), Upper Onon (UO), Tugur (TU), Konin/Assyni Junction (AJ), Konin (KO), and Munikan (MO) sampling sites. Plotting was completed using R software.

tree based on pairwise estimates of Nei's D among sampling localities provided and alternative visualization of hierarchical genetic structure consistent with all other population genetic analyses (Fig. 3d).

Genetic diversity varied across the sampled geographic regions. Mean genetic diversity across all nine sampling sites was 0.00164 for θ_π (range 0.00119–0.00209) and 0.00132 for θ_w (range 0.00077–0.00178). Genetic diversity was consistently higher in Pacific drainages (mean $\theta_\pi = 0.00181$, range 0.00165–0.00209; mean $\theta_w = 0.00152$, range 0.00119–0.00178) than in the Arctic drainage (mean $\theta_\pi = 0.00130$, range = 0.00119–0.00138; mean $\theta_w = 0.00093$, range = 0.00077–0.00101), as was H_E (Arctic mean = 0.1195, range = 0.118–0.121; Pacific

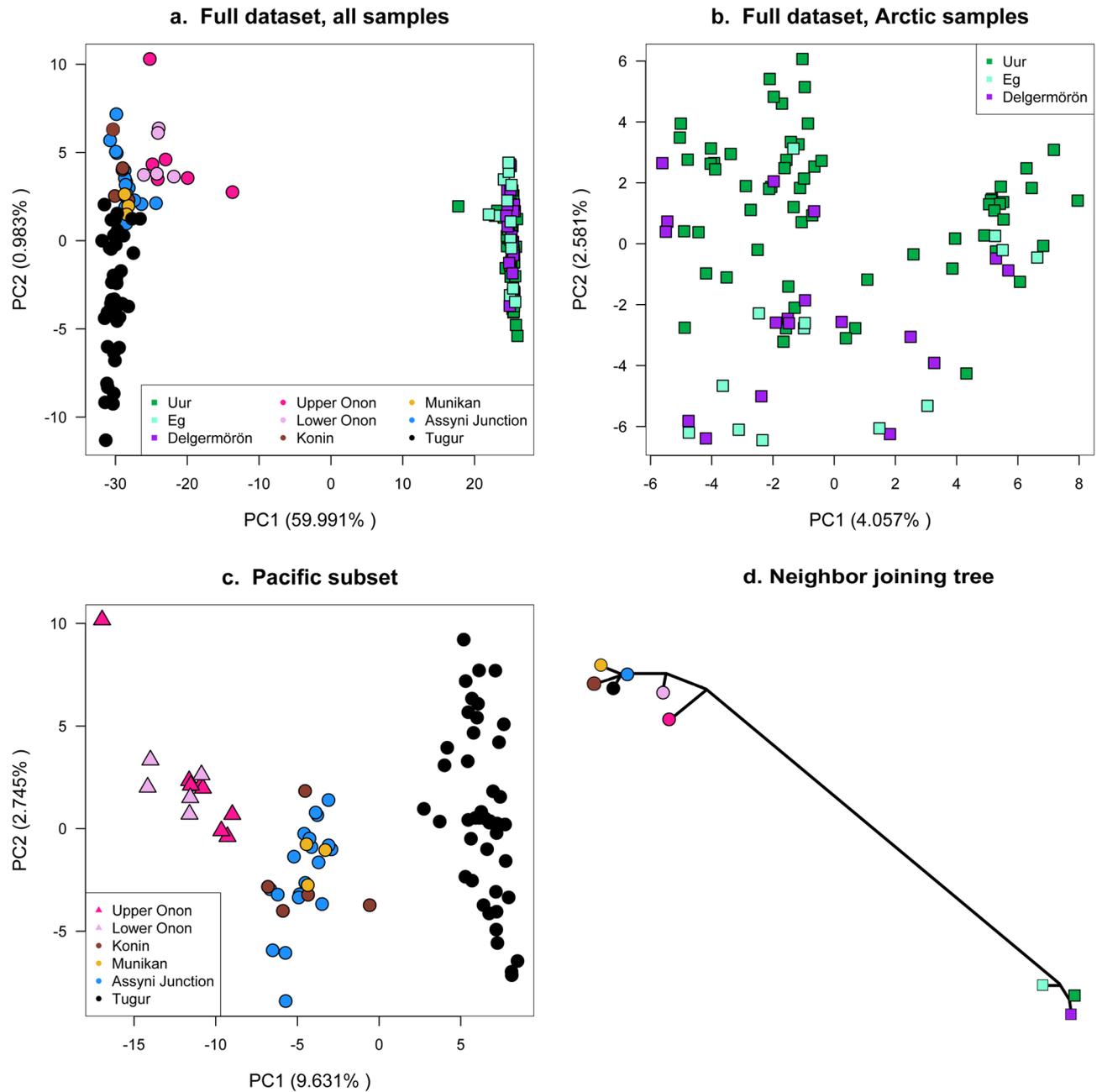


Figure 3. Genetic variation among taimen illustrated with PCA (calculated using R software) of 6046 SNPs called in all sampled individuals (a), Arctic individuals (b), and in 3961 SNPs called separately for individuals from the Pacific drainages (c). In panel (a), samples from the Pacific drainages are represented by circles, and those from the Arctic drainage by squares. In panel (c), samples from the Pacific drainages are represented by triangles while those from the Tugur basin are represented by circles. The neighbor joining tree in panel (d) supports the deep divergence between Arctic and Pacific drainages, where symbols represent populations shown in panel (a).

mean = 0.2203, range = 0.213–0.230). No populations from the Arctic drainage had confidence intervals that overlapped with those in the Pacific drainages. Importantly, for taimen localities sampled in the current work, sample size was unrelated to both genetic diversity metrics (sample size vs. heterozygosity: $r = -0.327$, $P = 0.323$; sample size vs. nucleotide diversity: $r = -0.424$, $P = 0.253$; Supplementary Fig. S1).

Analyses of the Pacific drainages. Given pronounced genetic differentiation between the Arctic and Pacific drainages and evidence for finer-scale differentiation among river systems within the Pacific drainages, we conducted assembly and variant calling separately for Pacific populations. Unique reads were assembled into a partial reference consisting of 160,858 contigs. Following reference mapping, variant calling, subsequent bioinformatic processing, and paralog filtering, we retained 3961 SNPs from 83 individuals (mean coverage = 10.4X

Outflow	Basin	Site	H_E	θ_π	θ_π low	θ_π high	θ_W	θ_W low	θ_W high	D	D low	D high
Arctic	Selenge	DL	0.118	0.001187	0.001177	0.001197	0.000774	0.000768	0.000779	-0.00665	-0.00668	-0.00662
		EG	0.121	0.001348	0.001338	0.001358	0.001006	0.00100	0.001013	-0.01024	-0.01026	-0.01021
		UR	0.120	0.001377	0.001367	0.001387	0.000999	0.000993	0.001004	-0.01266	-0.01270	-0.01263
Pacific	Amur	UO	0.227	0.002090	0.002077	0.002103	0.001777	0.001767	0.001787	0.00210	0.00207	0.00213
		LO	0.217	0.001757	0.001744	0.001770	0.001506	0.001496	0.001516	0.00475	0.00472	0.00478
	Tugur	KO	0.214	0.001646	0.001635	0.001658	0.001391	0.001382	0.001400	-0.00025	-0.00028	-0.00022
		MO	0.213	0.001875	0.001861	0.001889	0.001672	0.001661	0.001684	0.00835	0.00832	0.00838
		AJ	0.221	0.001775	0.001765	0.001786	0.001585	0.001578	0.001591	-0.02457	-0.02460	-0.02454
		TU	0.230	0.001694	0.001682	0.001705	0.001186	0.001179	0.001192	-0.00942	-0.00946	-0.00938

Table 2. Population genetic diversity for the full set of 174 individuals, including estimates based on population allele frequencies (H_E) and those based directly on DNA sequence variation across the sampled genomic regions (the average number of pairwise differences between sequences, or θ_π , the number of segregating sites, or θ_W , and the scaled difference between these two, or Tajima's D). High and low columns represent upper and lower 95% confidence intervals for each metric.

Model description	Model #	Growth parameters		N_e		Ancestral N_e	$T_{\text{divergence}}$ (generations)	Migration		Δ Model likelihood	Δ AIC
		TU	TT	TU	TT			TU \rightarrow TT	TT \rightarrow TU		
2 Population with gene flow	1	-	0	40,999 (38,052–43,946)	10,058 (7998–12,119)	446,164 (398,862–493,466)	28,480 (23,685–33,276)	0.00356 (0.00279–0.00434)	0.00435 (0.00386–0.00485)	31.13	-
	2	0	-	3302	83,980	734,783	5046	0.00251	0.59000	39.68	39.36
	3	-	-	47,370	81,411	253,462	53,643	0.00377	0.00027	36.32	23.88
	4	+	+	21,677	72,582	380,356	2587	0.00702	0.00574	35.73	21.17
	5	0	+	505	54,433	743,812	2274	0.00489	0.00605	38.05	31.84
	6	+	0	18,046	7769	816,466	5739	0.00170	0.00401	36.72	25.73
	7	0	0	687	1724	523,521	1247	0.00596	0.00464	35.67	20.88
2 Population	8	0	0	435	791	110,201	38	-	-	136.83	484.7

Table 3. Demographic modeling parameter estimates for each of eight 2-population models. TU refers to Tugur mainstem, while TT refers to Tugur tributaries. Growth parameters for each population represent expansion (+), contraction (-), and no change (0) following the ancestral divergence. Model numbers correspond to those in Fig. 4. For the best fit model (model 1), 95% confidence intervals are given in parentheses.

per locus per individual) for the Pacific subset. The vcf file associated with the Pacific subset is available at DRYAD (doi: <https://doi.org/10.5061/dryad.wstqj2kd>).

Although pairwise measures of genetic differentiation among sampling sites across this drainage were relatively low (Supplementary Table S2), analyses illustrated finer-scale patterns of differentiation that were less evident in analyses of the full data set. The $k=2$ entropy model fit the data best, although the $k=3$ model had similar DIC support (Supplementary Table S1), and further illustrated population structure. Across analyses, clear genetic differentiation was evident between the Amur (Upper and Lower Onon sites) and Tugur drainages (Figs. 2c, d, 3b). Both the $k=2$ and $k=3$ entropy models assigned individuals from the Amur and Tugur drainages to different ancestral clusters (Fig. 2c, d), and PCA clearly separated individuals from the two drainages (Fig. 3c).

There was unexpected genetic differentiation among sampling sites within the Tugur basin; taimen from the headwater tributaries (KO, MO, AJ; $n=37$) were differentiated from samples taken in the mainstem Tugur (TU; $n=43$; Fig. 1b). Ancestry was assigned differentially to separate clusters in both the $k=2$ and $k=3$ entropy models (Fig. 2c, d), and the tributary populations formed a non-overlapping cluster in PC space intermediate between samples from the Tugur and the Amur drainages (Fig. 3c). Although overall genetic differentiation was subtle (F_{ST} range = 0.012–0.029; Supplementary Table S2), these analyses nonetheless indicate the presence of finer-scale differentiation among sampling sites in the Tugur than in the other drainages we examined.

After the more stringent filtering with `vcftools v0.1.14`, we retained 1971 variants from which we constructed the unfolded SFS. Coalescent simulations using `fastsimcoal2` were run for eight models spanning variation in divergence and demography of the Tugur mainstem and Tugur tributary populations (see Table 3 for parameter estimates and model comparison metrics, and Fig. 4 for model schematics). Importantly, all models including migration had substantially better fits than the model without migration. Model likelihoods were similar across models including migration, but the best fit model included initial divergence, after which coalescent N_e remained constant in the tributary population but contracted in the mainstem population

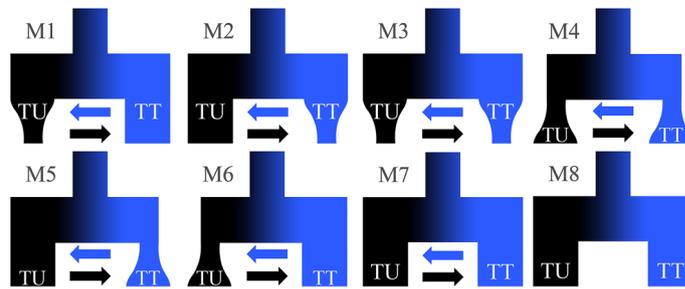


Figure 4. Representations of each two-population model tested with *fastsimcoal2*, where TU represents the mainstem population, TT represents the tributaries population, and arrows represent gene flow. Model numbers and parameters correspond to those listed in Table 3.

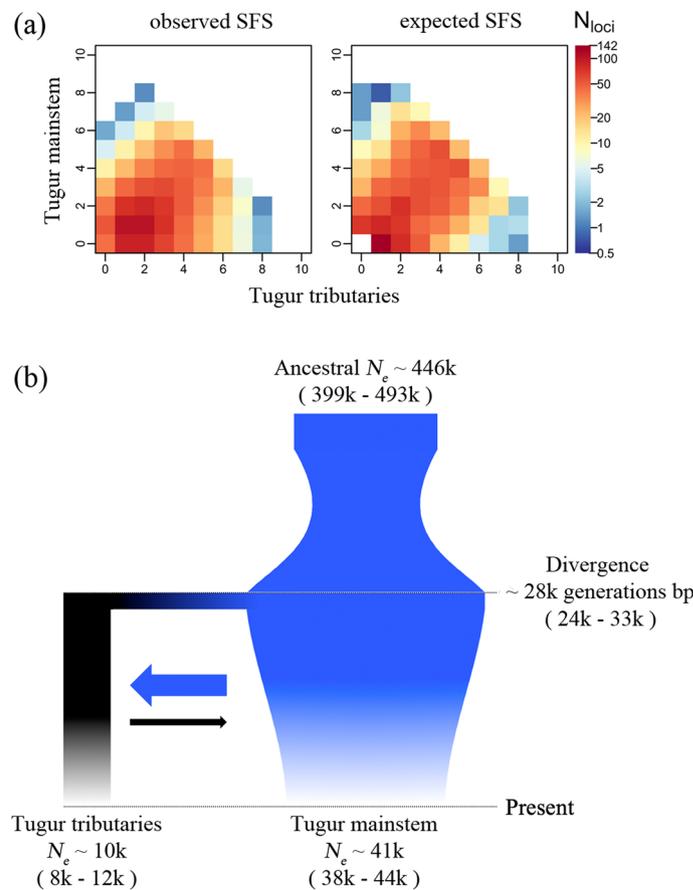


Figure 5. The expected versus observed SFS for the best fit model shows substantial overlap (a). Panel (b) shows the schematic representing the best fit model (model 1) from demographic inference using *fastsimcoal2*. Coalescent N_e is given for population, with branch and arrow width corresponding to population size and level of gene flow, respectively (but not drawn to scale). Numbers in parentheses represent 95% confidence intervals.

(Table 3; Fig. 5). Parameter estimates indicated divergence at ≈ 28 k generations ago based on a mutation rate of $8e-9$ (mean ≈ 393 kya based on mean generation time of 13.8 years [range $\approx 195-594$ kya based on generation time range 6.9–20.8 years (mean ± 2 standard deviations)]^{34,72}; Fig. 5). Generation time was calculated from age-frequency data from the Tugur River (M.R. Sloat, *unpublished data*) and fecundity-at-age data presented by Ref.³⁴. Following the ancestral divergence event, asymmetrical gene flow was inferred between mainstem and tributary clusters, with greater gene flow from the mainstem to tributaries. Coalescent N_e for tributaries was substantially lower than that of the mainstem ($\approx 10,000$ vs. $\approx 41,000$), and the model indicated population contraction in the latter.

Discussion

Using high throughput sequencing of reduced representation libraries, we documented hierarchical patterns of genetic differentiation among taimen sampled across multiple drainages. We found pronounced divergence among taimen from the Arctic and Pacific drainages, but also recovered more subtle patterns of differentiation within and among river systems that drain to the Pacific. Although our spatial sampling of the distribution was limited, our analyses indicate the potential for genetic differentiation at finer scales within basins and even within specific river systems. This aspect of our results differs from several past studies and was potentially influenced by different spatial and more thorough genomic sampling. Below, we discuss our results in the context of past genetic work on taimen, with consideration of how current and future population genetic analyses could inform their management and conservation.

Deep genetic divergence between Arctic and Pacific drainages. Our population genetic analyses illustrate substantial genetic differentiation consistent with significant historical isolation between taimen from the Arctic and Pacific drainages (mean $F_{ST} = 0.24$; Fig. 1, Supplementary Table S2), similar to past studies based on smaller sets of genetic markers^{41,45,46}. Additionally, maximum likelihood phylogenetic analyses (RAxML⁷³) based on 1229 SNPs in a concatenated alignment similarly illustrated deep divergence among taimen sampled from the Pacific and Arctic drainages (Supplementary Methods, Supplementary Fig. S2). Genetic differentiation across systems would be expected given the geologic barriers separating these drainages. Though the Amur and Selenge basins are separated by only a few kilometers (Fig. 1), the Khentii Mountains arising from the Baikal Rift system (last active during the Pliocene⁷⁴) form a continental divide that likely severed any recent connectivity between the Arctic and Pacific river systems. This geologic feature is associated with deep phylogeographic breaks for other taxa occurring in this region, including other salmonids (*Thymallus*^{75,76}, *Brachymystax*^{46,77}), as well as other groups of freshwater fishes (*Cottus*⁷⁸, *Esox*⁷⁹). Divergence time estimates from mtDNA analyses indicate divergence across this divide in the range of 1–2.5 mya for *Cottus*⁷⁸ and 0.5–2 mya for taimen^{41,45}.

Taimen from the Arctic and Pacific outflows do differ in a few key morphological traits, including average mass per length, and density and size of spot patterns (Mikhail Skopets and M.R. Sloat, *unpublished data*). This variation is consistent with isolation and independent evolution, and could possibly be due in part to ecological differences among drainages on either side of this divide, such as riverscape and landscape topography^{35,72}, constituent riparian species composition³⁵, and marine food web subsidies from seasonal spawning runs of chum salmon in the Tugur basin, but not the Selenge basin⁷². Unfortunately, the geographically sparse and clumped nature of our sampling led to a strong correlation between geographic and environmental distances (based on 19 BioClim variables) among sampling localities, which precluded formal tests of environmental influences on spatial genetic structure.

We did not detect genetic differentiation among sites within the Arctic drainage (Figs. 2, 3), though only three locations, all within the Selenge basin, were sampled in this drainage. Future studies could benefit from additional sampling across the Arctic drainage, particularly upstream and downstream of Lake Baikal, which may have presented a migration barrier to the predominantly riverine taimen. This apparent absence of genetic structure is consistent with past studies based on microsatellite and mtDNA data^{45,46}, and with the possibility of population connectivity among sites separated by hundreds of kilometers. Limited genetic differentiation across broad spatial scales would not be surprising given the large size, long lifespan, and substantial opportunity for movement in taimen³⁵. Moreover, the region's glacial history suggests that populations in the Arctic are likely younger than those in the Pacific drainages. Northwestern Siberia experienced repeated Pleistocene glaciations which blocked north-flowing rivers and formed ice dam lakes^{80–82}, whereas the more southern Pacific drainages are thought to have been less influenced by glaciation⁸³ (but see Ref.⁸⁰). A younger taimen lineage in the Arctic drainage would be consistent with limited differentiation among populations often seen in Arctic fishes^{84,85}, and also with variation in genetic diversity among the regions we sampled.

For all metrics, levels of genetic diversity were consistently lower for the three Arctic drainage sites than those from Pacific drainage sites (Table 2). Measures of nucleotide diversity were relatively low, but well within the range of published nucleotide diversity estimates across salmonids^{86–89}. For example, our values for taimen were similar to those reported for Atlantic salmon (*Salmo salar*, overall nucleotide diversity of 3.99e–4; Ref.⁹⁰), a species that has also been strongly influenced by recent glacial periods⁹⁰. Lower diversity in the Arctic drainage is consistent with the hypothesis that Arctic populations are younger, and/or suffered recent bottlenecks during Pleistocene glacial activity in this region⁸².

Fine-scale genetic structure within the Pacific drainages. Our analyses revealed clear, though less pronounced, genetic differentiation between the two Pacific drainages (Amur and Tugur). Taimen from the Onon River (Amur basin) were differentiated from those occurring 1800–2000 km to the east in the Tugur basin (mean $F_{ST} = 0.033$). This pattern could be consistent with the Yablonovy and Stanovoi Mountain ranges acting as a divide for aquatic fauna separating the headwaters of the Lena (just north of the Tugur) and Amur basins⁹¹, although a lack of sampling over a large area between the Onon and Tugur sites limits our understanding of geographic features associated with spatial genetic structure. As in the Arctic drainage, we did not detect evidence for differentiation among sampling sites within the Amur basin. However, given the size and complexity of the Amur basin combined with the population structure documented across the basin in other salmonids (e.g., *Thymallus*⁷⁶), it is possible that more thorough sampling could reveal additional structure.

Our analyses revealed unexpected evidence for fine-scale genetic differentiation between two groups of taimen sampled in the Tugur basin. Though sampling locations within this basin are highly proximate (often separated by less than 10 km, Fig. 1) and well within average home ranges of taimen^{35,92}, we observed distinct genetic differentiation between the Tugur River mainstem and its tributaries (Figs. 1b, 2c, d, 3c). Differentiation among these

groups was subtle but clear (Supplementary Table S2), as individuals from the mainstem and tributary groups were completely identifiable and formed non-overlapping groups in PCA and ancestry-based analyses (Figs. 2d, 3c). No evident barriers to movement (and thus spawning) have been observed in this region, and taimen from these sites are not known to differ morphologically. Nonetheless, the pattern of consistent differentiation among these sites indicates that some barrier to gene flow likely exists. Ecological variation among tributaries and the mainstem has been noted, including differences in water flow, water level, and chemical composition, as well as the timing of food availability (e.g., variation in chum salmon spawning⁹³), suggesting ecological factors may underlie isolation. Reproductive isolation has evolved within a number of salmonid species due to spatiotemporal differences in spawning (spring vs. fall Chinook salmon²⁶), or morphological specialization (dwarf and normal *Coregonus clupeaformis*⁹⁴; benthic and limnetic *Salvinus alpinus*⁹⁵; pelagic and littoral feeding *Thymallus nigrescens*⁹⁶). Additionally, although taimen populations typically do not differentiate at smaller spatial scales, anadromous Sakhalin taimen (*Parahucho perryi*) demonstrate population genetic structure arising from differences in spawning grounds and homing behavior⁹⁷. While we are unaware of such mechanisms underlying divergence in taimen, our results indicate the potential for differentiation at smaller geographic scales than previously detected and the need for further study to understand its evolutionary causes and consequences for management.

Given the unexpected differentiation within the Tugur basin, we compared empirical and simulated SFS for models representing different divergence scenarios to consider the timing and demographic context of this divergence. All seven of the models incorporating migration had strongly improved fit compared to the model without migration (Table 3), consistent with a history of allopatric divergence followed by secondary contact and gene flow. The best fit model included bidirectional gene flow, with population size constant in the Tugur tributaries and contracting in the mainstem. The divergence time estimate for this model of ~28 k generations would correspond to divergence at 195–594 kya, depending on generation time estimates. The estimated coalescent N_e was substantially larger in the Tugur mainstem than the tributaries, while migration probabilities were higher from the tributaries into the mainstem. It is worth noting that parameter estimates for these models can be affected by mutation rate, changes in migration and population size over time, and bioinformatic processing of sequence data^{98,99}, and also that the additional models including migration had similar likelihoods to the best fit model yet very different parameter estimates (Table 3). Importantly, coalescent N_e estimates can be heavily affected by changes in these modeling input values, as seen in our substantially higher coalescent N_e for the best fit model than empirical N_e estimates for taimen⁷². Denser data and sampling would improve our ability to evaluate such models and characterize this history.

A possible, if not likely, scenario underlying divergence in the Tugur basin is that ancestral tributary taimen diverged in isolation in the eastern Amur, before paleohydrological connections allowed colonization of the Tugur basin. Consistent with this hypothesis, taimen in the Tugur tributaries are intermediate with respect to those from the Tugur mainstem and the western Amur basin samples in both PCA and ancestry-based analyses (Figs. 2, 3). Although the Tugur and Amur basins have no contemporary hydrologic connectivity and the hydrologic history of this region is not well understood, faunal similarities are consistent with paleohydrological connections among these drainages during the Pleistocene. For example, mtDNA haplotype sharing among the Amur and Tugur basins has been documented in blunt-nosed lenok (*Brachymystax lenok*), suggesting a history of such connectivity⁷⁷. The location of a potential paleo connection between the Tugur and Nimelen, a lower Amur tributary, is apparent in a low-relief area where active channels in tributaries of the two rivers are separated by < 1 km and by a drainage divide of < 10 m. One hypothesized scenario is that the Tugur was forced south and joined the lower Amur during Pleistocene “back-arc glaciation” of a Sea of Okhotsk marine ice sheet and the Stanovoi glacier complex to the west before drainages reorganized and became independent during glacier retreat^{80,100}. A similar pattern of glaciation has been hypothesized to generate paleohydrological connections between the proto Tugur drainage and mid-Amur drainages of the Bureya and Zeya Rivers (see Fig. 6 of Ref.¹⁰⁰).

Understanding the historical and geographic context of differentiation in the Tugur basin as well as the mechanisms potentially maintaining it will require further study. Our results are based on relatively sparse spatial sampling only in the western Amur basin. More thorough sampling of the eastern Amur basin, especially where it nears the Tugur, will be necessary to more thoroughly characterize the geographical context and origin of divergence among taimen in the Tugur basin. Similarly, additional sampling within the Tugur basin could improve understanding of the spatial distribution of the genetically differentiated groups detected here. As importantly, an understanding of ecological, morphological, and life history variation among taimen in this system will be critical for understanding potential mechanisms underlying and maintaining differentiation. Further work here is warranted as the Tugur River continues to support the largest individuals of the world’s largest salmonid¹⁰¹, and the majority of the watershed is protected within the Tugursky Nature Reserve, a regional zakaznik (equivalent to an IUCN Category IV protected area).

Implications for conservation and management. Given substantial and widespread declines of taimen populations, an understanding of genetic diversity and spatial genetic structure could be critical for informing taimen conservation. The demarcation of taimen management units could become important as anthropogenic influences known to impact fish populations (e.g., see Refs.^{102,103}) increase in frequency and intensity, and as translocation efforts are considered for regions where populations have declined or have been extirpated. Genetic diversity is often considered a key parameter for conservation and restoration, as it is commonly viewed as a proxy for population resilience and evolutionary potential^{104,105}. Nucleotide diversity estimates for all of our sampling sites were in the range of estimates published for other salmonids^{86–89}, and do not reflect severely reduced diversity. Although, Tajima’s D estimates were slightly positive for the Amur basin, which could be consistent with population declines here. However, our sampling was limited to regions with healthy river systems

and those that have yet to exhibit strong population declines, and may not represent standing variation in other areas of the distribution.

Evolutionarily significant units (ESUs^{106,107}) are often used to designate lineages of conservation importance. Kaus et al.⁴⁶ proposed two taimen ESUs across northern Mongolia that largely correspond to our Selenge (Arctic) and Amur (Pacific) basin sites. Our results indicating the presence of substantial genetic divergence among these basins lend support for two separate units that might require separate conservation and management strategies. The hierarchical population structure recovered in this study suggests that ecologically relevant genetic variation might be partitioned at smaller spatial scales than previously considered, and we thus caution against the translocation of taimen among geographically and ecologically distinct populations before more thorough genetic sampling can be completed. The differentiated groups we detected in the Tugur basin could reflect ecological and historical variation warranting unique conservation consideration, though further study is clearly needed. Indeed, the additional evidence for population structure within the Pacific drainages suggests that our understanding of taimen genetic structure across different riverscapes is likely limited by both the geographic and genomic extent of sampling. Due to difficulty in sampling via fly fishing, our sampling was opportunistic, geographically sparse, and less than ideal for quantifying how environmental and hydrological variation may influence spatial genetic structure across the distribution. Future studies with more comprehensive sampling within and across basins could be essential for developing a finer scale understanding of the factors influencing evolutionary history of taimen across Siberia and the rest of its range.

Data availability

The datasets generated for this study are available at the Dryad Digital Repository (doi: <https://doi.org/10.5061/dryad.wstjq2kd>; <https://doi.org/10.5061/dryad.wstjq2kd>) and NCBI's Short Read Archive (accession PRJNA745962; <https://dataview.ncbi.nlm.nih.gov/object/PRJNA745962>).

Received: 23 April 2021; Accepted: 20 September 2021

Published online: 15 October 2021

References

1. Avise, J. C. Perspective: Conservation genetics enters the genomics era. *Conserv. Genet.* **11**, 665–669 (2010).
2. Luikart, G., Ryman, N., Tallmon, D. A., Schwartz, M. K. & Allendorf, F. W. Estimation of census and effective population sizes: The increasing usefulness of DNA-based approaches. *Conserv. Genet.* **11**, 355–373 (2010).
3. Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G. & Hohenlohe, P. A. Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat. Rev. Genet.* **17**, 81–92 (2016).
4. Wagner, C. E. et al. Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Mol. Ecol.* **22**, 787–798 (2013).
5. Vendrami, D. L. J. et al. RAD sequencing resolves fine-scale population structure in a benthic invertebrate: Implications for understanding phenotypic plasticity. *R. Soc. Open Sci.* **4**, 160548 (2017).
6. Xuereb, A. et al. Asymmetric oceanographic processes mediate connectivity and population genetic structure, as revealed by RAD seq, in a highly dispersive marine invertebrate (*Parastichopus californicus*). *Mol. Ecol.* **27**, 2347–2364 (2018).
7. DeSaix, M. G. et al. Population assignment reveals low migratory connectivity in a weakly structured songbird. *Mol. Ecol.* **28**, 2122–2135 (2019).
8. Hohenlohe, P. A., Amish, S. J., Catchen, J. M., Allendorf, F. W. & Luikart, G. Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Mol. Ecol. Resour.* **11**, 117–122 (2011).
9. Mandeville, E. G. et al. Inconsistent reproductive isolation revealed by interactions between *Catostomus* fish species. *Evol. Lett.* **1**, 255–268 (2017).
10. Bay, R. A. et al. Genomic signals of selection predict climate-driven population declines in a migratory bird. *Science* **359**, 83–86 (2018).
11. Razgour, O. et al. Considering adaptive genetic variation in climate change vulnerability assessment reduces species range loss projections. *Proc. Natl. Acad. Sci.* **116**, 10418–10423 (2019).
12. Larson, W. A. et al. Genotyping by sequencing resolves shallow population structure to inform conservation of Chinook salmon (*Oncorhynchus tshawytscha*). *Evol. Appl.* **7**, 355–369 (2014).
13. Bernos, T. A., Jeffries, K. M. & Mandrak, N. E. Linking genomics and fish conservation decision making: A review. *Rev. Fish Biol. Fish.* **30**, 587–604 (2020).
14. Montgomery, D. R. *King of Fish: The Thousand-Year Run of Salmon* (Westview Press, 2003).
15. Costello, M. J. How sea lice from salmon farms may cause wild salmonid declines in Europe and North America and be a threat to fishes elsewhere. *Proc. R. Soc. B* **276**, 3385–3394 (2009).
16. Sepulveda, A. J., Rutz, D. S., Dupuis, A. W., Shields, P. A. & Dunker, K. J. Introduced northern pike consumption of salmonids in Southcentral Alaska. *Ecol. Freshw. Fish* **24**, 519–531 (2015).
17. Otero, J. et al. Basin-scale phenology and effects of climate variability on global timing of initial seaward migration of Atlantic salmon (*Salmo salar*). *Glob. Change Biol.* **20**, 61–75 (2014).
18. Clavero, M. et al. Historical citizen science to understand and predict climate-driven trout decline. *Proc. R. Soc. B* **284**, 20161979 (2017).
19. Utter, F. Population genetics, conservation and evolution in salmonids and other widely cultured fishes: Some perspectives over six decades. *J. Fish Biol.* **65**, 323–324 (2004).
20. Bernatchez, L. et al. Harnessing the power of genomics to secure the future of seafood. *Trends Ecol. Evol.* **32**, 665–680 (2017).
21. Ryman, N. Conservation genetics considerations in fishery management. *J. Fish Biol.* **39**, 211–224 (1991).
22. Hohenlohe, P. A. et al. Genomic patterns of introgression in rainbow and westslope cutthroat trout illuminated by overlapping paired-end RAD sequencing. *Mol. Ecol.* **22**, 3002–3013 (2013).
23. Bay, R. A., Taylor, E. B. & Schluter, D. Parallel introgression and selection on introduced alleles in a native species. *Mol. Ecol.* **28**, 2802–2813 (2019).
24. Briec, M. S. O., Ono, K., Drinan, D. P. & Naish, K. A. Integration of Random Forest with population-based outlier analyses provides insight on the genomic basis and evolution of run timing in Chinook salmon (*Oncorhynchus tshawytscha*). *Mol. Ecol.* **24**, 2729–2746 (2015).
25. Hecht, B. C., Matala, A. P., Hess, J. E. & Narum, S. R. Environmental adaptation in Chinook salmon (*Oncorhynchus tshawytscha*) throughout their North American range. *Mol. Ecol.* **24**, 5573–5595 (2015).

26. Prince, D. J. *et al.* The evolutionary basis of premature migration in Pacific salmon highlights the utility of genomics for informing conservation. *Sci. Adv.* **3**, e1603198 (2017).
27. Habicht, C. *et al.* *Harvest and Harvest Rates of Sockeye Salmon Stocks in Fisheries of the Western Alaska Salmon Stock Identification Program (WASSIP), 2006–2008* (Alaska Department of Fish and Game, Division of Sport Fish, Research and Technical Services, 2012).
28. Ocock, J. *et al.* *Mongolian Red List of Fishes*. Volume 3 (2006).
29. Hofmann, J. *et al.* Initial characterization and water quality assessment of stream landscapes in northern Mongolia. *Water* **7**, 3166–3205 (2015).
30. Mercado-Silva, N. *et al.* Fish community composition and habitat use in the Eg-Uur River system, Mongolia. *Mong. J. Biol. Sci.* **6**, 21–30 (2008).
31. Jensen, O. P. *et al.* Evaluating recreational fisheries for an endangered species: A case study of taimen, *Hucho taimen*, Mongolia. *Can. J. Fish. Aquat. Sci.* **66**, 1707–1718 (2009).
32. Crête-Lafrenière, A., Weir, L. K. & Bernatchez, L. Framing the Salmonidae family phylogenetic portrait: A more complete picture from increased taxon sampling. *PLoS ONE* **7**, e46662 (2012).
33. Shedko, S. V., Miroshnichenko, I. L. & Nemkova, G. A. Phylogeny of salmonids (Salmoniformes: Salmonidae) and its molecular dating: Analysis of nuclear RAG1 gene. *Russ. J. Genet.* **48**, 575–579 (2012).
34. Holčík, J., Hensel, K., Nieslanik, J. & Skacel, L. *The Eurasian Huchen, Hucho Hucho: Largest Salmon of the World* Vol. 5 (Dr. W. Junk Publishers, 1988).
35. Gilroy, D. J. *et al.* Home range and seasonal movement of taimen, *Hucho taimen*, Mongolia. *Ecol. Freshw. Fish* **19**, 545–554 (2010).
36. Esteve, M. Observations of spawning behaviour in Salmoninae: *Salmo*, *Oncorhynchus* and *Salvelinus*. *Rev. Fish Biol. Fish.* **15**, 1–21 (2005).
37. Esteve, M., Gilroy, D. & McLennan, D. A. Spawning behaviour of taimen (*Hucho taimen*) from the Uur River, Northern Mongolia. *Environ. Biol. Fishes* **84**, 185–189 (2009).
38. He, F. *et al.* The global decline of freshwater megafauna. *Glob. Change Biol.* **25**, 3883–3892 (2019).
39. Hogan, Z. & Jensen, O. *Hucho taimen*. *The IUCN Red List of Threatened Species*. 2013: e.T188631A22605180.
40. Rand, P. S. Current global status of taimen and the need to implement aggressive conservation measures to avoid population and species-level extinction. *Arch. Pol. Fish.* **21**, 119–128 (2013).
41. Froufe, E., Alekseyev, S., Knizhin, I. & Weiss, S. Comparative mtDNA sequence (control region, ATPase 6 and NADH-1) divergence in *Hucho taimen* (Pallas) across four Siberian river basins. *J. Fish Biol.* **67**, 1040–1053 (2005).
42. Tong, G., Kuang, Y., Yin, J., Liang, L. & Sun, X. Isolation of microsatellite DNA and analysis on genetic diversity of endangered fish, *Hucho taimen* (Pallas). *Mol. Ecol. Notes* **6**, 1099–1101 (2006).
43. Kuang, Y., Tong, G., Xu, W., Sun, X. & Yin, J. Analysis on population genetic structure of taimen (*Hucho taimen*) in the Heilongjiang River. *J. Fish. Sci.* **6**, 1208–1217 (2010).
44. Balakirev, E. S., Romanov, N. S., Mikheev, P. B. & Ayala, F. J. Mitochondrial DNA variation and introgression in Siberian taimen *Hucho taimen*. *PLoS ONE* **8**, e71147 (2013).
45. Marić, S. *et al.* First mtDNA sequencing of Volga and Ob basin taimen *Hucho taimen*: European populations stem from a late Pleistocene expansion of *H. taimen* out of western Siberia and are not intermediate to *Hucho hucho*. *J. Fish Biol.* **85**, 530–539 (2014).
46. Kaus, A. *et al.* Fish conservation in the land of steppe and sky: Evolutionarily significant units of threatened salmonid species in Mongolia mirror major river basins. *Ecol. Evol.* **9**, 3415–3433 (2019).
47. Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S. & Hoekstra, H. E. Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE* **7**, e37135 (2012).
48. Parchman, T. L. *et al.* Genome-wide association genetics of an adaptive trait in lodgepole pine. *Mol. Ecol.* **21**, 2991–3005 (2012).
49. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357 (2012).
50. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
51. Puritz, J. B., Hollenbeck, C. M. & Gold, J. R. *dDocent*; a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ* **2**, e431 (2014).
52. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
53. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
54. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
55. McKinney, G. J., Waples, R. K., Seeb, L. W. & Seeb, J. E. Paralogs are revealed by proportion of heterozygotes and deviations in read ratios in genotyping-by-sequencing data from natural populations. *Mol. Ecol. Resour.* **17**, 656–669 (2017).
56. Gompert, Z. *et al.* Admixture and the organization of genetic diversity in a butterfly species complex revealed through common and rare genetic variants. *Mol. Ecol.* **23**, 4555–4573 (2014).
57. Shastry, V. *et al.* Model-based genotype and ancestry estimation for potential hybrids with mixed-ploidy. *Mol. Ecol. Res.* **21**, 1434–1451 (2021).
58. Falush, D., Stephens, M. & Pritchard, J. K. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* **164**, 1567–1587 (2003).
59. Buerkle, C. A. & Gompert, Z. Population genomics based on low coverage sequencing: How low should we go?. *Mol. Ecol.* **22**, 3028–3035 (2013).
60. Fumagalli, M. *et al.* Quantifying population genetic differentiation from next-generation sequencing data. *Genetics* **195**, 979–992 (2013).
61. R Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, <https://www.R-project.org>, 2020).
62. Hudson, R. R., Slatkin, M. & Maddison, W. P. Estimation of levels of gene flow from DNA sequence data. *Genetics* **132**, 583–589 (1992).
63. Nei, M. Genetic distance between populations. *Am. Nat.* **106**, 192–283 (1972).
64. Tajima, F. Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**, 437–460 (1983).
65. Watterson, G. A. On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **7**, 256–276 (1975).
66. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595 (1989).
67. Korneliussen, T. S., Moltke, I., Albrechtsen, A. & Nielsen, R. Calculation of Tajima's D and other neutrality test statistics from low depth next-generation sequencing data. *BMC Bioinform.* **14**, 289 (2013).
68. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of next generation sequencing data. *BMC Bioinform.* **15**, 356 (2014).
69. Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C. & Foll, M. Robust demographic inference from genomic and SNP data. *PLoS Genet.* **9**, e1003905 (2013).
70. Beichman, A. C., Huerta-Sánchez, E. & Lohmueller, K. E. Using genomic data to infer historic population dynamics of nonmodel organisms. *Annu. Rev. Ecol. Evol. Syst.* **49**, 433–456 (2018).

71. Rougemont, Q. *et al.* Demographic history shaped geographical patterns of deleterious mutation load in a broadly distributed Pacific Salmon. *PLoS Genet.* **16**, e1008348 (2020).
72. Kul'bachnyi, S. E. & Kul'bachnaya, A. V. Some features of biology of the Siberian Taimen *Hucho taimen* (Salmonidae) from the Tugur River basin. *J. Ichthyol.* **58**, 765–768 (2018).
73. Stamatakis, A. *The RAxML v8.2. X Manual* (Heidleberg Institute for Theoretical Studies, 2016).
74. Petit, C. & Deverchere, J. Structure and evolution of the Baikal rift: A synthesis. *Geochem. Geophys. Geosyst.* **7**, Q11016 (2006).
75. Koskinen, M. T., Knizhin, I., Primmer, C. R., Schlotterer, C. & Weiss, S. Mitochondrial and nuclear DNA phylogeography of *Thymallus spp.* (grayling) provides evidence of ice-age mediated environmental perturbations in the world's oldest body of fresh water, Lake Baikal. *Mol. Ecol.* **11**, 2599–2611 (2002).
76. Froufe, E., Knizhin, I. & Weiss, S. Phylogenetic analysis of the genus *Thymallus* (grayling) based on mtDNA control region and ATPase 6 genes, with inferences on control region constraints and broad-scale Eurasian phylogeography. *Mol. Phylogenet. Evol.* **34**, 106–117 (2005).
77. Froufe, E., Alekseyev, S., Alexandrino, P. & Weiss, S. The evolutionary history of sharp- and blunt-snouted lenok (*Brachymystax lenok* (Pallas, 1773)) and its implications for the paleo-hydrological history of Siberia. *BMC Evol. Biol.* **8**, 1–18 (2008).
78. Yokoyama, R., Sideleva, V. G., Shedko, S. V. & Goto, A. Broad-scale phylogeography of the Palearctic freshwater fish *Cottus poecilopus* complex (Pisces: Cottidae). *Mol. Phylogenet. Evol.* **48**, 1244–1251 (2008).
79. Skog, A., Vøllestad, L. A., Stenseth, N. C., Kasumyan, A. & Jakobsen, K. S. Circumpolar phylogeography of the northern pike (*Esox lucius*) and its relationship to the Amur pike (*E. reichertii*). *Front. Zool.* **11**, 67 (2014).
80. Grosswald, M. G. Late-Weichselian ice sheets in arctic and Pacific Siberia. *Quat. Int.* **45**, 3–18 (1998).
81. Karabanov, E. B., Prokopenko, A. A., Williams, D. F. & Colman, S. M. Evidence from Lake Baikal for Siberian glaciation during oxygen-isotope substage 5d. *Quat. Res.* **50**, 46–55 (1998).
82. Mangerud, J. *et al.* Ice-dammed lakes and rerouting of the drainage of northern Eurasia during the Last Glaciation. *Quat. Sci. Rev.* **23**, 1313–1332 (2004).
83. Simonov, E. A. & Dahmer, T. D. *Amur-Heilong River Basin Reader* (Ecosystems, 2008).
84. Hewitt, G. M. Some genetic consequences of ice ages, and their role in divergence and speciation. *Biol. J. Linn. Soc.* **58**, 247–276 (1996).
85. Bernatchez, L. & Wilson, C. C. Comparative phylogeography of nearctic and palearctic fishes. *Mol. Ecol.* **7**, 431–452 (1998).
86. Ferchaud, A.-L., Laporte, M., Perrier, C. & Bernatchez, L. Impact of supplementation on deleterious mutation distribution in an exploited salmonid. *Evol. Appl.* **11**, 1053–1065 (2018).
87. Kijas, J. *et al.* Evolution of sex determination loci in Atlantic salmon. *Sci. Rep.* **8**, 1–11 (2018).
88. Leitwein, M. *et al.* Genome-wide nucleotide diversity of hatchery-reared Atlantic and Mediterranean strains of brown trout *Salmo trutta* compared to wild Mediterranean populations. *J. Fish Biol.* **89**, 2717–2734 (2016).
89. Matthaues, W. J. *Investigation of Fine Spatial Scale Population Genetic Structure in Two Alaskan Salmonids*. (Doctoral dissertation, 2016).
90. Ryyänen, H. J. & Primmer, C. R. Single nucleotide polymorphism (SNP) discovery in duplicated genomes: Intron-primed exon-crossing (IPEC) as a strategy for avoiding amplification of duplicated loci in Atlantic salmon (*Salmo salar*) and other salmonid fishes. *BMC Genom.* **7**, 1–11 (2006).
91. Banarescu, P. *Zoogeography of Fresh Waters. Volume 2: Distribution and Dispersal of Freshwater Animals in North America and Eurasia* 519–1091 (Aula-Verlag, 1991).
92. Kaus, A. *et al.* Seasonal home range shifts of the Siberian taimen (*Hucho taimen* Pallas 1773): Evidence from passive acoustic telemetry in the Onon River and Balj tributary (Amur River basin, Mongolia). *Int. Rev. Hydrobiol.* **101**, 147–159 (2016).
93. Kul'bachnyi, S. E. & Ivankov, V. N. Temporal differentiation and conditions of reproduction of chum salmon *Oncorhynchus keta* (Salmoniformes, Salmonidae) from the Tugur River basin (Khabarovsk Krai). *J. Ichthyol.* **51**, 63–72 (2011).
94. Bernatchez, L. *et al.* On the origin of species: Insights from the ecological genomics of lake whitefish. *Philos. Trans. R. Soc. B Biol. Sci.* **365**, 1783–1800 (2010).
95. Gordeeva, N. V., Alekseyev, S. S., Matveev, A. N. & Samusenok, V. P. Parallel evolutionary divergence in Arctic char *Salvelinus alpinus* complex from Transbaikalia: Variation in differentiation degree and segregation of genetic diversity among sympatric forms. *Can. J. Fish. Aquat. Sci.* **72**, 96–115 (2015).
96. Olson, K. W., Krabbenhoft, T. J., Hrabik, T. R., Mendsaikhan, B. & Jensen, O. P. Pelagic-littoral resource polymorphism in Hovsgol grayling *Thymallus nigrescens* from Lake Hovsgol, Mongolia. *Ecol. Freshw. Fish* **28**, 411–423 (2019).
97. Zhivotovsky, L. A. *et al.* Eco-geographic units, population hierarchy, and a two-level conservation strategy with reference to a critically endangered salmonid, Sakhalin taimen *Parahucho perryi*. *Conserv. Genet.* **16**, 431–441 (2015).
98. Han, E., Sinsheimer, J. S. & Novembre, J. Characterizing bias in population genetic inferences from low-coverage sequencing data. *Mol. Biol. Evol.* **31**, 723–735 (2014).
99. Shafer, A. B. A. *et al.* Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. *Methods Ecol. Evol.* **8**, 907–917 (2017).
100. Grosswald, M. G. & Hughes, T. J. 'Back-arc' marine ice sheet in the Sea of Okhotsk. *Russ. J. Earth Sci.* **7**, ES5004 (2005).
101. Zolotukhin, S. F. & Shcherbovich, I. V. Maximum weight of the Siberian *Hucho taimen* (Pallas) in the area. *Fish. Russia* **1**, 47–51 (2021).
102. Reid, S. M., Wilson, C. C., Mandrak, N. E. & Carl, L. M. Population structure and genetic diversity of black redhorse (*Moxostoma duquesnei*) in a highly fragmented watershed. *Conserv. Genet.* **9**, 531 (2008).
103. Wu, H. *et al.* Effects of dam construction on biodiversity: A review. *J. Clean. Prod.* **221**, 480–489 (2019).
104. Allendorf, F. W. & Luikart, G. *Conservation and the Genetics of Populations* (Wiley, 2009).
105. Weeks, A. R. *et al.* Assessing the benefits and risks of translocations in changing environments: A genetic perspective. *Evol. Appl.* **4**, 709–725 (2011).
106. Ryder, O. A. Species conservation and systematics: The dilemma of subspecies. *Trends Ecol. Evol.* **1**, 9–10 (1986).
107. Moritz, C. Defining 'evolutionarily significant units' for conservation. *Trends Ecol. Evol.* **9**, 373–375 (1994).

Acknowledgements

This research was funded by The Taimen Fund (previously The Tributary Fund), the Wild Salmon Center, and the University of Nevada, Reno. We thank The Taimen Fund and the Taimen Conservation Fund for accessibility and facilitation of cooperative sampling across fly fishing outfitters. We thank the staff and fishing guides of Konin Lodge, Sweetwater Travel, Hovsgol Travel, and Mongolia River Outfitters for substantial aid in logistics of lodging, sampling, and local permitting. We thank Iain McClaren for his aid in acquiring samples from the Tugur River basin. We thank the Global Water Center Science Team and the University of Nevada, Reno, for support in sampling and outreach. Finally, we thank Michael R. Miller from the University of California, Davis, for helpful discussion and guidance on demographic inference.

Author contributions

S.C., Z.H., and J.B.S. identified the need for this research and secured funding. L.M.G., J.B.S., S.C., and Z.H. collected specimens. L.M.G. extracted DNA, generated ddRADseq libraries, and conducted population genetic analyses. A.R.L.-N. executed demographic inference modeling. M.R.S. contributed natural history of taimen and geological history of the Tugur River basin and helped interpret demographic modeling results. L.M.G. and T.L.P. wrote the manuscript, with substantial input from J.P.J. All authors contributed to the final revisions of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-99530-3>.

Correspondence and requests for materials should be addressed to L.M.G.

Reprints and permissions information is available at www.nature.com/reprints.

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021