Chromosome-level genome assembly for Sichuan taimen (*hucho bleekeri*) reveals the extraordinary tandem repeat proportions and its persistent population shrinkage

Xinmiao Zhang[1], Dongmei Xiong[1], Shenglong Jian[2], Yu Jiang[3*], Lixin Wang[1*]

[1]College of Animal Science and Technology, Northwest A&F University, Yangling, 712100, China

[2]Qinghai Provincial Fishery Technology Extension Center, Xining, China

[3]Center for Ruminant Genetics and Evolution, College of Animal Science and Technology, Northwest A&F University, Yangling 712100, China

*Correspondence authors. E-mail: fisherwanglx@nwsuaf.edu.cn (Lixin Wnag) and yu.jiang@nwafu.edu.cn (Yu Jiang).

ABSTRACT:

Salmonid fishes are globally renowned and valuable, with most members of the Salmonidae family living in seawater and exhibiting migratory behavior. In contrast, huchonid fishes (*Hucho* spp.and *Brachymystax* spp.), an ancient evolutionary branch within Salmoninae, are entirely landlocked. The Sichuan taimen (*Hucho bleekeri* Kimura) is a critically endangered fish that has attracted widespread concern and is one of eleven national first-level protected fishes in China. However, genome resources for all *Hucho* spp., including *H. bleekeri*, are scarce, and the genomic characteristics of this ancient evolutionary lineage remain unclear, hindering conservation biology efforts. Here, we present the first chromosome-level genome for the Sichuan taimen, with a final genome size of approximately 3.45 Gb across 44 chromosomes. The Sichuan taimen genome contains 44.15% tandem repetitive sequences, exceeding those of all sequenced salmonid fishes. We also identified 44 Ss4R homeologous block pairs in the assembled genome. Genome synteny analysis suggested a ~5 fold tandem repeat expansion in the Sichuan taimen compared to the Salmonidae ancestor Northern pike. Phylogenetic analysis estimated the divergence time between huchonid and other Salmoninae fishes at approximately 33.29 million years ago (Mya). The divergence time between Siberia taimen and Sichuan taimen was estimated at around 2.29 Mya, with their effective population size declining from around 1 Mya. The genomic resource provided in this article will promote the protection of the Sichuan taimen and evolutionary genetics studies of salmonids.

Keywords: *Hucho bleekeri*; Genome assembly; Tandem repeats; Salmonid

INTRODUCTION

The Sichuan taimen (*Hucho bleekeri* Kimura, 1934) belonging to the subfamily Salmoninae, is a glacial remnant species and endemic to China (Wang et al., 2016). It has the lowest distribution latitude among all taimen species, and is primarily found in the upper tributaries of the Yangtze River and Wei River (Hu et al., 2008). Sichuan taimen, characterized by their large body size and predatory behavior, occupy the top of the aquatic food chain (Wang et al., 2016). However, their late sexual maturation makes them particularly vulnerable to human activities and ecological changes (Allendorf & Hard, 2009; Sadovy et al., 2013). Individuals of Sichuan taimen are found sporadically and scattered in their distribution, and have been listed in China red data book of endangered animals since 1998. Despite this, they continues to be illegally captured by poachers for their high nutritional value and taste (Sung et al., 1998). Due to the continuous decline of the population size, the Sichuan taimen was listed as critically endangered (CR) level in the International Union for Conservation of Nature (IUCN) Red List of Threatened Species in 2012 (Song, 2012). As of 2021, the Sichuan taimen is under the highest level of protection from the Chinese government, and is the only salmonid fish with national protected animal in "class I".

Genomic data provide significant opportunities for advancements in the conservation and evolutionary research of threatened animals (Allendorf et al., 2010; Kohn et al., 2006). As sequencing costs have decreased, the genomes of many endangered fish species have been published, including *Trachidermus fasciatus* Heckel (Xue et al., 2022), *Myxocyprinus asiaticus* (Liu et al., 2022), and *Brachymystax tsinlingensis* (Zhu

et al., 2022), etc. Based on genome data, scientific and rational conservation policies have been formulated, enhancing the protection benefits of these endangered animals (Brandies et al., 2019; Waples et al., 2020). However, the lack of genome data for *Hucho bleekeri* creates a gap in nuclear-based phylogenetic research of Salmonidae and affects genomic studies, such as taxonomic studies based on SNPs (Chen et al., 2020; Zhang et al., 2020) and researches on transcriptome (Chen et al., 2021).

Due to the whole genome duplication and lineage-specific rediploidization, the genome of salmonids have became complex and variable, increasing the difficulty of assembly (Dysin et al., 2022; Robertson et al., 2017). The circular consensus sequencing (CCS) mode in single-molecule real-time (SMRT) sequencing by Pacific Biosciences can produce long and high fidelity (HiFi) reads, which facilitates the genome assembly process and improves quality (Wenger et al., 2019). In this research, we utilized HiFi and Hi-C sequencing technology to construct a high-quality genome for Sichuan taimen. From the genome we identified a high proportion of tandem repetitive sequences. Additionally, we analyzed the phylogenetic relationships of salmonid fishes and the divergence history of *H. bleekeri* and *H. taimen*, another species found in Heilongjiang. Our work provides insights into the genome evolution of salmonid fishes and the conservation of the endangered Sichuan taimen.

MATERIALS AND METHODS

Sample collection and sequencing

Two wild individuals of *H. bleekeri* were collected from Taibai river (Baoji, Shaanxi province), one for genome assembly the other for resequencing. Additionally, a

78 cultured individual of *H. taimen* was collected from Heilongjiang. All the experiments
79 in this study were performed in accordance with the guidelines of China Council on
80 Animal Care, and approved by the Experimental Animal Ethics Committee of
81 Northwest A&F University (NWAFU-314020038).
82 Genomic DNA was extracted from kidney tissue with cetyltrimethylammonium
83 bromide (CATB) method (Stefanova et al., 2013). The reads for genome size
84 estimation were generated on Illumina NovaSeq-6000 platform after the constructed
85 of 150 bp library. For genome assembly, a 15 kb library preparation was performed by
86 SMRTbell Express Template Prep Kit 2.0 (Pacific Biosciences, CA, USA), and
87 sequenced on the PacBio Sequel II platform (Pacific Biosciences, CA, USA) with two
88 cells of single molecule real time sequencing (SMRT). The CCS software v.6.0.0
89 (https://github.com/PacificBiosciences/ccs) was performed to integrate HiFi reads
90 with the parameters "--min-passes 3 --min-length 10 --min-rq 0.99". The Hi-C library
91 was constructed and sequenced on Illumina NovaSeq-6000 platform. The RNA
92 samples were extracted from fin， brain, gill, gonad, and eye, then sequenced on
93 Illumina NovaSeq-6000 platform. The two wild samples for demographic history
94 analysis were sequenced on DNB-seq platform (BGI).

95 **Genome size estimation**

96 The raw Illumina paired-end reads of Sichuan taimen were filtered using Fastp
97 version 0.20.0 (Chen et al., 2018) with default parameters. The genome size
98 estimation of Sichuan taimen was performed by GCE software version 1.0.2 (Liu et
99 al., 2013). A total of 58.89 Gb clean reads were used to calculate the distribution of

100 k-mer frequency using kmerfreq with k=17. Then the gce software was ran in both
101 homozygous and heterzygous mode to evaluate the genome size and heterozygosity.
102 The genome heterozygous was calculated using the formula "a[1/2]/k".

103 **Chromosome level genome assembly**

104 The PacBio HiFi reads were assembled using Hifiasm version 0.13-r308 (Cheng et al.,
105 2021) to obtain *de novo* contigs. After quality control by Fastp (Chen et al., 2018), the
106 clean Hi-C reads were used to integrate the contigs. The Hi-C interaction matrix was
107 generated using Juicer (version 1.6)(Durand et al., 2016b)scripts with -s MboI
108 parameter, then, the 3D-DNA (version 201008)(Dudchenko et al., 2017)with 0 rounds
109 was used to link the contigs into chromosome level based on the Hi-C interaction
110 matrix. The initial assembly was manually corrected according to the Hi-C heatmap
111 using Juicebox version 1.11.08 (Durand et al., 2016a) and run the 3D-DNA again to
112 generate the final assembly. The BUSCO (Benchmarking universal single-copy
113 orthologues) version 3.0.2 (Simão et al., 2015) was used to assess the accuracy and
114 integrity of the genome with the gene database of ray-finned fishes
115 (actinopterygii_odb9).

116 **Repeat sequence detection and gene annotation**

117 To recognize repeat sequences of the Sichuan taimen genome, we build a *de novo*
118 library of transposable elements (TEs). First, the RepeatModeler version 2.1 (Flynn et
119 al., 2020) with default parameters was ran to constructed a repeat library; then an
120 additional long terminal repeat (LTR) library of the Sichuan taimen was built by
121 LTR_FINDER version 1.07 (Xu & Wang, 2007) and LTR_RETRIEVER version 2.9.6

122 (Ou & Jiang, 2018); third the final repeat library was integrated by replacing the LTR
123 elements of the first repeat library with the LTR library. To further classify the
124 "Unknown" sequence, the BLAST version 2.14.0 (Ye et al., 2006) was performed
125 between the final repeat library and public repeat library
126 (http://www.hrt.msu.edu/uploads/535/78637/Tpases020812.gz).
127 The TEs of the genome were described and masked by RepeatMasker version 4.1.2
128 (Tarailo-Graovac & Chen, 2009) with parameters "-xsmall -s -cutoff 255 -frag 20000
129 –gff -engine ncbi". The tandem repeats (TRs) were detected through Tandem Repeats
130 Finder (TRF) version 4.09 (Benson, 1999) with parameters "2 7 7 80 10 50 500 -f -d
131 -m". The TRs were further divided into short tandem repeats (STRs) and variable
132 number tandem repeats (VNTRs) according to the length of repeat unit (STR: unit
133 number at 1 to 6; VNTR: unit number more than 6).
134 The gene annotation of the Sichuan taimen genome was carried out using BRAKER2
135 pipeline version 2.1.5 (Hoff et al., 2016; Hoff et al., 2019). For transcriptome based
136 prediction, the RNAs were mapping to the assembled genome using STAR version
137 2.7.1a (Dobin et al., 2013) with default parameters, then the RNA sequence data were
138 used to predict genes through auto BRAKER2 pipeline. For homology based
139 prediction, the protein sequence of six species including *Lepisosteus oculatus*
140 (GCF_000242695.1)(Braasch et al., 2016), *Esox Lucius* (GCF_011004845.1)(Ishiguro
141 et al., 2003), *Oncorhynchus mykiss* (GCF_013265735.2)(Gao et al., 2021),
142 *Oncorhynchus nerka* (GCF_006149115.2) , *Salmo salar* (GCF_000233375.1)(Lien et
143 al., 2016), and *Brachymystax tsinlingensis* (GCA_031305455.1) which represent the

144 sister clade or the ancestor species of taimen were download, and predicted through
145 BRAKER2 pipeline. The results from RNA and protein-based annotation were
146 combined using TSEBRA version 1.0.3 (Gabriel et al., 2021) with default parameters
147 to generate the final gene prediction file. The gene expression level was measured by
148 transcript per million (TPM) which were calculated using Kallisto (Bray et al., 2016)
149 with default parameters. To remove the pseudogenes caused by tandem repetitive
150 sequence, the gene set was filtered based on the following criteria: 1) contain only one
151 exon; 2) TPM<1; 3) no matched BLAST results with *B. tsinlingensis*.

152 **Chromosome synteny identification**

153 Homology blocks were reconstructed using whole-genome duplicated genes.
154 Diamond (v0.8.24.86) with default parameter was used to align against the predicted
155 genes. Then selected paired genes with highest identity and identity ≥ 75% from the
156 Diamond results. Finally, one-to-one paralogs were extracted and conducted manual
157 linkage. The *Esox lucius* genome was download from NCBI (GCF_011004845.1).

158 **Phylogenetic analysis**

159 Protein sequences of 10 species *Esox lucius*, *Coregonus* sp. *'balchen'*
160 (GCA_902810595.1)(Christensen et al., 2020), *Thymallus thymallus*
161 (GCA_023634145.1), *Oncorhynchus mykiss*, *Oncorhynchus nerka*, *Salvelinus*
162 *namaycush* (GCF_016432855.1)(Smith et al., 2022), *Salmo salar*, *Salmo trutta*
163 (GCF_901001165.1)(Hansen et al., 2021), *Hucho bleekeri*, and *Brachymystax*
164 *tsinlingensis* were used to construct phylogenetic tree. The single copy orthologous
165 proteins among the 10 species were recognized by Orthfinder version 2.3.12 (Emms

166 & Kelly, 2019) with "msa" method. After alignment using MUSCLE version 3.8.31
167 (Edgar, 2004), the species tree was calculated through RAxML version 8.2.12
168 (Stamatakis, 2014) with 1000 bootstraps. The expansions and contractions of gene
169 family was estimated using CAFE5 (Mendes et al., 2020).
170 The divergence time among all these species were calculated using MCMCtree
171 program of PAML software version 4.9d (Yang, 2007). Four fossil records were
172 downloaded from TimeTree (Kumar et al., 2022) and calibrated the results, the
173 Markov chain Monte Carlo (MCMC) was sampled with every 1,000 step and conduct
174 10,000 samples, the pre-burnin number was set at 20,000.

175 **Divergence time estimation of two taimen species**

176 The raw resequencing data of *H. taimen* was mapping to the assembled genome using
177 BWA-MEM version 0.7.17 after quality control. The whole genome sequence for *H.*
178 *taimen* was generated using angsd version 0.918. To obtain the gene annotation
179 information of *H. taimen*, the liftoff (v1.6.3) was operated with default parameters,
180 and further removed genes with a proportion of unknown nucleotides above 1%. Then,
181 four species were selected to generated 1:1 single copy genes using blastn version
182 2.2.31+, including *H. taimen*, *H. bleekeri*, *B. tsinlingensis*, and the outgroup *Salmo*
183 *salar*. After the construction of phylogenetic tree using the methods described above,
184 we estimated the divergence time using BEAST version 2.6.60, the nucleotide
185 substitution model F81 was predicted by jmodeltest version 2.1.10, the chain length
186 was set to 50,000,000 and pre-burnin 10%.

187 **Demographic history estimation**

188 The pairwise sequentially Markovian coalescent (PSMC) method version 0.6.5-r67
189 (Li & Durbin, 2011) was utilized to predict the effective population size. Besides the
190 two wild taimen sample, we used two lenok samples with highest sequencing depths
191 from Heilongjiang River (*B.tumernsis*) and Qinling Mountains (*B. tsinlingensis*). Due
192 to the different sequencing depth, we used different strategies to generate the diploid
193 consensus sequences for taimen and lenok, -D 60 for taimen and -D 20 for lenok at
194 vcfutils.pl packages of bcftools version 1.12 (Danecek et al., 2021). The psmc was run
195 with parameters "-N25 -t30 -r4 -p 8+4+25*2+4+6" to control the divergence at left of
196 the results.

197 **RESULTS**

198 **Genome size estimation and assembly**

199 A total of 426,685,475 paired-end reads were obtained from the Illumina
200 NovaSeq-6000 platform, yielding a total length of 64.00 Gb. The genome size
201 estimated by GCE (k=17) was ~3.21 Gb with a low heterozygous ratio of 0.44%
202 (Supplementary Figure S1; Supplementary Table S1). Based on the estimated genome
203 size, the PacBio HiFi and Illumina Hi-C sequencing produced 76.48 Gb (23.8×) and
204 350.63 Gb (109.23×) of raw reads for genome assembly, respectively. Besides, 155.88
205 Gb of resequencing data from two additional wild samples (~22 ×) were generated.
206 After *de novo* genome assembly using the HiFi reads, a 3.48 Gb primary genome was
207 generated, which included 4,373 high quality contigs (>2 Kb) with a contig N50 of
208 2.4 Mb (Table 1). Following Hi-C assisted genome assembly and scaffold

209 deduplication, the contigs were anchored into 44 chromosome-level scaffolds,
210 covering 95.26% of the total scaffolds length, with a scaffold N50 of 73.71 Mb
211 (Figure 1A; Supplementary Figure S2; Table 1). The final Sichuan taimen genome,
212 closely matching the estimated size, had a total size of 3.45 Gb with a GC content of
213 41.99% (Figure 1A, B; Table 1). The BUSCO results showed that the assembled
214 genome included 95.3% complete genes from the actinopterygii_odb9 database,
215 demonstrating high completeness (Supplementary Table S2).

216 **Landscape of repeat sequences**

217 The non-redundancy length of the total repeat sequences recognized by three
218 annotation methods was 2.60 Gb, covering 74.21% of the assembled genome, which
219 represents the highest proportion of repetitive sequences among salmonids (Figure 1C;
220 Figure 2; Supplementary Table S3, S4 and S5). The interspersed repeats reported by
221 the Repeatmasker and LTR_FINDER composed 36.18% of the genome, the lowest
222 proportion among salmonids; however, the total length of interspersed repeats in the
223 Sichuan taimen genome was similar to that of other salmonids (Supplementary Table
224 S3 and S4). The proportion of retroelements in the Sichuan taimen genome was
225 13.11%, with 7.03% LINEs and 5.98% LTR elements (Supplementary Table S3).
226 DNA transposons composed 9.93% of the genome, with the Tc1-IS630-Pogo family
227 being the most abundant TE, similar to other salmonids, covering 9.38% of the
228 Sichuan taimen genome (Supplementary Table S3).

229 Tandem repetitive sequences covered 44.15% of the Sichuan taimen genome,
230 representing highest proportion among salmonids (Figure 2A; Supplementary Table

231 S5). The lengths of STRs and VNTRs were 45,561,499 bp and 1,553,872,291,
232 accounting for 1.30% and 44.40% of the genome, respectively. Additionally, 59.12%
233 of STRs were less than 1 Kb away from the surrounding VNTRs (Figure 2A). The
234 distribution of TR was not uniform. We found that 29.44% of the regions in the
235 Sichuan taimen genome had low TR ratios (One-tenth of the genome-wide scale;
236 Figure 2A), and some chromosomes had an enrichment of VNTR at their ends (Figure
237 2A). The distribution of STRs in the Sichuan taimen genome was relatively scattered,
238 with only the top 200 STRs by length exceeding the average length (16,620 bp) of the
239 HiFi reads (Figure 2B). Dinucleotide repeats accounted for 81.53% of STRs
240 (Supplementary Table S6), which is higher than the average level for vertebrates
241 (Adams et al., 2016). We identified 3,337 VNTRs with lengths over 100 Kb, totaling
242 0.97 Gb and covering 28.20% of the genome (Figure 2C). In addition, the presence of
243 TRs affected GC content in local regions of the genome; however, the GC content of
244 the TR regions was 41.16%, similar to the genome-wide range (Figure 2A).

245 **Gene annotation and genomic synteny**

246 A total of 40.00 Gb of RNA reads were generated from five tissues for RNA-based
247 gene annotation. After filtering pseudogenes, the BRAKER2 annotation pipeline
248 predicted 49,052 genes, with an average length of 14,577 bp, containing 338,448
249 exons in total. The average number of exons per gene was ~7 (Figure 1D). After
250 mapping to the NR databases, only 79.02% of the predicted genes were functionally
251 annotated. The TR ratio in intron regions was 22.74%, significantly lower than the
252 genome-wide level (Z-test; $P<0.01$). Based on kernel density estimation, 161 gene

253 hotspots were identified (Supplementary Figure S3). NR database annotation results
254 showed that the top three regions with the highest gene density were located in
255 histone gene family loci on chromosome 8 and 22. The BUSCO analysis for the
256 predicted genes indicated that 88.1% were complete BUSCOs (C: 88.1%[S: 43.3%, D:
257 44.8%], F: 4.3%, M: 7.6%, n: 4584).
258 After the protein mapping process, we identified 15,787 paired homeologous genes,
259 accounting for 64.37% of the total predicted genes. Eventually, 44 homeologous
260 blocks were located based on the genomic synteny of the genes (Figure 1E;
261 Supplementary Supplementary Table S7). The circos plot of homeologous blocks
262 showed that the longer chromosomes were more likely to be composed of multiple
263 homologous blocks, especially for the first five chromosomes larger than 120 Mb
264 (Figure 1E). Notably, more than half of the chromosomes (27/44) were composed of a
265 single homologous block, and the repeat distribution pattern was similar between
266 blocks (Figure 1E; Figure 2A).

267 **Analysis of phylogenetic and evolution of gene family**

268 We identified 35,589 orthologous gene families between Sichuan taimen and nine
269 other species, including 481 gene families specific to the Sichuan taimen and 342
270 single-copy orthologues. The results of the phylogenetic study showed that, as an
271 ancient branch of Salmoninae, the divergence time between huchonids (*Hucho* spp.
272 and *Brachymystax* spp.) and other Salmoninae fishes was estimated at approximately
273 33.29 Mya (Million years ago)(Figure 3; Supplementary Figure S4), which is
274 consistent with previous studies (Horreo, 2017; Lecaudey et al., 2018). The

275 divergence time estimated by PAML was 15.81 Mya between the taimen and its sister
276 group, lenok (*Brachymystax* spp.)(Figure 3).
277 We identified 235 significant expanded genes in the Sichuan taimen genome ($P <$
278 0.01). Gene Ontology analysis showed the highest enrichment in nucleosome
279 (GO:0000786) and chromosome (GO:0005694) categories (Supplementary Table S8).
280 Similarly, a large number of histone genes were found in the expanded gene set,
281 which we speculated might be influenced by the expansion of repetitive sequences.
282 Besides, the GO term fin development (GO:0033333) was annotated, implying that
283 the Sichuan taimen might have evolved enhanced swimming or predation abilities
284 (Supplementary Table S8). KEGG pathway analysis showed enrichment in the
285 Hedgehog signaling pathway (dre04340) and Insulin signaling pathway (dre04910),
286 suggesting changes in embryonic development, energy metabolism and growth, which
287 might be related to its special body size and habits (Supplementary Table S9).

288 **Ancestral chromosome status**

289 To evaluate the chromosome evolution of Sichuan taimen, genomic synteny blocks
290 were recognized against the Northern pike (*Esox lucius*), which serves as an
291 appropriate outgroup preceding the Ss4R. A total of 12,077 orthologs (1:2 or 1:1 for *E.*
292 *lucius* to *H. bleekeri*) were identified, resulting in the linkage of 80 homologous
293 blocks covering 90.89% of the length of the 44 Sichuan taimen chromosomes
294 (Supplementary Table S10). We recovered Sichuan taimen chromosome
295 equivalents for all Northern pike chromosomes, with each Northern pike chromosome
296 corresponding to two Sichuan taimen homologues (Figure 4A). The results showed

297 that half of the Sichuan taimen chromosomes (26/44) mapped completely to
298 individual Northern pike chromosomes (Figure 4A).
299 The total chromosome lengths were 3.32 Gb and 0.92 Gb for Sichuan taimen and
300 Northern pike, respectively. However, the length of the Sichuan taimen genome far
301 exceeded what would be expected from whole-genome duplication alone. The
302 expansion of TR in the Sichuan taimen genome was also evident in the circos plot
303 (Figure 4A). We further evaluated the TR expansion in the Sichuan taimen by
304 comparing the ratio of total length to TR length in the 80 homologous blocks. The
305 results revealed that all blocks showed TR expansion in the Sichuan taimen genome,
306 with the TR length ratio increasing proportionally with total length growth (Figure
307 4B). Linear regression analysis indicated that the TR ratio in the Sichuan taimen
308 genome was approximately fivefold higher than that of the Northern pike (Figure 4B).

309 **Demographic history and divergence for two taimen species**

310 The geographical distribution pattern of the other landlocked huchonid fish, lenok, is
311 similar to that of taimen. Both are found in the Qinling Mountain regions and the
312 Heilongjiang River, and the competitive relationship is existed between them (Galland
313 et al., 2021; Kaus et al., 2019). To reveal the divergence time between the two taimen
314 species, one Siberia taimen (*H. taimen*) sample were collected from the Heilongjiang
315 River (Figure 5A). A total of 44,120 genes were captured from the Siberia taimen
316 genome, and we identified 4,860 1:1 orthologous genes among the four species. The
317 result of BEAST showed that two taimen species separated at 2.29 Mya (Figure 5B),
318 which was similar to the previous inference (Kucinski & Fopp-Bayat, 2022).

319 To illustrate the trends in historical effective population sizes (N_e) changes of taimen
320 and lenok in Shaanxi and Heilongjiang, four resequencing samples from these regions
321 were used for joint analysis. The results showed that the histories of the four species
322 trace back to different times: ~9 Mya for the two Heilongjiang samples (*H. taimen*
323 and *B.tumernsis*), and ~1 Mya for the two Shaanxi samples (*H. bleekeri* and *B.*
324 *tsinlingensis*)(Figure 5C). *H. taimen* had a higher N_e level than *H. bleekeri* throughout
325 their entire history (Figure 5C). The N_e curve suggested that a large-scale population
326 expansion event occurred around 2 Mya in both *H. taimen* and *B.tumernsis*, followed
327 by a continuous decline in population size over millions of years (Figure 5C). In
328 summary, the Heilongjiang populations had higher N_e level than the Shaanxi
329 populations since ~200 Kya. Although the biomass of Sichuan taimen is currently
330 smaller than that of Qinling lenok, there might have been times in history when the
331 opposite was true (Figure 5C), which might reflect the competitive dynamics between
332 them.

333 **DISCUSSION**

334 **Complete genome assembly under high tandem repeat ratio**

335 The combination of HiFi and Hi-C reads allowed us to achieve the chromosome-level
336 assembly of the largest salmonid genome to date. The final size of the Sichuan taimen
337 genome was approximately 3.45 Gb, which exceeds the size of the Atlantic salmon
338 genome (~2.97 Gb)(Lien et al., 2016). The chromosome count of the Sichuan taimen
339 genome was 44, which approached to other salmonids such as *Salvelinus namaycush*
340 (42), *B. tsinlingensis* (41), and *Salmo trutta* (40)(Supplementary Table S11). The

341 assembled genome recovered 95.20% complete BUSCO genes, with a relatively high
342 percentage of duplicated BUSCOs (46.20%), similar to its sister group *B. tsinlingensis*
343 (Smith et al., 2022). Additionally, the chromosomal-level gene synteny implied the
344 integrity of the genome assembly and annotation.

345 The genome assembly process and outcome were affected by the ratio of repetitive
346 sequences (Tørresen et al., 2017; Treangen & Salzberg, 2011). The highly accurate
347 reads produced by CCS technology improved both the process and quality of the
348 assembly, especially inrepetitive sequences regions (Logsdon et al., 2020; Wenger et
349 al., 2019). We found that the assembled Sichuan taimen genome contains a high
350 proportion of repetitive sequences, with nearly half of the genome covered by tandem
351 repeat sequences. This may contribute to the increased genome size and pose
352 challenges for genome assembly. Fortunately, most of the TR regions were covered
353 and assembled by the HiFi reads, demonstrating the advantage of long reads.
354 However, some long TRs accounted for a relatively large proportion of the total
355 genome length, causing gaps around them. Generally, the number of gaps in the
356 assembled Sichuan taimen genome was similar to or even smaller than other salmonid
357 genomes, except for GCF_013265735.2, indicating that our assembly was not
358 severely disturbed by TRs (Supplementary Table S11).

359 Long repetitive regions can affect gene annotation, leading to the appearance of false
360 and redundant genes (Lang et al., 2020). In our preliminary gene annotation, we found
361 many tandem pseudogenes in long TR regions, which interfered with subsequent
362 analysis. Despite the presence of TRs in specific protein domains (Simon & Hancock,

363 2009), each of these pseudogenes usually had only one exon according to our manual
364 inspection. Therefore, we removed these single-exon genes located within repetitive
365 sequence regions based on expression levels (TPM<1). In parallel, we blasted these
366 pseudogenes to confirm their absence in other species.

367 **Phylogenetic relationships of Salmoninae based on single-copy orthologous genes**

368 Salmonidae contains three subfamilies: Coregoninae, Thymallinae, and Salmoninae.
369 However, their evolutionary relationships remains uncertain (Dysin et al., 2022).
370 Lineage-specific rediploidization, incomplete lineage sorting, hybridization, and
371 introgression have resulted in a complex and disordered phylogenetic relationship
372 within Salmonidae (Campbell et al., 2020; Lecaudey et al., 2018; Robertson et al.,
373 2017). Some studies suggest that Thymallinae is the sister group of Salmoninae
374 (Crête-Lafrenière et al., 2012; Mckay et al., 2004; Wang et al., 2011), while other
375 studies place Thymallinae at the root of the Salmonidae phylogenetic tree (Horreo,
376 2017; Wang et al., 2022). Based on the phylogenetic tree constructed using the
377 orthologous genes, we suggest that Coregoninae and Thymallinae have a sister
378 relationship, supported by 97% bootstraps (Supplementary Figure S4). This is
379 consistent with other trees based on orthologues (Macqueen & Johnston, 2014;
380 Robertson et al., 2017), mitochondrial genome (Campbell et al., 2013; Shedko et al.,
381 2013), and ultra-conserved elements (Campbell et al., 2020). The inferred topology
382 within Salmoninae aligns with other studies, with the genera *Hucho* and
383 *Brachymystax* forming the ancient branch of the subfamily, constituting a sister group
384 to the genera *Salmo, Salvelinus*, *and Oncorhynchus* (Crespi & Fulton, 2004; Kucinski

385 & Fopp-Bayat, 2022). The divergence time between taimen and lenok was estimated
386 at 9.41 to 23.51 Mya, which aligns with previous results based on mitochondria
387 (Kucinski & Fopp-Bayat, 2022), and coincides with the Miocene Climate Optimum
388 (MCO) at 15-18 Mya (Böhme, 2003; Sun & Zhang, 2008).

389 **Effective population size changes and the urgency of conservation**

390 All species of the genera *Hucho* and *Brachymystax* are endangered and protected by
391 Chinese government law (Sung et al., 1998). The results of PSMC revealed that the
392 effective populations size of four *Hucho* and *Brachymystax* species have declined
393 over the past million years, following the separation of *H. bleekeri* and *H. taimen*.
394 Similar patterns of N_e curves were observed in species with the same geographical
395 distribution. The curves of *H. bleekeri* and *B. tsinlingensis* in Shaanxi showed a close
396 coalescent endpoint, we speculated that these two salmonids shared the same
397 immigration route during the glacial period and were retained in low-latitude areas
398 during the interglacial period. The Sichuan taimen had a lower N_e level than Siberia
399 taimen in our calculation result, and current resource surveys indicate that Sichuan
400 taimen are very rare in the Qinling Mountains, having disappeared from most of their
401 original distribution areas (Shen et al., 2006; Wang et al., 2016). Consequently, more
402 effective conservation measures should be implemented to protect the Sichuan
403 taimen.

CONCLUSIONS

This study provides a comprehensive genome assembly of the Sichuan taimen, uncovering a novel tandem repeat sequences expansion phenomenon and offering valuable insights into the evolutionary history of this species. The detailed analysis of phylogenetic relationships and effective population size changes highlights the historical population dynamics of the Sichuan taimen and its relatives. The assembled genome serves as an essential resource for future research on the evolutionary biology and conservation genetics of salmonids. As the Sichuan taimen remains critically endangered, this genomic resource provides a foundation for developing effective conservation measures to ensure the survival of this ancient and unique species.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHOR CONTRIBUTIONS

Xin-miao Zhang, Li-xin Wang and Yu Jiang conceived and supervised the project. Dong-mei Xiong, and Sheng-long Jian the samples. Xin-miao Zhang, and Dong-mei Xiong carried out the bioinformatic analyses. Xin-miao Zhang wrote the manuscript. Sheng-long Jian, Yu Jiang and Li-xin Wang revised the manuscript. All authors have read and approved the final manuscript.

ACKNOWLEDGEMENTS

This study was financially supported by the National Natural Science Foundation of

China (32273138) and the latitudinal project foundation of China Power Construction Group Beijing Survey, Design and Research Institute Co., Ltd. (K4050422184) . We thank the High-Performance Computing Platform of Northwest A&F University and the National Supercomputing Center in Xi'an and Hefei for providing computing resources.

REFERENCE

Adams RH, Blackmon H, Reyes-Velasco J, et al. 2016. Microsatellite landscape evolutionary dynamics across 450 million years of vertebrate genome evolution. *Genome*, **59**(5): 295-310.

Allendorf FW & Hard JJ. 2009. Human-induced evolution caused by unnatural selection through harvest of wild animals. *Proceedings of the National Academy of Sciences of the United States of America*, **106**(1): 9987-9994.

Allendorf FW, Hohenlohe PA, Luikart G. 2010. Genomics and the future of conservation genetics. *Nature Reviews Genetics*, **11**(10): 697-709.

Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Researchearch*, **27**(2): 573-580.

Böhme M. 2003. The Miocene climatic optimum: evidence from ectothermic vertebrates of Central Europe. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **195**(3-4): 389-401.

Braasch I, Gehrke AR, Smith JJ, et al. 2016. The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. *Nature Genetics*, **48**(4): 427-437.

Brandies P, Peel E, Hogg CJ, et al. 2019. The value of reference genomes in the conservation of threatened species. *Genes (Basel)*, **10**(11).

Bray NL, Pimentel H, Melsted P, et al. 2016. Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology*, **34**(5): 525-527.

Campbell MA, Buser TJ, Alfaro ME, et al. 2020. Addressing incomplete lineage sorting and paralogy in the inference of uncertain salmonid phylogenetic relationships. *PeerJ*, **8**: e9389.

Campbell MA, López JA, Sado T, et al. 2013. Pike and salmon as sister taxa: detailed intraclade resolution and divergence time estimation of Esociformes + Salmoniformes based on whole mitochondrial genome sequences. *Gene*, **530**(1): 57-65.

Chen S, Zhou Y, Chen Y, et al. 2018. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, **34**(17): i884-i890.

Chen Y, Yang H, Chen Y, et al. 2021. Full-length transcriptome sequencing and identification of immune-related genes in the critically endangered *Hucho bleekeri*. *Developmental & Comparative Immunology*, **116**: 103934.

Chen Y, Yang H, Gong Q, et al. 2020. Isolation and characterization of 34 SNP markers in *Hucho bleekeri*. *Conservation Genetics Resources*, **12**: 157-160.

Cheng H, Concepcion GT, Feng X, et al. 2021. Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm. *Nature Methods*, **18**(2): 170-175.

461 Christensen KA, Rondeau EB, Minkley DR, et al. 2020. The sockeye salmon genome, transcriptome,
462 and analyses identifying population defining regions of the genome. *PLoS One*, **15**(10): e0240935.
463 Crespi BJ & Fulton MJ. 2004. Molecular systematics of Salmonidae: combined nuclear data yields a
464 robust phylogeny. *Molecular Phylogenetics and Evolution*, **31**(2): 658-679.
465 Crête-Lafrenière A, Weir LK, Bernatchez L. 2012. Framing the Salmonidae family phylogenetic
466 portrait: a more complete picture from increased taxon sampling. *PLoS One*, **7**(10): e46662.
467 Danecek P, Bonfield JK, Liddle J, et al. 2021. Twelve years of SAMtools and BCFtools. *Gigascience*,
468 **10**(2):giab008.
469 Dobin A, Davis CA, Schlesinger F, et al. 2013. STAR: ultrafast universal RNA-seq aligner.
470 *Bioinformatics*, **29**(1): 15-21.
471 Dudchenko O, Batra SS, Omer AD, et al. 2017. De novo assembly of the *Aedes aegypti* genome using
472 Hi-C yields chromosome-length scaffolds. *Science*, **356**(6333): 92-95.
473 Durand NC, Robinson JT, Shamim MS, et al. 2016a. Juicebox provides a visualization system for Hi-C
474 contact maps with unlimited zoom. *Cell Systems*, **3**(1): 99-101.
475 Durand NC, Shamim MS, Machol I, et al. 2016b. Juicer provides a one-click system for analyzing
476 loop-resolution Hi-C experiments. *Cell Systems*, **3**(1): 95-98.
477 Dysin AP, Shcherbakov YS, Nikolaeva OA, et al. 2022. Salmonidae genome: features, evolutionary
478 and phylogenetic characteristics. *Genes (Basel)*, **13**(12):2221.
479 Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput.
480 *Nucleic Acids Research*, **32**(5): 1792-1797.
481 Emms DM & Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics.
482 *Genome Biology*, **20**(1): 1-14.
483 Flynn JM, Hubley R, Goubert C, et al. 2020. RepeatModeler2 for automated genomic discovery of
484 transposable element families. *Proceedings of the National Academy of Sciences of the United States of*
485 *America*, **117**(17): 9451-9457.
486 Gabriel L, Hoff KJ, Brůna T, et al. 2021. TSEBRA: transcript selector for BRAKER. *BMC*
487 *Bioinformatics*, **22**(1): 566.
488 Galland LM, Simmons JB, Jahner JP, et al. 2021. Hierarchical genetic structure and implications for
489 conservation of the world's largest salmonid, *Hucho taimen*. *Sci Rep*, **11**(1): 20508.
490 Gao G, Magadan S, Waldbieser GC, et al. 2021. A long reads-based de-novo assembly of the genome
491 of the Arlee homozygous line reveals chromosomal rearrangements in rainbow trout. *G3 (Bethesda)*,
492 **11**(4):jkab052.
493 Hansen T, Fjelldal PG, Lien S, et al. 2021. The genome sequence of the brown trout, *Salmo trutta*
494 Linnaeus 1758. *Wellcome Open Research*, **6**: 108.
495 Hoff KJ, Lange S, Lomsadze A, et al. 2016. BRAKER1: Unsupervised RNA-Seq-Based Genome
496 Annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics*, **32**(5): 767-769.
497 Hoff KJ, Lomsadze A, Borodovsky M, et al. 2019. Whole-Genome Annotation with BRAKER.
498 *Methods in Molecular Biology*, **1962**: 65-95.
499 Horreo JL. 2017. Revisiting the mitogenomic phylogeny of Salmoninae: new insights thanks to recent
500 sequencing advances. *PeerJ*, **5**: e3828.
501 Hu M, Wang Y, Cao L, et al. 2008. Threatened fishes of the world: *Hucho bleekeri* Kimura, 1934
502 (Salmonidae). *Environmental Biology of Fishes*, **82**: 385-386.
503 Ishiguro NB, Miya M, Nishida M. 2003. Basal euteleostean relationships: a mitogenomic perspective
504 on the phylogenetic reality of the "Protacanthopterygii". *Molecular Phylogenetics and Evolution*, **27**(3):

476-488.
Kaus A, Michalski S, Hänfling B, et al. 2019. Fish conservation in the land of steppe and sky: Evolutionarily significant units of threatened salmonid species in Mongolia mirror major river basins. *Ecology and Evolution*, **9**(6): 3416-3433.
Kimura S. 1934. Description of the fishes collected from the Yangtze-Kiang China by late Dr. K. Kishinouye and his party in 1927–1929. Shanghai: Journal of the Shanghai Science Institute.
Kohn MH, Murphy WJ, Ostrander EA, et al. 2006. Genomics and conservation genetics. *Trends in Ecology Evolution*, **21**(11): 629-637.
Kucinski M & Fopp-Bayat D. 2022. Phylogenetic analysis of *Brachymystax* and *Hucho* genera-Summary on evolutionary status within the Salmoninae subfamily. *Journal of Applied Ichthyology*, **38**(4): 403-411.
Kumar S, Suleski M, Craig JM, et al. 2022. TimeTree 5: An Expanded Resource for Species Divergence Times. *Molecular Biology and Evolution*, **39**(8).
Lang D, Zhang S, Ren P, et al. 2020. Comparison of the two up-to-date sequencing technologies for genome assembly: HiFi reads of Pacific Biosciences Sequel II system and ultralong reads of Oxford Nanopore. *Gigascience*, **9**(12):giaa123.
Lecaudey LA, Schliewen UK, Osinov AG, et al. 2018. Inferring phylogenetic structure, hybridization and divergence times within Salmoninae (Teleostei: Salmonidae) using RAD-sequencing. *Molecular Phylogenetics and Evolution*, **124**: 82-99.
Li H & Durbin R. 2011. Inference of human population history from individual whole-genome sequences. *Nature*, **475**(7357): 493-496.
Lien S, Koop BF, Sandve SR, et al. 2016. The Atlantic salmon genome provides insights into rediploidization. *Nature*, **533**(7602): 200-205.
Liu B, Shi Y, Yuan J, et al. 2013. Estimation of genomic characteristics by analyzing k-mer frequency in *de novo* genome projects. *arXiv.org arXiv*: 1308.2012.
Liu X, Zeng H, Wang C, et al. 2022. Improved genome assembly of Chinese sucker (*Myxocyprinus asiaticus*) provides insights into the identification and characterization of pharyngeal teeth related maker genes in Cyprinoidei. *Water Biology and Security*, **1**(3): 100049.
Logsdon GA, Vollger MR, Eichler EE. 2020. Long-read human genome sequencing and its applications. *Nature Reviews Genetics*, **21**(10): 597-614.
Macqueen DJ & Johnston IA. 2014. A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. *Proceedings of the Royal Society B-Biological Sciences*, **281**(1778): 20132881.
Mckay SJ, Trautner J, Smith MJ, et al. 2004. Evolution of duplicated growth hormone genes in autotetraploid salmonid fishes. *Genome*, **47**(4): 714-723.
Mendes FK, Vanderpool D, Fulton B, et al. 2020. CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics*, **36**(22-23): 5516-5518.
Ou S & Jiang N. 2018. LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiology*, **176**(2): 1410-1422.
Robertson FM, Gundappa MK, Grammes F, et al. 2017. Lineage-specific rediploidization is a mechanism to explain time-lags between genome duplication and evolutionary diversification. *Genome Biology*, **18**(1): 111.
Sadovy De Mitcheson Y, Craig MT, Bertoncini AA, et al. 2013. Fishing groupers towards extinction: a global assessment of threats and extinction risks in a billion dollar fishery. *Fish and Fisheries*, **14**(2):

549 119-136.
550 Shedko SV, Miroshnichenko IL, Nemkova GA. 2013. Phylogeny of salmonids (Salmoniformes,
551 Salmonidae) and molecular dating: analysis of mtDNA data. *Genetika*, **49**(6): 718-734.
552 Shen Z, Tang W, Li K. 2006. The analysis of population dynamics of *Hucho bleekeri* in Markehe River,
553 Qinghai Province. *Reservoir Fisheries*, **26**: 71-73 (In Chinese).
554 Simão FA, Waterhouse RM, Ioannidis P, et al. 2015. BUSCO: assessing genome assembly and
555 annotation completeness with single-copy orthologs. *Bioinformatics*, **31**(19): 3210-3212.
556 Simon M & Hancock JM. 2009. Tandem and cryptic amino acid repeats accumulate in disordered
557 regions of proteins. *Genome Biology*, **10**(6): R59.
558 Smith SR, Normandeau E, Djambazian H, et al. 2022. A chromosome-anchored genome assembly for
559 Lake Trout (*Salvelinus namaycush*). *Molecular Ecology Resources*, **22**(2): 679-694.
560 Song Z. 2012. *Hucho bleekeri*. The IUCN Red List of Threatened Species. Version 2014.3.
561 Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
562 phylogenies. *Bioinformatics*, **30**(9): 1312-1313.
563 Stefanova P, Taseva M, Georgieva T, et al. 2013. A modified CTAB method for DNA extraction from
564 soybean and meat products. *Biotechnology & Biotechnological Equipment*, **27**(3): 3803-3810.
565 Sun J & Zhang Z. 2008. Palynological evidence for the Mid-Miocene Climatic Optimum recorded in
566 Cenozoic sediments of the Tian Shan Range, northwestern China. *Global and Planetary Change*, **64**(1):
567 53-68.
568 Sung W, Peiqi Y, Yiyu C, et al. 1998. China Red data book of endangered animals: Pisces. Beijing:
569 Science press.
570 Tarailo-Graovac M & Chen N. 2009. Using RepeatMasker to identify repetitive elements in genomic
571 sequences. *Current Protocols in Bioinformatics*, **4**: 1-4.10.
572 Tørresen OK, Star B, Jentoft S, et al. 2017. An improved genome assembly uncovers prolific tandem
573 repeats in Atlantic cod. *BMC Genomics*, **18**(1): 95.
574 Treangen TJ & Salzberg SL. 2011. Repetitive DNA and next-generation sequencing: computational
575 challenges and solutions. *Nature Reviews Genetics*, **13**(1): 36-46.
576 Wang K, Zhang SH, Wang DQ, et al. 2016. Conservation genetics assessment and phylogenetic
577 relationships of critically endangered *Hucho bleekeri* in China. *Journal of Applied Ichthyology*, **32**(2):
578 343-349.
579 Wang Y, Guo R, Li H, et al. 2011. The complete mitochondrial genome of the Sichuan taimen (*Hucho*
580 *bleekeri*): repetitive sequences in the control region and phylogenetic implications for Salmonidae.
581 *Marine Genomics*, **4**(3): 221-228.
582 Wang Y, Xiong F, Song Z. 2022. Molecular phylogeny and adaptive mitochondrial DNA evolution of
583 salmonids (Pisces: Salmonidae). *Frontiers in Genetics*, **13**: 903240.
584 Waples RS, Naish KA, Primmer CR. 2020. Conservation and management of salmon in the age of
585 genomics. *Annual Review of Animal Biosciences*, **8**: 117-143.
586 Wenger AM, Peluso P, Rowell WJ, et al. 2019. Accurate circular consensus long-read sequencing
587 improves variant detection and assembly of a human genome. *Nature Biotechnology*, **37**(10):
588 1155-1162.
589 Xu Z & Wang H. 2007. LTR_FINDER: an efficient tool for the prediction of full-length LTR
590 retrotransposons. *Nucleic Acids Researchearch*, **35**(2): W265-W268.
591 Xue DX, Xing TF, Liu JX. 2022. A high-quality chromosome-level genome of the endangered
592 roughskin sculpin provides insights into its evolution and adaptation. *Molecular Ecology Resources*,

22(5): 1892-1905.
Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, **24**(8): 1586-1591.
Ye J, Mcginnis S, Madden TL. 2006. BLAST: improvements for better sequence analysis. *Nucleic Acids Research*, **34**(Web Server issue): W6-9.
Zhang Y, Luan P, Ren G, et al. 2020. Estimating the inbreeding level and genetic relatedness in an isolated population of critically endangered Sichuan taimen (*Hucho Bleekeri*) using genome-wide SNP markers. *Ecology and Evolution*, **10**(3): 1390-1400.
Zhu W, Wang Z, Li H, et al. 2022. A chromosome-level genome of *Brachymystax tsinlingensis* provides resources and insights into salmonids evolution. *G3 (Bethesda)*, **12**(8):jkac162.

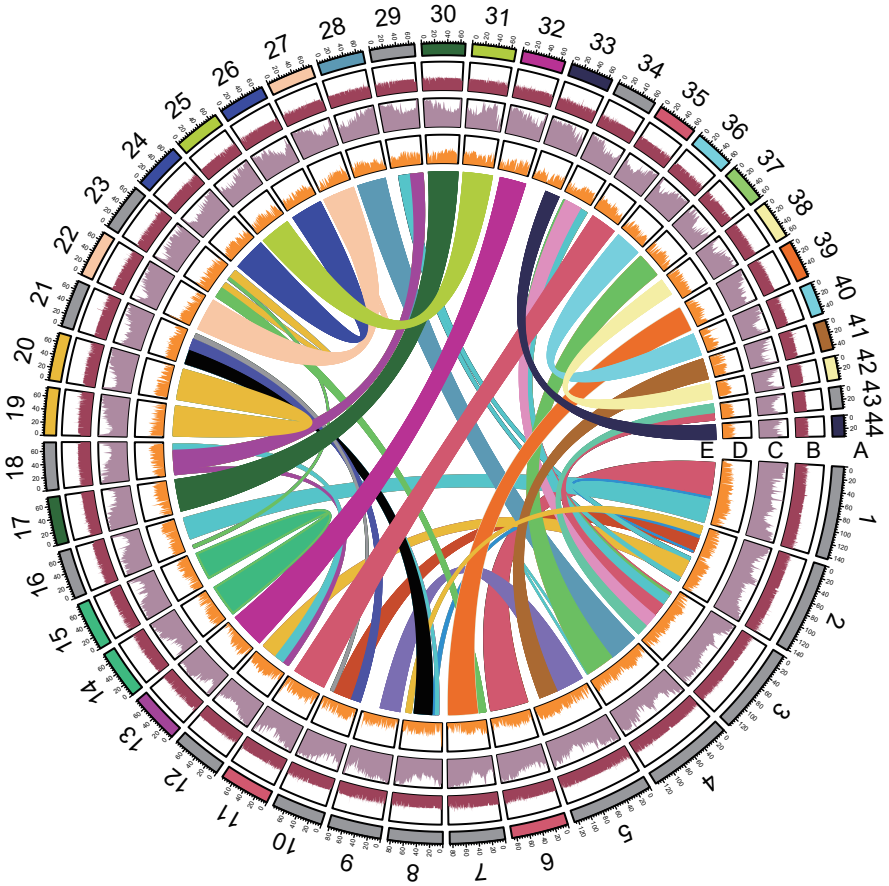
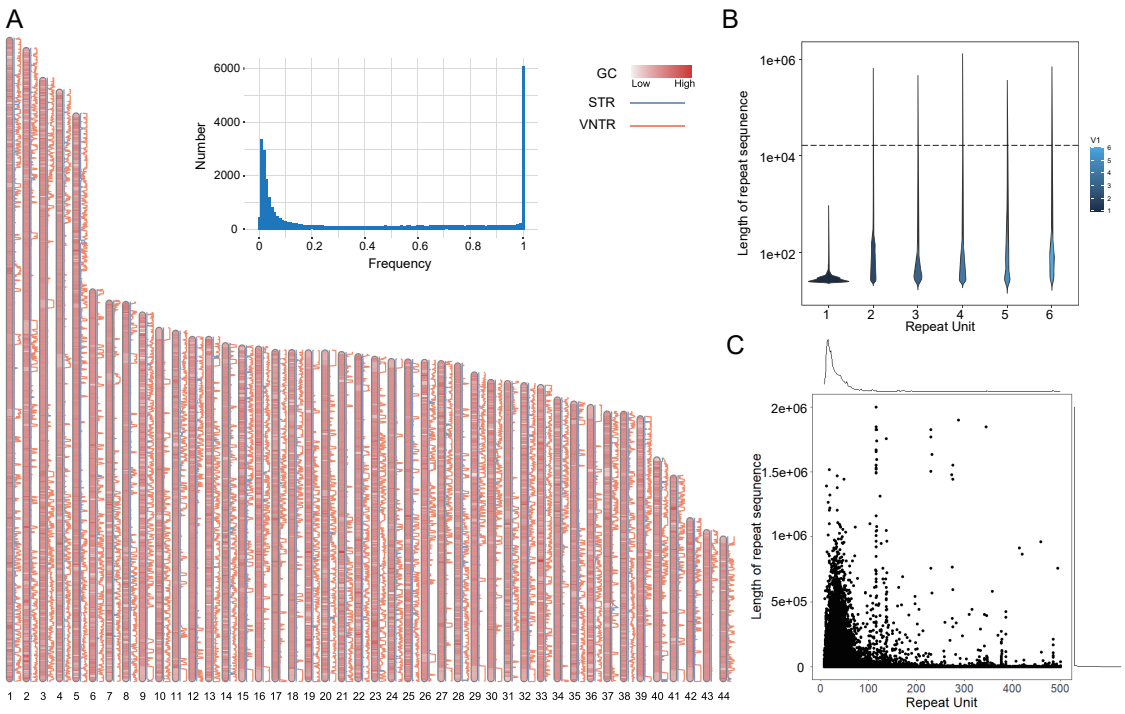


Figure 1 Circos plot of the genome assembly statistics of Sichuan taimen

A: Chromosome names and length scales. Chromosomes consisting of multiple homologous regions were shown in gray (The same below). B: GC ratio (scale: 0-100%). C: Proportion of total repeats (TR and TE)(scale: 0-100%). D: Gene density after log transformation (scale: 0-7). E: Homeologous regions in the Sichuan taimen genome recognized by 1:1 duplicated proteins.



608 **Figure 2 Distribution and statistics of repetitive sequence in the Sichuan taimen genome.**

609 A: Visualization of genomic GC content, STR, and VNTR distribution based on 100 Kb windows;
610 the bar graph reflects the TR frequency distribution in genome-wide 100 Kb windows. B and C
611 were statistics of tandem repeat unit length and total tandem repeat sequence length of STRs and
612 VNTRs respectively; The average length of the HiFi reads (16620 bp) was labeled as dashed line.

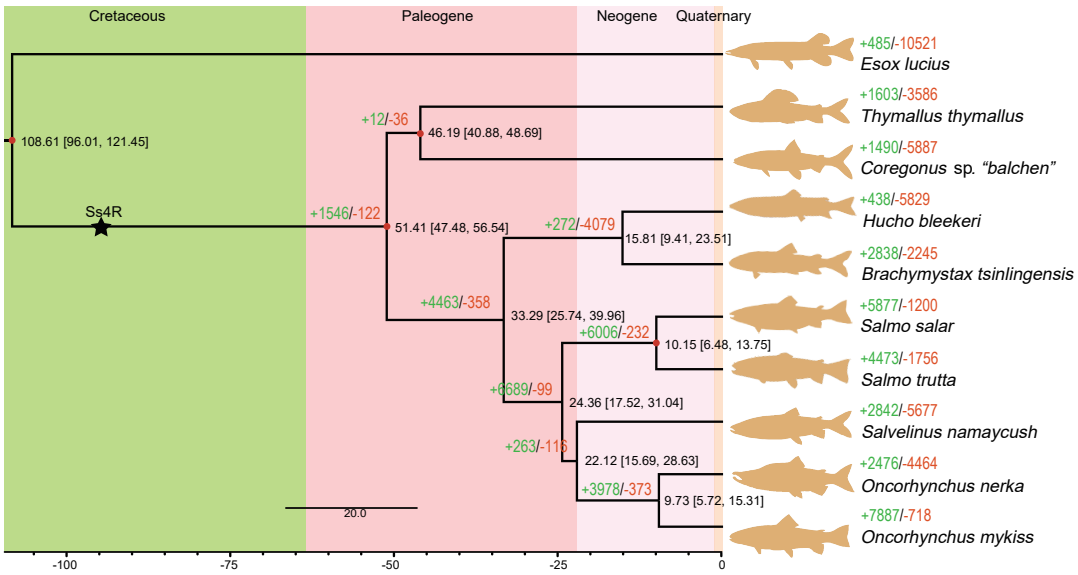
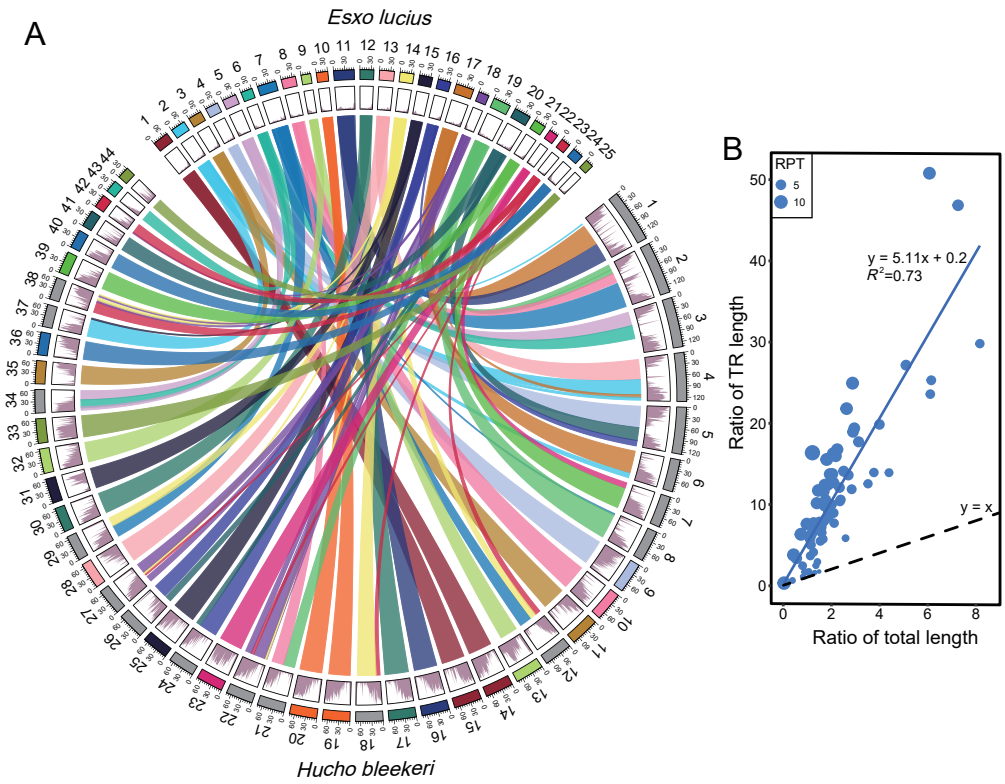


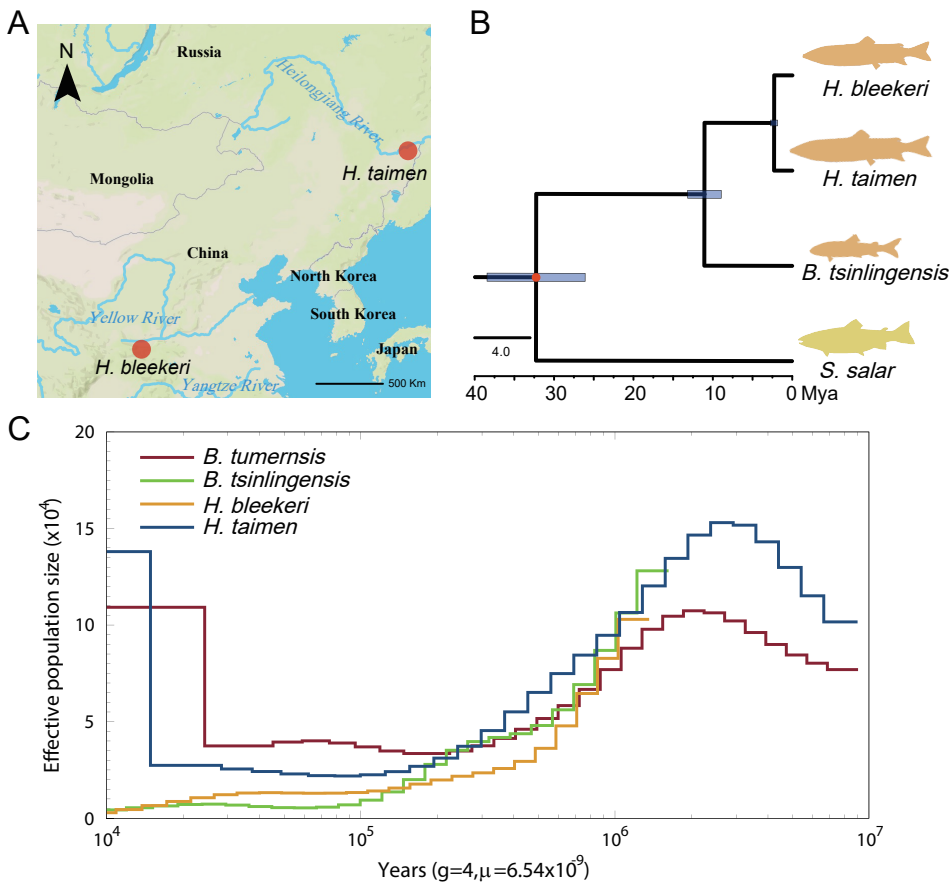
Figure 3 Maximum likelihood phylogenetic tree constructed based on single-copy orthologous genes.

Fossil calibration points used by MCMC are marked in red. Black numbers next to each nodes are the divergence times (Mya) and the 95% confidential intervals estimated by PAML. The green positive and red negative integers indicate the number of genes at expanded or contracted gene families at that node, respectively ($P < 0.05$). Ss4R: salmonid-specific fourth vertebrate whole-genome duplication.



620 **Figure 4 Chromosome evolution of *H. bleekeri***

621 A: Circos plot of homologous blocks and TR ratio in *E. lucius* and *H. bleekeri*. From the outside
622 to the inside were chromosome length, TR ratio, and homologous block connections. B: TR
623 expansion in taimen genome. The abscissa and ordinate were the ratio of the total length and TR
624 length in homologous blocks between Sichuan taimen and Northern pike, respectively. The size of
625 the circles was expressed as the ratio of the proportion of TR (RPT) in homologous blocks
626 between them.



627 **Figure 5 Species divergence history between *H. taimen* and *H. bleekeri***

628 A: Map of the two taimen species distribution. B: Phylogeny and divergence time of huchonid.
629 95% confidence interval was labeled as blue frame. Fossil calibration points were marked in red.
630 C: Demographic history of two Heilongjiang sample (*H. taimen* and *B.tumernsis*) and two Qinling
631 sample (*H. bleekeri* and *B.tsinlingensis*). The generation (g) time was set at 4 years and the neutral
632 mutation rate per generation (μ) was set at 6.54 × 10-9.

633 **Table 1. Summary statistics of the Sichuan taimen genome assembly**

Category	Length (bp)		Number	
	Contig	Scaffold	Contig	Scaffold
Max	25,857,850	143,121,144	-	-
Min	3,656	18,648	-	-
N50	2,400,826	73,705,547	354	19
N90	418,871	59,067,955	1680	39
Total	3,479,724,502	3,451,243,059	4373	505

634 **SUPPLEMENTARY FIGURE AND TABLE LEGEND**

635 Supplementary Figure S1. Frequency distribution map in 17-mer.

636 Supplementary Figure S2. Hi-C interaction heatmap of Hucho bleekeri genome
637 between 44 chromosomes and unmapped scaffolds.

638 Supplementary Figure S3. Gene hotspots of the hucho genome.

639 Supplementary Figure S4. Phylogenetic trees under 1,000 bootstraps.

640 Supplementary Table S1. Genome size estimation table at k=17.

641 Supplementary Table S2. Summarized benchmarking in BUSCO notation.

642 Supplementary Table S3. Summary of TE in 10 salmonids (%).

643 Supplementary Table S4. Summary of TE in 10 salmonids (length).

644 Supplementary Table S5. Statistics of TR for 10 salmonids.

645 Supplementary Table S6. Summary of STRs' unit and total length.

646 Supplementary Table S7. Genomic synteny of 44 homeologous blocks.

647 Supplementary Table S8. The Gene Ontology annotation of the 235 positive selected
648 genes.

649 Supplementary Table S9. The KEGG pathway annotation of the 235 positive selected
650 genes.

651 Supplementary Table S10. Genomic synteny of 80 homeologous blocks between
652 Northern pike (luc) and Sichuan taimen.

653 Supplementary Table S11. Summary of genome assembly results of 10 salmonids.