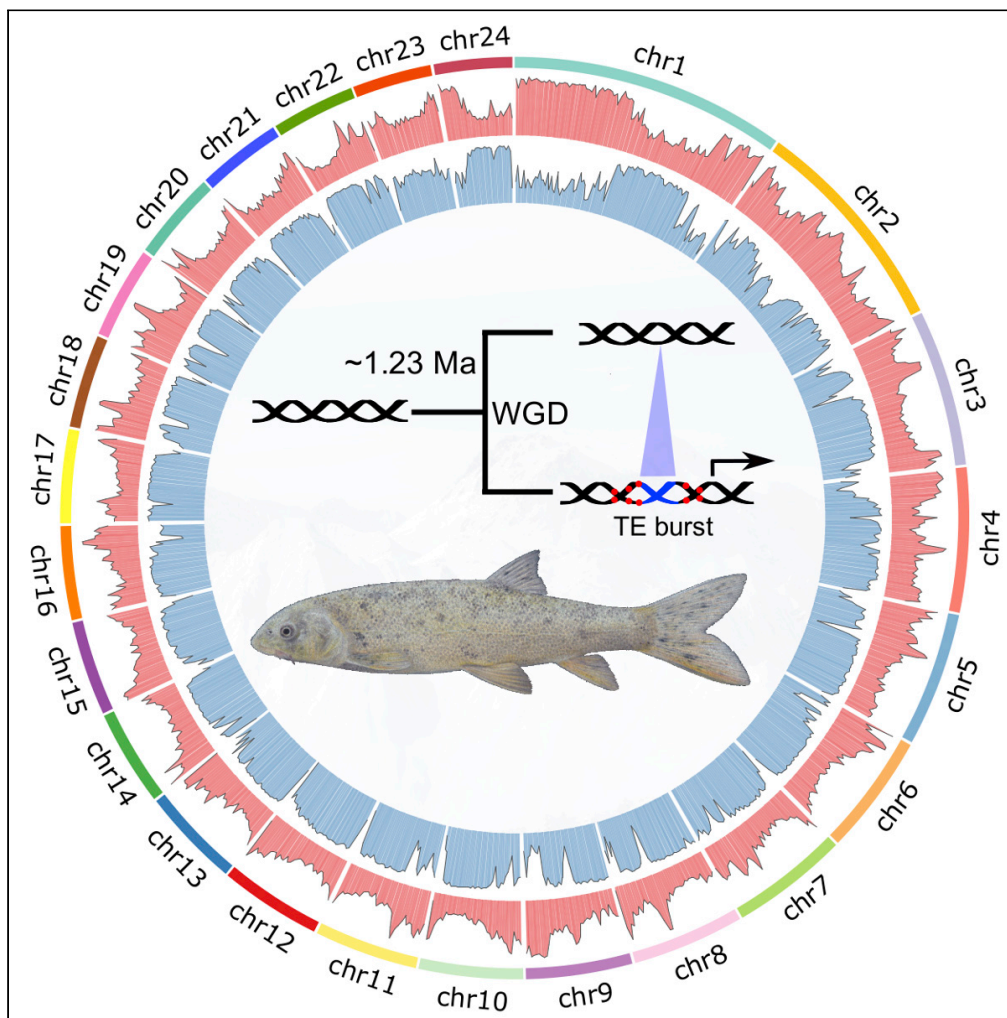


Article

Genome of Tetraploid Fish *Schizothorax o'connori* Provides Insights into Early Re-diploidization and High-Altitude Adaptation

Shijun Xiao,
Zhenbo Mou,
Dingding Fan, ...,
Hong Li, Di Wang,
Haiping Liu

luihappy@163.com

HIGHLIGHTS

Schizothorax o'connori underwent a recent WGD later than ~1.23 mya

Transposon bursts prompted sequence divergence after duplication

Transposon can induce sequence mutagenesis and regulate gene expression

Many genes in DNA repair and folate transport/metabolism were naturally selected

Xiao et al., iScience 23, 101497
September 25, 2020 © 2020
The Authors.
<https://doi.org/10.1016/j.isci.2020.101497>

Article

Genome of Tetraploid Fish *Schizothorax o'connori* Provides Insights into Early Re-diploidization and High-Altitude Adaptation

Shijun Xiao,^{1,2,6,8} Zhenbo Mou,^{1,8} Dingding Fan,^{3,8} He Zhou,⁴ Ming Zou,² Yu Zou,² Chaowei Zhou,^{1,5} Ruibin Yang,³ Jiaqi Liu,² Shilin Zhu,² Yajuan Li,⁴ Yanchao Liu,¹ Fei Liu,¹ Wanliang Wang,¹ Benhe Zeng,¹ Hong Li,⁷ Di Wang,⁷ and Haiping Liu^{1,9,*}

SUMMARY

Whole-genome duplications (WGDs) of Schizothoracinae are believed to have played a significant role in speciation and environmental adaptation on the Qinghai-Tibet Plateau (QTP). Here, we present a genome for *Schizothorax o'connori*, a QTP endemic fish and showed the species as a young tetraploid with a recent WGD later than ~1.23 mya. We exhibited that massive insertions between duplicated genomes caused by transposon bursts could induce mutagenesis in adjacent sequences and alter the expression of neighboring genes, representing an early re-diploidization process in a polyploid genome after WGD. Meanwhile, we found that many genes involved in DNA repair and folate transport/metabolism experienced natural selection and might contribute to the environmental adaptation of this species. Therefore, the *S. o'connori* genome could serve as a young tetraploid model for investigations of early re-diploidization in polyploid genomes and offers an invaluable genetic resource for environmental adaptation studies of the endemic fish of the QTP.

INTRODUCTION

The Qinghai-Tibet plateau (QTP) is the youngest, largest, and highest plateau on Earth, with an average altitude higher than 4,000 m (Pan et al., 2012). The plateau has been uplifting continuously for the past 50 Ma (Jiang and Li, 2014; Li et al., 2015). The high altitude, low temperatures, dramatic temperature fluctuations, and high UV radiation of the QTP make the organisms endemic to this region ideal models for studies of the molecular mechanisms underlying adaptation to harsh environments (Zhang et al., 2016). Schizothoracinae (order Cypriniformes, family Cyprinidae) is an important subfamily of endemic fish on the QTP (Qi et al., 2012), which includes 12 genera and more than 70 species/subspecies (Xu et al., 2016). The overwhelming majority of Schizothoracinae species are polyploid, including examples from tetraploid to hexaploid (Zan et al., 1985), implying that the ancestors of these species underwent the additional subfamily-specific rounds of whole-genome duplication (4R WGD) after the teleost-specific third round WGD (3R WGD). Scientists believed that the WGD of Schizothoracinae played a central role in their speciation and adaptation during QTP uplifts (Zheng et al., 2002). *Schizothorax o'connori* is a typical tetraploid Schizothoracinae species that is distributed widely in the middle section of the Yarlung Tsangpo River, a core area of the QTP (Ma et al., 2010; Zhang, 2011). The complex age structure, slow growth, late sexual maturity, low fecundity, and short reproductive period of this population of *S. o'connori* make them very sensitive to human activities and overfishing (Guo et al., 2016). In a recent investigation of fishery resources in the middle reaches of the Yarlung Zangbo river, it was suggested that *S. o'connori* is a threatened species (Ma et al., 2020). Recent successes in the artificial breeding of *S. o'connori* suggest that this fish may be a promising resource for local aquaculture (Zhang, 2011). Therefore, *S. o'connori* is a potential model for investigations of genome evolution, specifically the re-diploidization process after WGD and the molecular mechanisms underlying high-altitude adaptations of the species of the QTP.

WGD is a major driver of genome variation, as WGD can increase genomic diversity, facilitating possible environmental adaptations and speciation (Berthelot et al., 2014; Lien et al., 2016; MacKintosh and Ferrier,

¹Institute of Fisheries Science, Tibet Academy of Agricultural and Animal Husbandry Sciences, Lhasa, China

²Department of Computer Science, Wuhan University of Technology, Wuhan, China

³College of Fisheries, Huazhong Agricultural University, Wuhan, China

⁴College of Fisheries and Life Science, Dalian Ocean University, Dalian, China

⁵Department of Aquaculture, College of Animal Science, Southwest University, Chongqing, China

⁶College of Plant Protection, Jilin Agriculture University, Changchun, Jilin, China

⁷Novogene Bioinformatics Institute, Beijing, China

⁸These authors contributed equally

⁹Lead Contact

*Correspondence: luihappy@163.com

<https://doi.org/10.1016/j.isci.2020.101497>



2017; Xu et al., 2014; Yang et al., 2016). The teleost 3R WGD, which occurred about 320 mya, played an important role in the diversification of ancient teleosts (Yang et al., 2016). Therefore, polyploid fish species can be used to study post-WGD genome evolution and to determine the contribution of WGD to speciation and environmental adaptation (Chen et al., 2019). Re-diploidization is an essential post-WGD evolutionary process, which differentiates and stabilizes duplicated genomes (Lien et al., 2016). The characterization of polyploid genomes might widen our understanding of the post-WGD re-diploidization process and reveal possible mechanisms of speciation and adaptation after WGD. Massive sequence divergences, chromosome rearrangements, and large-scale transposon bursts have been identified as the main features of genome re-diploidization, based on investigations of the *Salmo salar* and *Oncorhynchus mykiss* genomes, which were duplicated 80–103 mya (Berthelot et al., 2014; Lien et al., 2016; Macqueen and Johnston, 2014). However, genome characterizations of the early re-diploidization process of recent WGD, such as the genomic arrangements and transposon expansions along the genome, have not been revealed yet.

In this work, we generated an *S. o'connori* polyploidy genome. We found that *S. o'connori* is a young tetraploid, and its ancestral WGD occurred around 1.23 mya. Although the synteny of the *S. o'connori* genome was largely conserved with respect to that of *Danio rerio*, we detected massive sequence insertion differences between duplicated regions in the *S. o'connori* genome, which are caused primarily by transposon bursts. Those insertions influenced adjacent sequence differentiations and nearby gene expressions. Therefore, we propose that transposon insertions, coupled with the accumulation of mutations and changes in expression patterns in duplicated genes, represent the genomic features of the early stage of re-diploidization after this recent WGD. Meanwhile, a large number of genes associated with DNA repair, folate transport/metabolism, and energy metabolism were positively selected. These genes might participate in the environmental adaptations of *S. o'connori*.

RESULTS

Genome Assembly and Annotation

To collect samples for genome sequencing, wild female *S. o'connori* (see Table S1 for species abbreviations) individuals were captured in the Yarlung Tsangpo river in Tibet (Figures 1A and 1B). The genome of *S. o'connori* was sequenced using next-generation sequencing (NGS) on an Illumina HiSeq XTen platform and using Pacific Bioscience's single-molecule real-time (SMRT) sequencing. We obtained 295.2 Gb of short reads (Table S2) and estimated genome features with a K-mer-based method (Table S3). The *S. o'connori* genome is 1.94 Gb, comprising 56.7% repetitive elements. This genome is larger than those of most other teleosts (~700 Mb) and is similar in size to related genomes that have undergone additional rounds of WGDs, such as *Cyprinus carpio* (1.83 Gb) (Xu et al., 2014) and *Carassius auratus* (1.85 Gb) (Chen et al., 2019). We obtained a coverage depth of approximately 68X, and in total 141 Gb sub-reads were generated using the Pacific Sequel platform (Table S2). An *S. o'connori* genome of 2.07 Gb length was assembled using FALCON-unzip (Chin et al., 2016) with a contig N50 of 241.9 kb (Tables 1 and S4).

A comprehensive gene prediction strategy, combining *de novo* and homologous proteins-based method, as well as RNA sequencing (RNA-seq)-derived evidence from 15 tissues (Table S5), was performed for the *S. o'connori* genome. The homologous proteins were culled from other Cypriniformes of *Astyanax mexicanus*, *D. rerio*, *Ctenopharyngodon idellus*, *Sinocyclocheilus grahami*, *C. carpio*, and *C. auratus*; a Salmoniforme of *S. salar*; and a Siluriforme of *Glyptosternum maculatum* (Table S6). The common ancestor of the Cypriniforme species is estimated to live about 49–54 mya (Wang et al., 2015). As a result, we identified 43,731 predicted protein-coding genes from the *S. o'connori* genome (Tables 1 and S6). In addition, about 96.4% of all the 2,586 Benchmarking Universal Single-Copy Orthologs (BUSCOs) (Waterhouse et al., 2018) were annotated in the *S. o'connori* genome (Table S7). More than 99.97% of the protein-coding genes predicted in the *S. o'connori* genome could be functionally annotated successfully using public databases (Table S8). This high-quality genome assembly and annotations provide a valuable reference for genome evolution and population genetic studies of *S. o'connori*.

S. o'connori WGD and Chromosomal Comparisons to *Danio rerio*

Previous cytogenetic studies have shown that most species of Schizothoracinae are polyploid (Zan et al., 1985). The karyotype of *S. o'connori* is still controversial and has been reported as $2n = 92$ or $2n = 106$ in previous studies (Yunfei et al., 1999). The chromosome number is roughly twice as large as diploid cyprinid

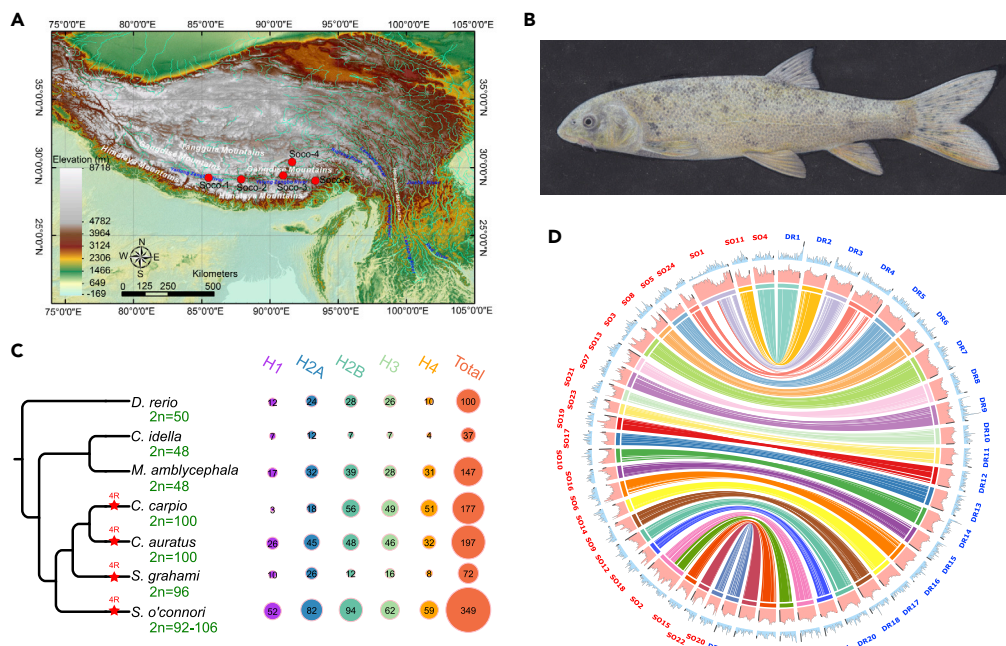


Figure 1. Samples from the Qinghai-Tibet Plateau and Genome Assembly for *Schizothorax o'connori*

(A) The sample distribution for the five populations used in this work.

(B) A photograph of *Schizothorax o'connori*.

(C) The expansion of core histone genes in *Schizothorax o'connori* (family Cyprinidae, subfamily Schizothoracinae), compared with *Megalobrama amblycephala* (family Cyprinidae, subfamily Cultrinae), *Ctenopharyngodon idella* (family Cyprinidae, subfamily Leuciscinae), *Cyprinus carpio* (family Cyprinidae, subfamily Cyprininae), *Carassius auratus* (family Cyprinidae, subfamily Cyprininae), and *Sinocyclocheilus grahami* (family Cyprinidae, subfamily Barbinae). Chromosome numbers are shown in green near species names, and red stars illuminated 4R WGDs. Note that positions of red stars did not represent timings of WGD.

(D) The pseudo-chromosome assembly of *Schizothorax o'connori* and conservation synteny relationship to *Danio rerio*.

species, including *D. rerio* ($2n = 50$) and *C. idellus* ($2n = 48$) (Wang et al., 2015), suggesting that *S. o'connori* is likely tetraploid (Dai and Han, 2018). Several additional features of the *S. o'connori* genome further support that this genome had been subjected to an extra 4R WGD. First, the number of protein-coding genes predicted from the *S. o'connori* genome (43,731) was comparable to numbers in other tetraploids, such as *C. carpio* (Xu et al., 2014), *C. auratus* (Chen et al., 2019), and *S. grahami* (Yang et al., 2016) (Figure 1C). Second, about 51.2% of BUSCO genes were duplicated (Table S7). Third, there was a clear two-to-one relationship in 25,670 orthologous gene triplets (58.70% of all of the *S. o'connori* genes) between *S. o'connori* and *D. rerio*, which was comparable with those in *C. carpio* (45.83%), *C. auratus* (53.45%), and *S. grahami* (67.74%) (Table S9). Fourth, we identified more than 16 Hox clusters in the *S. o'connori* genome (Figure S1), which was twice the number of Hox clusters of the *D. rerio* genome (Málaga-Trillo and Meyer, 2001).

To identify duplicated contigs in our final assembly, contig sequences were aligned to each other to identify WGD-generated contig pairs, namely, R1 and R2 contigs (Table S10). Our analyses identified 36,255 genes in R1 and R2 contigs and defined 16,264 gene pairs (89.72% of all the genes across both contigs, Table 1) as ohnologs (Berthelot et al., 2014), which suggested that synteny was highly conserved between these duplicated genomes. However, due to their high similarities (see the similarity analysis below), we could not accurately cluster R1 and R2 into two subgenomes using Hi-C (high-through chromosome conformation capture) read mapping. To investigate the genome synteny to other species, we created pseudo-chromosomes by organizing R1 and unpaired contigs using whole-genome Hi-C sequencing data to represent an "averaged" contig order for *S. o'connori* chromosomes. As a result, we constructed 24 pseudo-chromosomes for *S. o'connori* (Figures 1D and S2). We found that these pseudo-chromosomes exhibited a high degree of synteny with the chromosomes of *D. rerio* without obvious inter-chromosomal re-arrangements. A total of 338 collinearity blocks were identified, with an average block size of 3.07 and

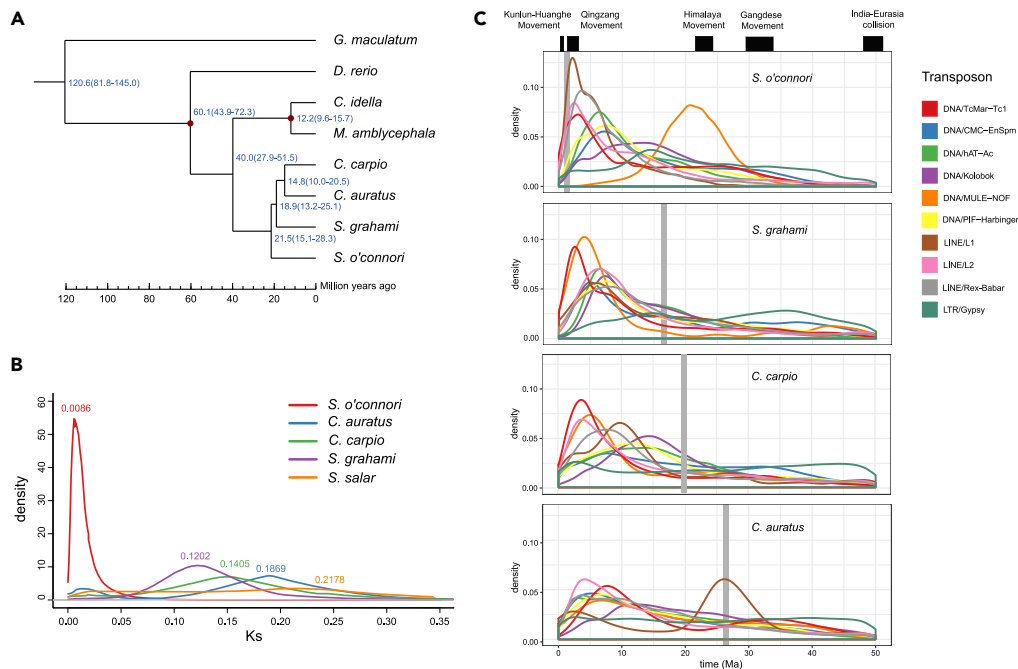


Figure 2. Whole-Genome Duplication and Transposon Burst in the *Schizothorax o'connori* Genome

(A) The time estimation for species divergence and whole-genome duplication for *Schizothorax o'connori* and evolutionarily close species. The time is estimated by 95% interval of the estimation.

(B) The Ks distribution of ohnologs for tetraploid fish species, including *Schizothorax o'connori*, *Carassius auratus*, *Cyprinus carpio*, *Sinocyclocheilus grahami*, and *Salmo salar*.

(C) The distribution of the insertion time of dominant transposon elements in *Schizothorax o'connori*, *Carassius auratus*, *Cyprinus carpio*, and *Sinocyclocheilus grahami*. The gray bars illuminated the time of subgenome divergence for species.

4.02 Mb for *S. o'connori* and *D. rerio* (Figure S3), respectively, suggesting that the duplicated genome of *S. o'connori* retained collinear conservations with the *D. rerio* genome after the WGD event.

Recent Subgenome Divergence for *S. o'connori*

Ohnologs identified from R1 and R2 sequences were used to determine the divergence time for the subgenomes of *S. o'connori*, *C. carpio*, *C. auratus*, and *S. grahami*. Using the 16,264 ohnologous gene pairs between the R1 and R2 contigs of *S. o'connori*, we identified 30,026 fourfold synonymous third-codon transversion (4DTv) sites. These gene pairs were used to reconstruct a phylogeny of these fish and to estimate divergence times between the sub-genomes of these species. Based on our phylogenetic analysis of orthologous genes after gene clustering in these related fish species (Tables S11 and S12 and Figures S4 and S5), all the fish subjected to 4R WGDs diverged from their most recent common ancestor (MRCA) about 21.5 mya (Figure 2A). The divergence times between *C. carpio* and *C. auratus* were about 10.0–20.5 mya, which was consistent with previous studies (Xu et al., 2014; Yang et al., 2016).

The divergence time for subgenomes was then estimated using the synonymous site substitutions (Ks) from ohnologs. The distribution of ohnolog Ks indicated that the peak Ks for *S. o'connori* subgenomes was 0.0086 (Figures 2B and Table S13), which is significantly smaller than those for other genomes. The smaller Ks value indicated that *S. o'connori* is a young tetraploid compared with other reported natural tetraploid fish species (Figure 2B). On the basis of a Ks rate of 3.51×10^{-9} substitutions per synonymous site per year (Xu et al., 2014), we estimated that the divergence time for the subgenomes of *S. o'connori* was about 1.23 mya.

Dynamics of Transposons in the *S. o'connori* Genome

WGDs are usually closely related to transposon dynamics (Rodriguez and Arkhipova, 2018), because WGDs and transposon bursts occurred simultaneously in many polyploid genomes (Belyayev, 2014). More than

	Genome Assembly
Total length (bp)	2,067,143,420
Sequence number	25,126
Max sequence	9,078,463
Number \geq 2000	24,148
N50 length	241,912
Gene number	43,731
Singletons	3,899
Ohnologs	25,670

Table 1. Assembly Statistics of the *Schizothorax O'connori* Genome and Annotations

50% of the *S. o'connori* genome is associated with transposons (Table S14). DNA (31.02%) and long interspersed nuclear elements (LINE) (7.6%) were the two most abundant transposons (Table S15). The relative abundance of DNA transposons in the *S. o'connori* genome was lower than that of *D. rerio*, but the relative abundance of LINE elements (XXX%) in the *S. o'connori* genome was significantly higher than those of other cyprinid fish (Tables S16 and S17).

Substitutions identified in a given genome relative to a consensus sequence can indicate the timing of transposon expansions (Albertin et al., 2015). We found that all tetraploid fish underwent a long period of persistent transposon activity along the Himalayan movement, followed by a recent burst around 3 mya (Figures 2C and S6). Although DNA transposons, especially for TcMar-Tc1 and MULE-NOF, mainly contributed to the latest transposon expansion (~3 mya) in *C. carpio*, *C. auratus*, and *S. grahami*, we observed a significant LINE element expansion, especially for L1, L2, and Rex-Babar, in the recent transposon expansion of the *S. o'connori* genome and L1 expansion with the 4R of *C. auratus* (Figure 2C). Meanwhile, we found that many transposons expanded before the WGD event in *S. o'connori* (Figures 2C and S7), suggesting that the expansion of those transposons occurred in their diploid ancestor lineages.

Transposon Dynamics May Have Prompted the Re-diploidization of the *S. o'connori* Genome

The sequence identities of duplicated genome regions reflect the degree of re-diploidization in polyploid genomes (Lien et al., 2016). In the *S. salar* genome that duplicated about 80–103 mya (Berthelot et al., 2014; Lien et al., 2016; Macqueen and Johnston, 2014), only 25.6% of high-scoring segment pairs (HSPs) exhibit more than 90% identity, and the peak HSP identity was 74.9% across the whole genome. However, in the *S. o'connori* genome, 98% of HSPs had more than 90% identity and the peak HSP identity was as high as 97.8% (Figure S8 and Table S18). This was consistent with our aforementioned analysis indicating that *S. o'connori* was a young tetraploid genome. Despite high sequence identities between HSP pairs, we identified 271,251 sequence insertions by direct sequence comparison between duplicated genomic regions (Figure 3A and Table S19). As shown in Figures 3A and 3B, these insertions caused obvious sequence divergence between duplicated genomes, leading to widespread gaps in genome comparisons.

Further analysis showed that the majority of those regions of insertions and deletions (>75%) between duplicated genomic regions were overlapped with transposons, suggesting that the transposon expansion played an important structural role in genomic sequence re-diploidization. More importantly, we found that *S. o'connori* HSP identity was significantly negatively correlated with the local density of transposons ($R^2 = -0.81$, p value < 0.001, Figure 3C), which indicated that transposon insertions may have prompted the differentiation of duplicated sequences. Indeed, we found that the mean SNP density within the 500 bases surrounding transposons was significantly higher than the SNP density in remote and random genomic regions (Figure 3D). The enrichments of SNPs around transposon insertion loci implied that transposon insertions might have induced mutagenesis and facilitated the mutation accumulation in nearby genomic regions.

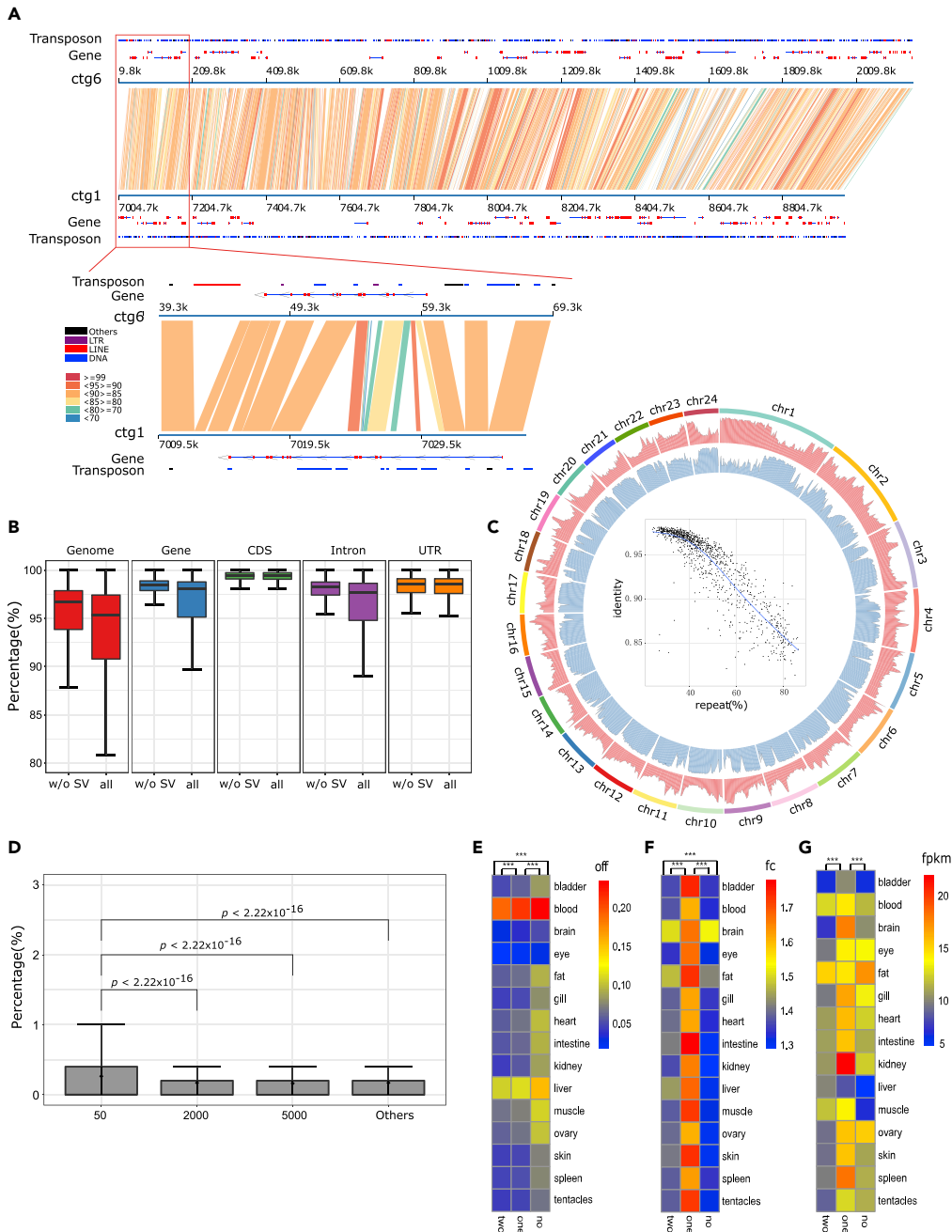


Figure 3. The Sequence Divergence and Structural Variants between Homologous Genomes after the Whole-Genome Duplication in *Schizothorax O'connori*

(A) The sequence comparison between ctg1 and ctg6 to illustrate the conserved sequence and transposon insertion after the duplication.

(B) The estimated similarity of HSP and regions for whole-genome, gene, coding sequence (CDS), intron, and untranslated regions (UTR).

(C) The distribution of sequence divergence and density of transposons along the pseudo-chromosomes for *Schizothorax o'connori*.

(D) The SNP density in the adjacent genomic sequences around transposon insertion sites.

(E–G) The on-and-off (E), fold changes (F), and FPKM (G) to illustrate the regulation of transposon insertion to the gene expression patterns.

Figure 3. Continued

Similarities and SNP densities for (B) and (D) are shown in boxplot with error bar for minimal and maximal value. The statistics for (D–G) were performed with Student t test, and *, **, and *** represent statistical p value smaller than 0.05, 0.01, and 0.001, respectively.

Transposons May Have Altered the Expression of Duplicated Genes in the *S. o'connori* Genome

Transposon expansions can even alter model genes in a genome (Bourque et al., 2018). We found that roughly 7.8% of all insertions and deletions were located in exon regions of gene loci. Transposon insertions around genes might have influenced the regulation of gene expression and even the sequences and functions of the encoded proteins. To investigate the possible influence of transposon bursts on the regulation of gene expression, the expression patterns of genes with transposons were compared with those of genes without transposon insertions. Based on their genomic coordinates, we categorized ohnologs into three groups, which represented two (T), one (O), and no (N) genes that overlapped with transposons. We first investigated how transposons affected on-and-off gene expression patterns. To this end, the off-gene ratio (number of silenced genes/total number of genes) was calculated for groups in all the tissues tested by RNA-seq. We found that the off-ratios differed significantly between the T, O, and N groups. The off-gene ratio was the lowest in the T group, followed by the O and N groups (Figure 3E), suggesting that transposon insertions could activate the expression of nearby genes. Then, we compared the normalized gene expression (fragments per kilobase million [FPKM]) and fold change (FC) levels within ohnologs between the T, O, and N groups. We found that the FPKM and FC in group O (only one of an ohnolog pair overlapped with a transposon) were significantly higher than those in groups T or N (Figures 3F and 3G).

Co-expression Patterns of Ohnologs

Gene duplications are important resources for genome evolution and provide potential adaptation for environmental changes because redundancy may alleviate selective pressure and prompt ohnolog differentiation and neofunctionalization (Cheng et al., 2018). The expression patterns of *S. o'connori* ohnologs were analyzed and compared with orthologs from *D. rerio* (Figure 4A, Tables S5 and S20). Based on gene expression profiles in 12 *S. o'connori* tissues, a weighted gene co-expression network analysis (WGCNA) was used to cluster 34,628 genes into 26 functional modules, with an average weight of 0.294 (Figure S9). Using identical types of tissues and RNA-seq data from *D. rerio*, WGCNA clustered 23,122 genes into 20 modules, with an average weight of 0.274 (Figure S10). Based on the correlations between gene models, we assigned expression modules to specific tissues. Interestingly, we identified four and five expression modules that were associated significantly with skin and gill tissues, respectively, which was higher than that of the expression modules from *D. rerio* (Figures 4B and S11). Both gill and skin are essential tissues to sense environmental changes, and we attribute the larger numbers of gene expression modules to possible complicated and elaborate gene expression regulation in gill and skin tissues to external environment factors, such as low temperature of water or UV radiation.

Meanwhile, we identified 232 genes in the core gill expression module that were shared between *S. o'connori* and *D. rerio* (Figure S12). We also identified 843 and 762 species-specific genes in the core gill modules of *S. o'connori* and *D. rerio*, respectively, suggesting that co-expression networks in gill tissues of these two species exhibited obvious differentiation. A Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of the species-specific genes for *S. o'connori* gill showed that most of these genes were associated with the adherens junction, vascular endothelial growth factor (VEGF) signaling, and mucin-type O-glycan biosynthesis pathways (Table S21). The adherens junction is needed for both strong cell-cell adhesion and rapid cell-cell contact remodeling during development and wound healing (Meng and Takeichi, 2009). The VEGF signaling pathway is specific to the vascular endothelium, is crucial in both physiologic and pathologic angiogenesis, and may play important roles in the formation of the unique microridges of gill tissues (Eichmann and Simons, 2012; Jing et al., 2004). In addition, roles for mucin-type O-glycans have been well documented for many cell surface and secreted proteins that modulate recognition, adhesion, and communication events occurring between cells and their surrounding environments (Tran and Ten Hagen, 2013). Mucin-type O-glycan biosynthesis generates the primary constituents of mucins that are expressed on various mucosal sites

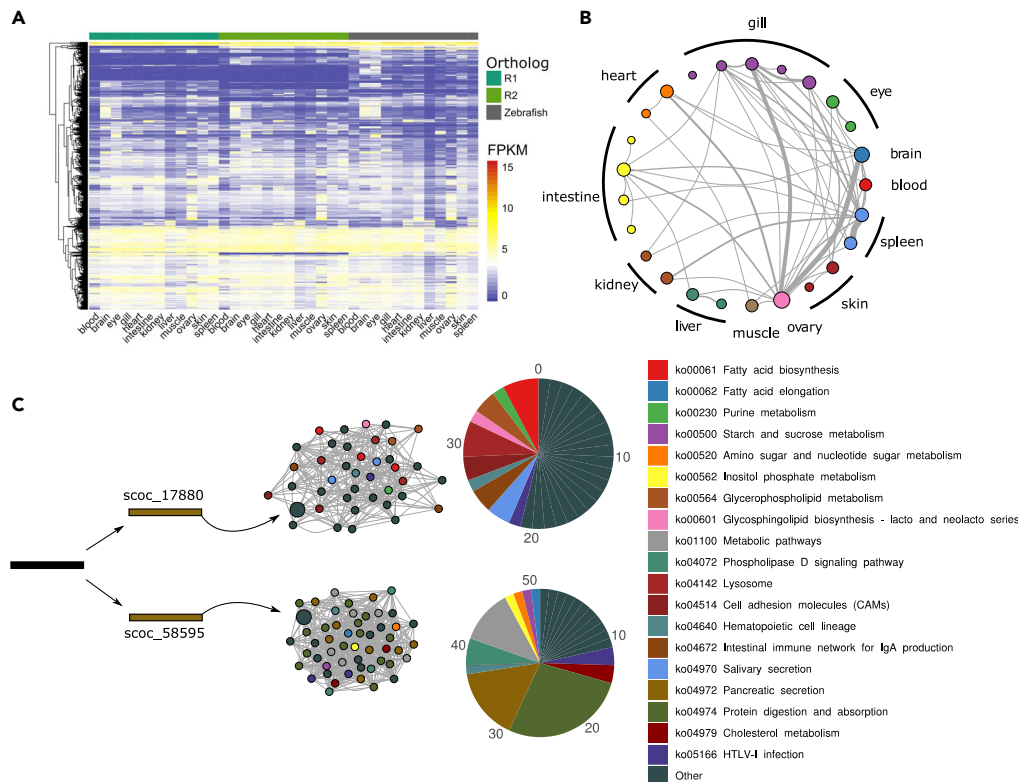


Figure 4. The Sequence and Functional Re-diploidization in the *Schizothorax o'connori* Genome

(A) The gene expression heatmap of ohnologs in *Schizothorax o'connori* and *Danio rerio*.

(B) The distribution of ohnolog pairs that were involved in distinct co-expression modules.

(C) An example for one ohnolog pair of SCOC_17,880 and SCOC_58,595 that was involved in the different co-expression networks and the biological KEGG pathway enrichments for neighboring co-expressed genes.

of the body (Jin et al., 2015), and they may play an important role in protecting fish from mechanical injury in rapid flow.

Remarkably, we also found that over 3,000 ohnologs were separated into different expression modules, especially between gill, brain, spleen, and ovary modules (Figure 4B). Genes from one ohnolog pair clustered in different modules might reflect functional divergences in duplicated genes. One example was the ohnologs of the neurotrypsin genes (SCOC_17,880 and SCOC_58,595). These genes were found in different expression modules, and functional analysis revealed that genes with 1° connections to SCOC_17,880 and SCOC_58,595 in the co-expression network enriched into entirely different biological pathways (Figure 4C). However, the mechanism of functional divergence of duplicated genes and their contribution to environmental adaptation and speciation for Schizothoracinae need further gene functional analysis.

Positively Selected Genes for Environmental Adaptation

Genes with more non-synonymous substitutions than synonymous substitutions (i.e., Darwinian selection) were termed positively selected genes (PSGs). To identify signals of positive selection compared with other fish species, we identified 177 PSGs (Table S22) from *S. o'connori* in the genome using codeml with the branch site model in PAML (Yang, 2007). Our functional analysis showed that cell-cell adhesion, DNA repair, immune response, Fanconi anemia pathway, and the phagosome were all enriched in PSGs (Tables 2, S23, and S24). To identify natural positive selection between ohnologs, we compared ohnolog sequences using the KaKs_Calculator (Zhang et al., 2006). As a result, we identified 503 pairs of positively selected ohnologs (Table S25). An enrichment analysis of those genes showed that cell-cell adhesion, mismatched DNA binding, mismatch repair, nucleosome assembly, and homologous recombination were over-represented (Tables S23 and S24). Meanwhile, we also identified 8.6–8.8 million SNPs in the

ID	Symbol	Method	Pathway	Type
SCOC_29972	<i>brca2</i>	Ohnologs	Fanconi anemia pathway, homologous recombination	I
SCOC_48653	<i>topbp1</i>	Ohnologs	Homologous recombination	I
SCOC_29848	<i>Atm</i>	Population	Homologous recombination	I
SCOC_41124	<i>tdp1</i>	Codon	DNA repair	I
SCOC_43553	<i>igfbp3</i>	Population	p53 signaling pathway: insulin-like growth factor-binding	I
SCOC_43551	<i>igfbp1a</i>	Population	p53 signaling pathway: insulin-like growth factor-binding	I
SCOC_25202	<i>mthfd1b</i>	Population	One-carbon pool by folate	II
SCOC_39892	<i>mthfd1b</i>	Population	One-carbon pool by folate	II
SCOC_42360	<i>slc19a1</i>	Population	Vitamin digestion and absorption	II
SCOC_23906	<i>Ftcd</i>	Ohnologs	One-carbon pool by folate	II
SCOC_17594	<i>hmgcl</i>	Population	Synthesis and degradation of ketone bodies	III
SCOC_47651	<i>abat</i>	Codon	Carbohydrate metabolism: Butanoate metabolism, propanoate metabolism	III
SCOC_55492	<i>Hyi</i>	Codon	Carbohydrate metabolism: Glyoxylate and dicarboxylate metabolism	III
SCOC_03692	<i>plpp4</i>	Population	Lipid metabolism: Glycerophospholipid metabolism, glycerolipid metabolism	III
SCOC_17661	<i>etnpl</i>	Population	Lipid metabolism: Glycerophospholipid metabolism	III
SCOC_30552	<i>adprm</i>	Population	Lipid metabolism: Glycerophospholipid metabolism	III
SCOC_45459	<i>dgke</i>	Population	Lipid metabolism: Glycerophospholipid metabolism, glycerolipid metabolism	III
SCOC_15952	<i>dgat1a</i>	Codon	Lipid metabolism: Glycerolipid metabolism	III
SCOC_50230	<i>gk3p</i>	Codon	Lipid metabolism: Glycerolipid metabolism	III
SCOC_39893	<i>gpx1a</i>	Population	Lipid metabolism: Arachidonic acid metabolism	III
SCOC_46334	<i>cyp2ad2</i>	Population	Lipid metabolism: Arachidonic acid metabolism, linoleic acid metabolism	III
SCOC_11670	<i>degs1</i>	Population	Lipid metabolism: Sphingolipid metabolism	III

Table 2. The Putative Positive Selected Function Genes in *Schizothorax o'connori* Genome for DNA Repairing, Folate Transport and Metabolism, and Energy Metabolism

Types I, II, and III denote possible PSGs in the DNA repairing, folate transport and metabolism, and energy metabolism, respectively.

S. o'connori genome from 199 individuals in five populations (Figures 1A and Tables S26–S28). Selective sweep analysis was performed using whole-genome resequencing data from different high-altitude populations (3,100–4,700 m) (Table S26), and 194 genomic regions that harbored 306 PSGs (Table S29) were determined to exhibit different genetic diversity among these populations. Interestingly, those genes were enriched in fatty acid biosynthesis, mucin-type O-glycan biosynthesis, and ion channel activity.

DISCUSSION

In this work, we present a reference genome for *S. o'connori* using long read data from PacBio sequencing. The continuity of the genome was assessed, and the contig N50 length of the *S. o'connori* genome was substantially greater than those of the previously constructed *O. mykiss* (7.7 kb), *S. grahami* (29.3 kb), *S. salar* (57.6 kb), and *C. carpio* (68.4 kb) genomes (Berthelot et al., 2014; Lien et al., 2016; Xu et al., 2014; Yang et al., 2016; Chen et al., 2019). We also noticed that the contig N50 of *C. auratus* (816.8 kb) was higher than that of *S. o'connori* (Chen et al., 2019). The *C. auratus* genome was also assembled using long reads, and the allopolyploid nature of *C. auratus* genome may have facilitated the sequence assembly of subgenomes. To further evaluate the quality of assembly, high-quality NGS short reads were aligned to our assembled genome. More than 98% of all the reads were properly mapped upon the genome, and only 0.0042% of detected SNPs were homozygous, indicating that the genome sequence was highly accurate at

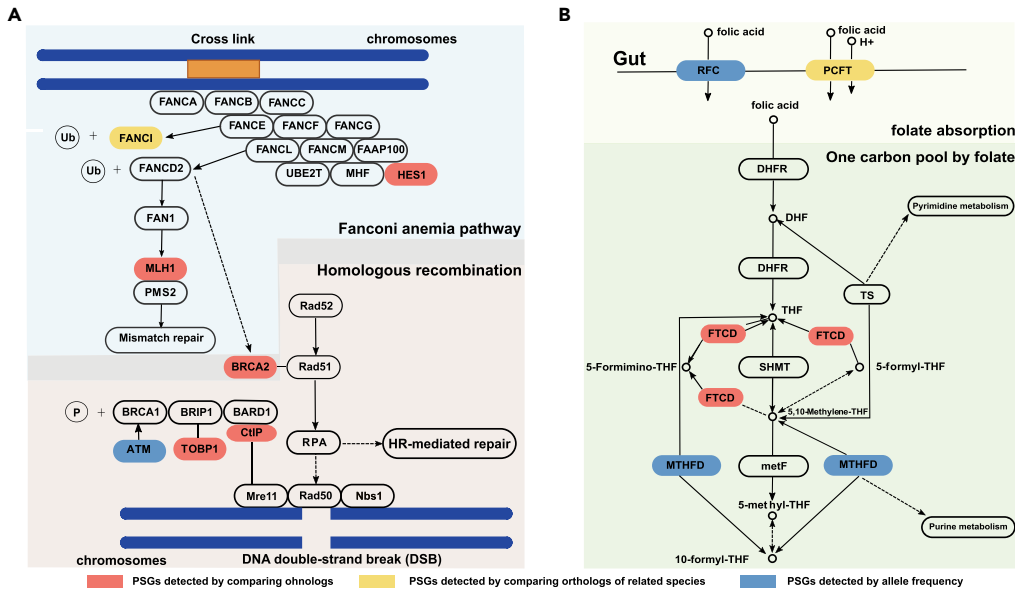


Figure 5. The Pathway that Illustrates the Positively Selected Genes in Folate Metabolism and DNA Repairs in *Schizothorax o'connori*

(A and B) The positively selected genes involved in folate transport and metabolism (A) and pathway for homologous recombination and fanconi anemia pathway (B). Genes that were positively selected from ohnolog pair comparisons and codon-, and population-based methods highlighted in red, yellow, and blue, respectively.

the base level (Table S30). In addition, the completeness of the genome was evaluated by BUSCO (Waterhouse et al., 2018) and the core eukaryotic genes mapping approach (CEGMA) (Parra et al., 2007). As a result, 2,586 BUSCO orthologs were present in the genome, of which 94.9% and 2.7% were completed and partial, respectively (Table S7). A total of 97.6% of the 248 core eukaryotic genes in the genome were also detected by CEGMA (Table S31).

Using gene annotation of the *S. o'connori* genome, we observed that the core histone genes, which included H1, H2A, H2B, H3, and H4, were significantly expanded in the *S. o'connori* genome, compared with other cyprinid species (Figure 1C). Histone genes play important roles in DNA packing and structural stability (Mariño-Ramírez et al., 2005), response to stress (e.g., to cold, Verleih et al., 2015), antibiotic stimulation (Lü et al., 2014; Noga et al., 2011), UV radiation (Pawlak and Deckert, 2007), and DNA-damage repairing (Schild-Poulter et al., 2003). The expanded histone genes might participate in genome stability and could contribute to adaptations to the harsh environments of cold stress, UV radiation, and DNA damage. However, the biological significances of these expanded histone genes in *S. o'connori* need further functional and evolutionary investigations.

The molecular mechanism of WGD in *S. o'connori* was not clear up to now. Although many tetraploid fish species in Cyprinidae lineages, such as *C. carpio* and *C. auratus*, are typical allotetraploid, *S. o'connori* exhibited obvious distinct genome features compared with *C. carpio* and *C. auratus*. First, a phylogenetic analysis showed that subgenomes of *C. carpio* and *C. auratus* were clustered into two clades across two species, indicating one *C. carpio* subgenome was evolutionarily closer to one *C. auratus* subgenome. However, the subgenomes of *S. o'connori* were clustered into a single clade (Figure S13). This result was consistent with the sequence identity between subgenomes of ~97.8% (Figure S8) and the divergence time of subgenomes for *S. o'connori* of 1.23 mya, which was much more recent than the divergence time of 14.4 mya for *C. carpio* and *C. auratus*. Second, we performed transposon analysis for homologous genomic regions. Transposons inserted into the same loci of homologous genomic regions should be ancient transposon duplications in diploid ancestors. We found that more than 50% of transposons were ancient in duplicated genomic regions (Table S32) and exhibited good synteny (Figure S14). In addition, we also performed a transposon family phylogenetic analysis to clarify the WGD origin of sterlet sturgeon (Du et al., 2020). We found that transposon families, including DNA/TcMar-Tc1, DNA/hAT-Ac, LINE/L2, and LINE/Rex-Babar, in the *S. o'connori* genome were monophyletic (Figures S15–S18), exhibiting no evidence of

an allopolyploid WGD. Therefore, our genomic analysis showed that the *S. o'connori* polyploidization was probably formed by autopolyploidization. However, this conclusion needs further confirmation by experiments such as cytogenetic studies.

We estimated 1.23 mya for the subgenome divergence time for *S. o'connori*. As we discussed, if *S. o'connori* was autotetraploid, the WGD of *S. o'connori* occurred around 1.23 mya. However, if *S. o'connori* was allotetraploid, the parental genomes were separated about 1.23 mya, and the 4R WGD of *S. o'connori* occurred even later. Regardless, *S. o'connori* is a young tetraploid compared with other reported natural tetraploid fish species, making this species a potential model for investigating the early post-WGD genome evolution and re-diploidization process. Meanwhile, the estimated subgenome divergence time was also coincident with the recent multiple rounds of uplift and deplanation of the QTP during the Tertiary period, leading to the overall height of the final plateau to ~3,000 to 3,500 m (Jiang and Li, 2014). In parallel with these plateau uplifts, Pleistocene glaciation began about 2.58 mya (Gibbard et al., 2010). The harsh environments jointly caused by these rapid QTP uplifts and Quaternary glaciation, including the arctic condition, nutritive deficiency, and high levels of UV exposure, might have been dominant factors driving the 4R WGD of *S. o'connori*. Interestingly, the historical effective population size of *S. o'connori* also declined from 1 to 2 mya but remarkably increased after about 0.9 mya (Figure S19). These results suggested that the ancestor of *S. o'connori* may have undergone a bottleneck effect after the rapid QTP uplifts and Quaternary glaciation, and a population expansion afterward (Figure S19).

We showed here that a genome-wide transposon burst occurred in the *S. o'connori* genome. Those transposons may have played important structural and functional roles in the genome evolution of *S. o'connori*. Our results showed that more than 75% of insertion and deletion differences between duplicated genomic regions were related to transposon elements (Figure 3A). Meanwhile, we analyzed base compositions around insertion sites and found an obvious TA-enriched pattern (Figure S20). Notably, the TA motif is a typical insertion site for DNA transposons, further confirming the contribution of transposon bursts to these sequence divergences. In addition, our gene expression analysis implied that asymmetric transposon insertions may have caused a differentiation in gene expression between ohnologs. These results illuminate that transposon insertions not only could have induced sequence mutagenesis between duplicated genomes but also prompted the differentiation of gene regulation and expression in ohnologous gene pairs.

WGD events double the number of genome-wide functional genes, generating massive redundant genes (Van de Peer et al., 2017). The duplication and expansion of functional genes may be crucial for environmental adaptation and species survival (Alix et al., 2017). Becoming a singleton is one fate of duplicated genes and might play an important role in maintaining cellular functions. Interestingly, we found that central genes in the main crossover pathway (ZMM pathway), namely, *msh4* and *msh5*, were both singletons. *Msh4* and *msh5* are important non-homology recombination suppressors during meiosis, and a singleton of *msh4* prevents non-homologous crossovers in the polyploid *Brassica napus* (Gonzalo et al., 2019). Therefore, these singleton genes in the ZMM pathways may have contributed to the stability of the *S. o'connori* genome after WGD, along the whole-genome of *S. o'connori*. Based on the ordering of genes in the syntenic blocks between homologous genomic regions (Figure S21), we found that about 10.99% of singletons could be identified using gene synteny information, which is slightly lower than that in *C. auratus* (11.7%) and *S. grahami* (18.59%) (Table S33). Considering that the WGD in the *S. o'connori* genome is much more recent, if *S. o'connori* is autotetraploid, our results showed that *S. o'connori* might have experienced rapid gene loss after the 4R WGD. However, if *S. o'connori* is allotetraploid, the gene loss might have occurred from the parental genomes.

We also identified putative naturally selected genes for *S. o'connori* through ortholog, ohnolog, and population comparisons (Tables 2, S22, S25, and S29). We performed functional analyses for PSGs based on various methods separately. Notably, many PSGs identified from ortholog and ohnolog comparison were involved in the biological function of DNA repair, cellular response to DNA damage stimulus, fanconi anemia, homologous recombination, and immune response (Table 2 and Figure 5A). In addition to the identified selected genes in aforementioned biological functions, a population comparison also detected many PSGs contributing to folate and energy metabolism (Table 2 and Figure 5B).

Homologous recombination and fanconi anemia pathway play an important role in the repair of double-stranded DNA and the removal of DNA interstrand cross-links (ICLs) (Li and Heyer, 2008; Rodríguez and

D'Andrea, 2017). The deleterious UV radiation exposure caused by rapid plateau uplifts might damage double-stranded DNA and induce ICLs (Cecchini et al., 2005). The *brca2* gene, crucial for the repair of damaged DNA and for stably maintaining cellular genetic information (Wooster et al., 1995), was identified from ohnolog comparisons (dN/dS of 1.08) (Figure 5A). The *tpd1* gene, participating in the repair of free-radical-mediated DNA double-strand breaks (Dong et al., 2012), was identified using a population comparison. The natural positive selection of key genes that participate in DNA repair might contribute to the maintenance of integrity of genetic information in the *S. o'connori* genome when exposed to intense UV radiation and dramatic temperature fluctuations.

We also identified three PSGs in the folate transport and metabolism pathways. Solute carrier family 19 member 1 (*slc19a1*) and methylenetetrahydrofolate dehydrogenase (*mthfd*) were identified from a population comparison, and formiminotetrahydrofolate cyclodeaminase (*fctd*) was identified from an ohnolog comparison (dN/dS of 1.35) (Figure 5B). Folate is a vitamin involved in the one-carbon transfer reaction, DNA synthesis, and methylation (Lucock, 2000). In addition, a previous association study of high-altitude populations revealed that methylenetetrahydrofolate reductase (*mthfr*) was selected in Tibetans, indicating the importance of folate in UV protection and DNA damage repair after UV exposure (Yang et al., 2017). Folate is vulnerable to degradation by UV radiation (Steindal et al., 2007), and increased exposure to solar UV radiation reduces the efficacy of folic acid supplementation (Borradale et al., 2014). *Slc19a1* plays a crucial role in folate absorption (Chiao et al., 1997) and intracellular folate concentration regulation (Hou and Matherly, 2014). The genes *mthfd* and *fctd* are important for folate metabolism and in the "one-carbon pool by folate" pathway for DNA biosynthesis (Watkins and Rosenblatt, 2012), and *mthfd* is also associated with gastric carcinogenesis (Wang et al., 2007) and Down syndrome (Scala et al., 2006). Natural selective pressure on *mthfd* and *fctd* might improve folate metabolism and DNA synthesis after exposure to UV radiation in *S. o'connori*.

Ketone bodies have been implicated not only in energy metabolism but also in lipogenesis (Chavan et al., 2016). We found that 3-hydroxy-3-methylglutaryl-CoA lyase (*hmgcl*), a key gene in ketogenesis and metabolism regulation during fasting or starvation (Kang et al., 2015), was under positive selection in populations living at different altitudes. This gene might play an important role in the environmental adaptation of *S. o'connori* to nutrient deficiency on the QTP. Carbohydrate metabolism provides energy through a series of complex chemical reactions, and lipid metabolism involves the fat breakdown and storage for energy. Of all PSGs, we identified two and nine PSGs that were associated with carbohydrate and lipid metabolism, respectively. Seven of these nine PSGs related to lipid metabolism were identified in a population comparison (Table S22, S25, and S29). The identified PSGs in those biological processes might reflect that *S. o'connori* underwent natural positive selection for energy metabolism and nutrient absorption to successfully adapt to ecological environments among populations.

Conclusion

Formed by the collision between India and Eurasia, the QTP is the youngest, largest, and highest plateau on Earth. Due to its increasing uplift during the last 50 Ma, organisms endemic to the QTP are useful models for investigating the underlying molecular mechanisms for adaptations to rigorous high-land environments. Many fish species in the subfamily Schizothoracinae are polyploids, and scientists believed that WGD played an important role in the environmental adaptation and speciation of Schizothoracinae.

S. o'connori is thought to be a representative tetraploid species in the subfamily Schizothoracinae. We assembled a reference genome for this species and confirmed its genomic features as a tetraploid fish. As this species has undergone a recent WGD (~1.23 mya), *S. o'connori* is an excellent model to study molecular mechanisms of early re-diploidization after WGD and environmental adaptations on the QTP. We demonstrated that transposon expansion can facilitate genome re-diploidization by prompting sequence mutagenesis and altering gene expression regulation, resulting in sequence divergence as well as function differentiations. In contrast to the tetraploid genome of *S. salar*, we did not observe massive inter-chromosomal rearrangements. We propose that the genome of *S. o'connori* represents an underway stage of re-diploidization for polyploidy after WGD and provides a golden opportunity to investigate post-WGD early genome evolution. To investigate putative functional genes contributing to the survival and environmental adaptation of *S. o'connori* to the harsh QTP environment, we identified PSGs through ortholog, ohnolog, and population comparisons and found that these

genes were involved in DNA repair, folate transport and metabolism, energy metabolism, and nutrient absorption.

Limitations of the Study

We confirmed genome features of the *S. o'connori* using reference genome sequences, but the molecular mechanism of WGD in *S. o'connori* was still not clear. The allotetraploid or autotetraploid nature of *S. o'connori* need further experiments. Meanwhile, we observed massive transposon expansion in the *S. o'connori* genome and found that insertions of transposon could induce the re-diploidization both on sequence and gene expression. However, we did not observe large-scale genomic rearrangements as in *S. salar* genome. Whether those genomic features are shared among endemic Schizothoracinae fish on the QTP need more genome investigations.

Resource Availability

Lead Contact

Further information for the research should be directed to the Lead Contact Haiping Liu (luihappy@163.com).

Materials Availability

This study did not generate new unique reagents.

Data and Code Availability

Genome and transcriptome sequencing data were submitted to the National Center for Biotechnology Information (NCBI) BioProject number PRJNA557578. Genomic sequencing data from Pacbio SMRT, Hi-C, and Survey are available from the NCBI Short Read Archive as SRR10018089–SRR10018098, SRR10018101–SRR10018116, and SRR9964271–SRR9964278, respectively. To facilitate the re-production of our results and codes re-use, we have released our scripts in the github: <https://github.com/DingDingFan/Whole-genome-duplication/tree/master/>.

METHODS

All methods can be found in the accompanying [Transparent Methods supplemental file](#).

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.isci.2020.101497>.

ACKNOWLEDGMENTS

This work was supported by the special finance of the Tibet Autonomous Region (No. XZNKY-2019-C-053), the Key Research and Development Projects in Tibet: Preservation of Characteristic Biological Germplasm Resources and Utilization of Gene Technology in Tibet (No. XZ202001ZY0016N), the National Natural Science Foundation of China of China (No. 31560144 and 31602207), and China Postdoctoral Science Foundation (2019M652724).

AUTHOR CONTRIBUTIONS

H.L. designed the research. H.L., S.X., Z.M., R.Y., C.Z., W.W., B.Z., Y.L., F.L., H.Z., Y.L., and S.X. collected samples and performed the experiments. S.X., Y.Z., Z.M., J.L., Y.Z., D.F., S.Z., H.L., D.W., and M.Z. performed genome assembly and chromosome comparisons. S.X., Y.Z., and D.F. performed repetitive element analyses and other bioinformatics analysis. S.X., D.F., and H.L. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: February 9, 2020

Revised: June 10, 2020

Accepted: August 19, 2020

Published: September 25, 2020

REFERENCES

- Albertin, C.B., Simakov, O., Mitros, T., Wang, Z.Y., Pungor, J.R., Edsinger-Gonzales, E., Brenner, S., Ragsdale, C.W., and Rokhsar, D.S. (2015). The octopus genome and the evolution of cephalopod neural and morphological novelties. *Nature* 524, 220–224.
- Alix, K., Gerard, P.R., Schwarzacher, T., and Heslop-Harrison, J.S.P. (2017). Polyploidy and interspecific hybridization: partners for adaptation, speciation and evolution in plants. *Ann. Bot.* 120, 183–194.
- Belyayev, A. (2014). Bursts of transposable elements as an evolutionary driving force. *J. Evol. Biol.* 27, 2573–2584.
- Berthelot, C., Brunet, F., Chalopin, D., Juanchich, A., Bernard, M., Noel, B., Bento, P., Da Silva, C., Labadie, K., Alberti, A., et al. (2014). The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nat. Commun.* 5, 3657.
- Borradaile, D., Isenring, E., Hacker, E., and Kimlin, M.G. (2014). Exposure to solar ultraviolet radiation is associated with a decreased folate status in women of childbearing age. *J. Photochem. Photobiol. B Biol.* 131, 90–95.
- Bourque, G., Burns, K.H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., Imbeault, M., Izsvak, Z., Levin, H.L., Macfarlan, T.S., et al. (2018). Ten things you should know about transposable elements. *Genome Biol.* 19, 199.
- Cecchini, S., Masson, C., La Madeleine, C., Huels, M.A., Sanche, L., Wagner, J.R., and Hunting, D.J. (2005). Interstrand cross-link induction by UV radiation in bromodeoxyuridine-substituted DNA: dependence on DNA conformation. *Biochemistry* 44, 16957–16966.
- Chavan, R., Feillet, C., Costa, S.S.F., Delorme, J.E., Okabe, T., Ripperger, J.A., and Albrecht, U. (2016). Liver-derived ketone bodies are necessary for food anticipation. *Nat. Commun.* 7, 10580.
- Chen, Z., Omori, Y., Koren, S., Shirokiya, T., Kuroda, T., Miyamoto, A., Wada, H., Fujiyama, A., Atsushi, T., Suiyuan, Z., et al. (2019). De novo assembly of the goldfish (*Carassius auratus*) genome and the evolution of genes after whole-genome duplication. *Sci. Adv.* 5, eaav0547.
- Cheng, F., Wu, J., Cai, X., Liang, J., Freeling, M., and Wang, X. (2018). Gene retention, fractionation and subgenome differences in polyploid plants. *Nat. Plants* 4, 258–268.
- Chiao, J.H., Roy, K., Tolner, B., Yang, C.-H., and Sirotnak, F. (1997). RFC-1 gene expression regulates folate absorption in mouse small intestine. *J. Biol. Chem.* 272, 11165–11170.
- Chin, C.-S., Peluso, P., Sedlazeck, F.J., Nattestad, M., Concepcion, G.T., Clum, A., Dunn, C., O'Malley, R., Figueroa-Balderas, R., Morales-Cruz, A., et al. (2016). Phased diploid genome assembly with single-molecule real-time sequencing. *Nat. Methods* 13, 1050.
- Dai, Y., and Han, H. (2018). Karyological analysis of two species in the subfamily Schizothoracinae (Cypriniformes: Cyprinidae) from China, with notes on karyotype evolution in Schizothoracinae. *Turk. J. Fish. Aquat. Sci.* 18, 175–186.
- Dong, C., Beecham, A., Wang, L., Blanton, S.H., Rundek, T., and Sacco, R.L. (2012). Follow-up association study of linkage regions reveals multiple candidate genes for carotid plaque in Dominicans. *Atherosclerosis* 223, 177–183.
- Du, K., Stock, M., Kneitz, S., Klopp, C., Woltering, J.M., Adolphi, M.C., Feron, R., Prokopov, D., Makunin, A., Kichigin, I., et al. (2020). The sterlet sturgeon genome sequence and the mechanisms of segmental rediploidization. *Nat. Ecol. Evol.* 4, 841–852.
- Eichmann, A., and Simons, M. (2012). VEGF signaling inside vascular endothelial cells and beyond. *Curr. Opin. Cell Biol.* 24, 188–193.
- Gibbard, P.L., Head, M.J., Walker, M.J.C., and Stratigraphy, T.S.o.Q. (2010). formal ratification of the quaternary system/period and the Pleistocene series/epoch with a base at 2.58 Ma. *J. Quat. Sci.* 25, 96–102.
- Gonzalo, A., Lucas, M.-O., Charpentier, C., Sandmann, G., Lloyd, A., and Jenczewski, E. (2019). Reducing MSH4 copy number prevents meiotic crossovers between non-homologous chromosomes in *Brassica napus*. *Nat. Commun.* 10, 2354.
- Guo, X.-Z., Zhang, G.-R., Wei, K.-J., Yan, R.-J., Ji, W., Yang, R.-B., Wei, Q.-W., and Gardner, J.P. (2016). Phylogeography and population genetics of *Schizothorax o'connori*: strong subdivision in the Yarlung Tsangpo River inferred from mtDNA and microsatellite markers. *Sci. Rep.* 6, 29821.
- Hou, Z., and Matherly, L.H. (2014). Biology of the major facilitative folate transporters SLC19A1 and SLC46A1. *Curr. Top. Membr.* 73, 175–204.
- Jiang, X.D., and Li, Z.X. (2014). Seismic reflection data support episodic and simultaneous growth of the Tibetan Plateau since 25 Myr. *Nat. Commun.* 5, 5453.
- Jin, C., Padra, J.n.T.s., Sundell, K., Sundh, H., Karlsson, N.G., and Linden, S.K. (2015). Atlantic salmon carries a range of novel O-glycan structures differentially localized on skin and intestinal mucins. *J. Proteome Res.* 14, 3239–3251.
- Jing, F., KangCheng, P., and Deng, T.H. (2004). A scanning electron microscopic observation of the gill of *Schizothorax (Schizothorax) prenanti* (tchang). *J. Sichuan Agric. Univ.* 22, 196–198.
- Kang, H.-B., Fan, J., Lin, R., Elf, S., Ji, Q., Zhao, L., Jin, L., Seo, J.H., Shan, C., and Arbiser, J.L. (2015). Metabolic rewiring by oncogenic BRAF V600E links ketogenesis pathway to BRAF-MEK1 signaling. *Mol. Cell* 59, 345–358.
- Lü, A., Hu, X., Wang, Y., Shen, X., Li, X., Zhu, A., Tian, J., Ming, Q., and Feng, Z. (2014). iTRAQ analysis of gill proteins from the zebrafish (*Danio rerio*) infected with *Aeromonas hydrophila*. *Fish Shellfish Immunol.* 36, 229–239.
- Li, J., Zhou, S., Zhao, Z., and Zhang, J. (2015). The Qingzang movement: the major uplift of the Qinghai-Tibetan Plateau. *Sci. China Earth Sci.* 58, 2113–2122.
- Li, X., and Heyer, W.-D. (2008). Homologous recombination in DNA repair and DNA damage tolerance. *Cell Res.* 18, 99.
- Lien, S., Koop, B.F., Sandve, S.R., Miller, J.R., Kent, M.P., Nome, T., Hvidsten, T.R., Leong, J.S., Minkley, D.R., Zimin, A., et al. (2016). The Atlantic salmon genome provides insights into rediploidization. *Nature* 533, 200–205.
- Lucock, M. (2000). Folic acid: nutritional biochemistry, molecular biology, and role in disease processes. *Mol. Genet. Metab.* 71, 121–138.
- Málaga-Trillo, E., and Meyer, A. (2001). Genome duplications and accelerated evolution of Hox genes and cluster Architecture in teleost Fishes1. *Am. Zool.* 41, 676–686.
- Ma, B.-S., Xie, C.-X., Huo, B., Yang, X.-F., and Huang, H.-P. (2010). Age and growth of a long-lived fish *Schizothorax o'connori* in the Yarlung Tsangpo River, Tibet. *Zool. Stud.* 49, 749–759.
- Ma, B., Wang, S., Shao, J., Li, D., Xie, C., Liu, X., and Huo, B. (2020). Population resources and fishery conservation policies of *Schizothorax oconnori* in the Yarlung Zangbo River. *J. Fish. Sci. China* 27, 106–117.
- MacKintosh, C., and Ferrier, D.E.K. (2017). Recent advances in understanding the roles of whole genome duplications in evolution. *F1000Res.* 6, 1623.
- Macqueen, D.J., and Johnston, I.A. (2014). A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. *Proc. Biol. Sci.* 281, 20132881.
- Mariño-Ramírez, L., Kann, M.G., Shoemaker, B.A., and Landsman, D. (2005). Histone structure and nucleosome stability. *Expert Rev. Proteomics* 2, 719–729.
- Meng, W., and Takeichi, M. (2009). Adherens junction: molecular architecture and regulation. *Cold Spring Harb. Perspect. Biol.* 1, a002899.
- Noga, E.J., Borrón, P.J., Hinshaw, J., Gordon, W.C., Gordon, L.J., and Seo, J.-K. (2011). Identification of histones as endogenous antibiotics in fish and quantification in rainbow trout (*Oncorhynchus mykiss*) skin and gill. *Fish Physiol. Biochem.* 37, 135–152.
- Pan, G., Wang, L., Li, R., Yuan, S., Ji, W., Yin, F., Zhang, W., and Wang, B. (2012). Tectonic evolution of the Qinghai-Tibet Plateau. *J. Asian Earth Sci.* 53, 3–14.
- Parra, G., Bradnam, K., and Korf, I. (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23, 1061–1067.
- Pawlak, S., and Deckert, J. (2007). Histone modifications under environmental stress. *Biol. Lett.* 44, 65–73.
- Qi, D., Chao, Y., Guo, S., Zhao, L., Li, T., Wei, F., and Zhao, X. (2012). Convergent, parallel and correlated evolution of trophic morphologies in the subfamily Schizothoracinae from the Qinghai-Tibetan plateau. *PLoS One* 7, e34070.

- Rodríguez, A., and D'Andrea, A. (2017). Fanconi anemia pathway. *Curr. Biol.* *27*, R986–R988.
- Rodríguez, F., and Arkhipova, I.R. (2018). Transposable elements and polyploid evolution in animals. *Curr. Opin. Genet. Dev.* *49*, 115–123.
- Scala, I., Granese, B., Sellitto, M., Salomè, S., Sammartino, A., Pepe, A., Mastroiacovo, P., Sebastio, G., and Andria, G. (2006). Analysis of seven maternal polymorphisms of genes involved in homocysteine/folate metabolism and risk of Down syndrome offspring. *Genet. Med.* *8*, 409.
- Schild-Poulter, C., Shih, A., Yarmowich, N.C., and Haché, R.J.G. (2003). Down-regulation of histone H2B by DNA-dependent protein kinase in response to DNA damage through modulation of octamer transcription factor 1. *Cancer Res.* *63*, 7197–7205.
- Steindal, A.H., Porojnicu, A.C., and Moan, J. (2007). Is the seasonal variation in cancer prognosis caused by sun-induced folate degradation? *Med. Hypotheses* *69*, 182–185.
- Tran, D.T., and Ten Hagen, K.G. (2013). Mucin-type O-glycosylation during development. *J. Biol. Chem.* *288*, 6921–6929.
- Van de Peer, Y., Mizrahi, E., and Marchal, K. (2017). The evolutionary significance of polyploidy. *Nat. Rev. Genet.* *18*, 411–424.
- Verleih, M., Borchel, A., Krasnov, A., Rebl, A., Korytář, T., Kühn, C., and Goldammer, T. (2015). Impact of thermal stress on kidney-specific gene expression in farmed regional and imported rainbow trout. *Mar. Biotechnol.* *17*, 576–592.
- Wang, L., Ke, Q., Chen, W., Wang, J., Tan, Y., Zhou, Y., Hua, Z., Ding, W., Niu, J., Shen, J., et al. (2007). Polymorphisms of MTHFD, plasma homocysteine levels, and risk of gastric cancer in a high-risk Chinese population. *Clin. Cancer Res.* *13*, 2526–2532.
- Wang, Y., Lu, Y., Zhang, Y., Ning, Z., Li, Y., Zhao, Q., Lu, H., Huang, R., Xia, X., Feng, Q., et al. (2015). The draft genome of the grass carp (*Ctenopharyngodon idellus*) provides insights into its evolution and vegetarian adaptation. *Nat. Genet.* *47*, 625–631.
- Waterhouse, R.M., Seppey, M., Simão, F.A., Manni, M., Ioannidis, P., Kliuchnikov, G., Kriventseva, E.V., and Zdobnov, E.M. (2018). BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* *35*, 543–548.
- Watkins, D., and Rosenblatt, D.S. (2012). Update and new concepts in vitamin responsive disorders of folate transport and metabolism. *J. Inher. Metab. Dis.* *35*, 665–670.
- Wooster, R., Bignell, G., Lancaster, J., Swift, S., Seal, S., Mangion, J., Collins, N., Gregory, S., Gumbs, C., and Micklem, G. (1995). Identification of the breast cancer susceptibility gene BRCA2. *Nature* *378*, 789.
- Xu, P., Zhang, X., Wang, X., Li, J., Liu, G., Kuang, Y., Xu, J., Zheng, X., Ren, L., Wang, G., et al. (2014). Genome sequence and genetic diversity of the common carp, *Cyprinus carpio*. *Nat. Genet.* *46*, 1212–1219.
- Xu, Q., Zhang, C., Zhang, D., Jiang, H., Peng, S., Liu, Y., Zhao, K., Wang, C., and Chen, L. (2016). Analysis of the erythropoietin of a Tibetan Plateau schizothoracine fish (*Gymnocypris dobula*) reveals enhanced cytoprotection function in hypoxic environments. *BMC Evol. Biol.* *16*, 11.
- Yang, J., Chen, X., Bai, J., Fang, D., Qiu, Y., Jiang, W., Yuan, H., Bian, C., Lu, J., He, S., et al. (2016). The *Sinocyclocheilus cavefish* genome provides insights into cave adaptation. *BMC Biol.* *14*, 1.
- Yang, J., Jin, Z.B., Chen, J., Huang, X.F., Li, X.M., Liang, Y.B., Mao, J.Y., Chen, X., Zheng, Z., Bakshi, A., et al. (2017). Genetic signatures of high-altitude adaptation in Tibetans. *Proc. Natl. Acad. Sci. U S A* *114*, 4189–4194.
- Yang, Z. (2007). Paml 4: phylogenetic analysis by maximum Likelihood. *Mol. Biol. Evol.* *24*, 1586–1591.
- Yunfei, W., Bin, K., Qiang, M., and Cuizhen, W. (1999). Chromosome diversity of Tibetan fishes. *Zool. Res.* *20*, 258–264.
- Zan, Z., Liu, K., and Song, W. (1985). Tetraploid-hexaploid relationship in Schizothoracinae. *Acta Genetica Sinica* *12*, 137–142.
- Zhang, L. (2011). Study on the large scale artificial propagation of *Schizothorax o'connori*. *Freshw. Fish.* *41*, 88–91.
- Zhang, Q., Gou, W., Wang, X., Zhang, Y., Ma, J., Zhang, H., Zhang, Y., and Zhang, H. (2016). Genome resequencing identifies unique adaptations of Tibetan chickens to hypoxia and high-dose ultraviolet radiation in high-altitude environments. *Genome Biol. Evol.* *8*, 765–776.
- Zhang, Z., Li, J., Zhao, X.-Q., Wang, J., Wong, G.K.-S., and Yu, J. (2006). KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. *Genomics Proteomics Bioinformatics* *4*, 259–263.
- Zheng, B., Xu, Q., and Shen, Y. (2002). The relationship between climate change and Quaternary glacial cycles on the Qinghai–Tibetan Plateau: review and speculation. *Quat. Int.* *97*, 93–101.

Supplemental Information

Genome of Tetraploid Fish *Schizothorax o'connori*

Provides Insights into Early

Re-diploidization and High-Altitude Adaptation

Shijun Xiao, Zhenbo Mou, Dingding Fan, He Zhou, Ming Zou, Yu Zou, Chaowei Zhou, Ruibin Yang, Jiaqi Liu, Shilin Zhu, Yajuan Li, Yanchao Liu, Fei Liu, Wanliang Wang, Benhe Zeng, Hong Li, Di Wang, and Haiping Liu

Supplementary Tables

Table S1 Short name and data resource used throughout the supplementary. Related to Figure 1.

Short name	Scientific name	Source	Note
SCHoco	<i>Schizothorax o'connori</i>	this study	this study
GLYmac	<i>Glyptosternon maculatum</i>	GigaScience	http://gigadb.org/dataset/100489
ICTpun	<i>Ictalurus punctatus</i>	Ensemble	release-94
ASTmex	<i>Astyanax mexicanus</i>	Ensemble	release-94
DANrer	<i>Danio rerio</i>	Ensemble	release-94
MEGamb	<i>Megalobrama amblycephala</i>	GigaScience	http://gigadb.org/dataset/100305
Ctransposonide	<i>Ctenopharyngodon idella</i>	-	http://www.ncgr.ac.cn/grasscarp
CARaur	<i>Carassius auratus</i>	NCBI	GCF_003368295.1
CYPcar	<i>Cyprinus carpio</i>	NCBI	GCF_000951615.1
SINgra	<i>Sinocyclocheilus grahami</i>	NCBI	GCF_001515645.1
SALsal	<i>Salmo salar</i>	NCBI	GCF_000233375.1

Table S2 High throughput data generated in this study and used for the assembly of the *Schizothorax o'connori* genome. The coverage was calculated using an estimated genome size of 1.94 Gb. Related to Figure 1 and Table 1.

Pair-end libraries	Insert size (bp)	Total data (Gb)	Read length (bp)	Sequence coverage (x)
Illumina reads	350	295.2	150	152.16
Pacbio reads	-	141	-	72.68
total	-	436.2	-	224.85

Table S3 Statistics of 31-mer analysis. The genome size of *Schizothorax o'connori* was estimated to be ~1,938.3 Mb. Related to Figure 1 and Table 1.

K-mer	K-mer number	K-mer depth	Genome size (Mb)	Revised Genome size (Mb)	Repeat (%)
31	209,524,858,560	99	2,116.41	1,938.30	56.65

Table S4 Summary of genome assembly of *Schizothorax o'connori*. Related to Figure 1 and Table 1.

	Contig length (bp)	Contig number
Total	2,067,143,420	25,126
N50	241,912	2,299
N60	190,413	3,265
N70	143,270	4,514
N80	100,040	6,236
N90	53,827	8,984

Table S5 Summary of RNA-seq for *Schizothorax o'connori*. Related to Figure 4.

Sample	Total reads	Total Bases (bp)	GC (%)	Q20 (%)	Q30 (%)
brain1	54,568,572	8,185,285,800	44.60	96.57	91.57
brain2	54,687,332	8,203,099,800	44.78	96.41	91.22
brain3	46,970,020	7,045,503,000	45.51	96.50	91.37
bladder1	50,779,150	7,616,872,500	46.97	96.06	90.45
bladder2	48,186,636	7,227,995,400	46.89	95.23	88.79
bladder3	52,519,810	7,877,971,500	47.04	96.32	90.97
blood1	43,376,542	6,506,481,300	45.7	96.32	91.02
blood2	44,359,962	6,653,994,300	47.27	96.52	91.31
blood3	58,230,356	8,734,553,400	46.71	96.47	91.28

eye1	51,750,466	7,762,569,900	46.2	96.61	91.61
eye2	57,056,460	8,558,469,000	46.18	96.31	91.00
eye3	53,709,888	8,056,483,200	46.11	96.35	91.09
fat1	59,750,130	8,962,519,500	50.19	96.97	92.28
fat2	46,084,572	6,912,685,800	49.55	96.4	91.22
fat3	41,815,914	6,272,387,100	51.42	95.75	89.42
gill1	52,421,886	7,863,282,900	45.78	96.55	91.52
gill2	49,374,968	7,406,245,200	45.92	96.68	91.72
gill3	44,221,190	6,633,178,500	46.03	96.22	90.82
heart1	58,951,096	8,842,664,400	45.86	96.75	91.93
heart2	46,783,092	7,017,463,800	46.00	96.47	91.25
heart3	46,769,590	7,015,438,500	46.16	96.20	90.76
intestine1	60,350,596	9,052,589,400	45.95	96.56	91.57
intestine2	56,615,052	8,492,257,800	46.61	96.65	91.66
intestine3	46,353,418	6,953,012,700	47.05	96.53	91.40
kidney1	50,731,352	7,609,702,800	45.90	96.38	91.16
kidney2	44,966,134	6,744,920,100	45.62	96.59	91.56
kidney3	51,202,892	7,680,433,800	45.90	96.35	91.08
liver1	67,776,260	10,166,439,000	46.41	96.84	92.06
liver2	50,256,982	7,538,547,300	46.22	96.34	90.98
liver3	51,840,710	7,776,106,500	46.67	96.39	91.09
mucle1	60,483,060	9,072,459,000	47.91	96.56	91.5

mucle2	48,829,500	7,324,425,000	47.52	96.56	91.48
mucle3	48,228,046	7,234,206,900	48.34	96.16	90.56
ovary1	59,078,620	8,861,793,000	48.12	94.97	88.56
ovary2	46,406,332	6,960,949,800	48.11	95.96	90.35
ovary3	44,446,606	6,666,990,900	47.90	95.90	90.30
skin1	42,995,272	6,449,290,800	47.19	96.32	91.00
skin2	48,854,244	7,328,136,600	47.32	96.06	90.43
skin3	60,968,926	9,145,338,900	46.50	96.66	91.73
spleen1	63,788,050	9,568,207,500	46.08	96.69	91.82
spleen2	51,046,618	7,656,992,700	46.18	96.31	90.99
spleen3	57,616,588	8,642,488,200	46.24	96.36	91.10
tentacles1	69,319,130	10,397,869,500	45.55	96.16	90.76
tentacles2	50,091,782	7,513,767,300	47.71	96.39	91.09
tentacles3	52,597,910	7,889,686,500	47.18	96.48	91.29

Table S6 Summaries of genes of *Schizothorax o'connori* and related species. ORF means open reading frame. Related to Figure 1 and Table 1.

Species	Gene number	ORF (%)	Single exon (%)	Average transcript length (bp)	Average CDS length (bp)	Average exons per gene	Average exon length (bp)	Average intron length (bp)
<i>G. maculatum</i>	22,846	94.73	6.65	17,782.28	1,762.04	10.31	170.85	1,720.14
<i>A. mexicanus</i>	23,042	67.88	5.53	20,457.70	1,588.71	9.81	161.97	2,142.10
<i>D. rerio</i>	25,591	84.74	7.20	25,474.69	1,613.27	9.33	172.86	2,863.59
<i>C. idellus</i>	32,691	59.86	10.44	10,400.72	1,387.00	7.66	181.02	1,353.01
<i>S. grahami</i>	45,778	91.33	6.30	16,259.73	1,587.47	9.25	171.70	1,779.42
<i>C. carpio</i>	48,287	91.46	5.45	10,990.49	1,266.81	7.71	164.22	1,448.24
<i>S. salar</i>	48,334	85.31	6.71	89,751.94	1,568.95	8.68	180.86	11,706.71
<i>C. auratus</i>	52,870	98.22	7.11	14,029.56	1,709.58	9.67	176.86	1,421.59
<i>S. o'connori</i>	43,731	83.61	7.94	13,389.89	1,508.89	8.85	170.49	1,513.45

Table S7 Summaries of BUSCO assessments for the gene annotation of *Schizothorax o'connori* and related species. Related to Figure 1 and Table 1.

Species	Complete (%)	Complete single-copy (%)	Complete duplicated (%)	Fragmented (%)	Missing (%)
<i>G. maculatum</i>	98.3	94.8	3.5	0.7	1.0
<i>S. salar</i>	98.1	47.5	50.6	1.2	0.7
<i>D. rerio</i>	97.0	77.0	20.0	1.6	1.4
<i>A. mexicanus</i>	93.0	88.7	4.3	5.1	1.9
<i>C. idellus</i>	85.4	81.6	3.8	11.6	3.0
<i>C. carpio</i>	83.8	47.3	36.5	11.7	4.5
<i>S. o'connori</i>	92.1	40.9	51.2	4.3	3.6
<i>C. auratus</i>	99.3	30.2	69.1	0.3	0.4

Table S8 Summaries of functional annotations for *Schizothorax o'connori*. Related to Figure 1 and Table 1.

Database		Annotated Number	Annotated Percent(%)
NR		43,400	99.24
Swiss-Prot		40,909	93.55
KEGG		37,442	85.62
InterPro	All	43,675	99.87
	Pfam	38,337	87.67
	GO	41,331	94.51
Annotated		43,719	99.97
Total		43,731	-

Table S9 Summary of RBH (reciprocal best hit) orthologs and Ohnologues for each species. All orthologs should have an identity >60%, coverage > 50% and alignment score >200. For ohnologues, the identity differences between the first and the second-best hit were less than 10%.

Related to Figure 1 and Table 1.

Species	Orthology	Orthology (%)	Orthology of zebrafish Gene	Orthology of zebrafish gene (%)	Ohnolog	Ohnolog (%)	Total genes
<i>G. maculatum</i>	12,009	47.93	12,009	47.15	-	-	25,054
<i>C. auratus</i>	51,105	96.65	20,619	80.95	28,264	53.45	52,875
<i>C. idellus</i>	17,240	52.74	17,240	67.68	-	-	32,691
<i>C. carpio</i>	46,431	96.16	19,995	78.50	22,130	45.83	48,287
<i>M. amblycephala</i>	16,839	71.06	16,839	66.11	-	-	23,696
<i>S. o'connori</i>	42,348	96.84	19,103	75.00	25,670	58.70	43,731
<i>S. grahami</i>	44,912	98.11	20,407	80.12	31,010	67.74	45,778

Table S10 Summary of homoeologous regions resulted from whole genome duplication. Related to Figure 1 and Table 1.

Classification	Bases (bp)	Number	N50 (bp)
R1	834,176,523	5,379	256,341
R2	927,726,323	4,237	298,351
Unclass	305,240,574	15,510	55,679

Table S11 Genomes selected for gene family constructions in the *Schizothorax o'connori* genome project. Related to Figure 1.

Species	order	Rounds of WGD
<i>D. rerio</i>	Cypriniformes	3
<i>C. idellus</i>	Cypriniformes	3
<i>M. amblycephala</i>	Cypriniformes	3
<i>C. carpio</i>	Cypriniformes	4
<i>C. auratus</i>	Cypriniformes	4
<i>S. grahami</i>	Cypriniformes	4
<i>S. o'connori</i>	Cypriniformes	4
<i>G. maculatum</i>	Siluriformes	3

Table S12 Summary of gene family in *Schizothorax o'connori* and related species. Related to Figure 1.

Species	Total genes	Unclustered genes	Families	Unique families	Average genes per family
<i>D. rerio</i>	25,471	486	16,160	48	1.55
<i>M. amblycephala</i>	23,696	354	15,460	27	1.51
<i>C. idella</i>	32,691	5,406	17,738	239	1.54
<i>C. auratus</i>	52,875	557	17,575	57	2.98
<i>C. carpio</i>	48,287	3,124	19,107	446	2.36
<i>S. grahami</i>	45,778	1,316	18,384	89	2.42
<i>S. o'connori</i>	43,731	385	16,576	101	2.61
<i>G. maculatum</i>	25,054	5,382	17,976	441	1.09

Table S13 Divergence time estimation based on the Ks values of orthologs (ohnologs) between (sub)genomes. On the basis of a Ks rate of 3.51×10^{-9} substitutions per synonymous site per year and the obtained Ks value of 0.0086 of *S. o'connori*, we estimated that the latest WGD (4R) happened 1.23 million years ago. The Confidence Intervals were calculated using t.test function in R program with 95% confidence level. For *Salmo salar* we cited from previous study. Related to Figure 2.

Species	Ks peak	Divergence time (Ma)	Confidence Intervals	
<i>S. o'connori</i>	0.0086	1.23	1.15	1.31
<i>C. auratus</i>	0.1869	26.62	26.47	26.78
<i>C. carpio</i>	0.1405	20.02	19.82	20.22
<i>S. grahami</i>	0.1202	17.12	17.02	17.24
<i>S. salar</i>	0.2187	80	--	--

Table S14 Summaries of transposable element (transposon) annotations for *Schizothorax o'connori*. Related to Figure 2.

Method	Repeat Size	% of genome
Repeatmasker	533,537,803	25.81
RepeatModeler	903,565,528	43.71
RepeatScout	856,195,906	41.42
Total	1,039,571,554	50.29

Table S15 Summaries of transposon classifications according to each method for *Schizothorax o'connori*. Related to Figure 2.

Type	Rebase transposons		<i>De novo</i>		Combined transposons	
	Length (bp)	% in genome	Length (bp)	% in genome	Length (bp)	% in genome
DNA	341,560,285	16.52	605,805,104	29.31	641,238,766	31.02
LINE	116,765,649	5.65	147,181,163	7.12	157,195,043	7.60
SINE	9,223,933	0.45	14,065,750	0.68	14,307,778	0.69
LTR	56,239,694	2.72	79,008,023	3.82	88,886,613	4.30
Other	3,125	0.00	0	0.00	2,710	0.00
Unknown	0	0.00	132,197,018	6.40	132,197,018	6.40
Total	519,069,683	25.11	954,550,195	46.18	1,000,999,455	48.42

Table S16 Summary of transposons content in 14 genomes from Repbase (2018.12) and repeatScout result. The species with background orange color means the genome was tetraploid genome. Related to Figure 2.

Order	Species	DNA	LINE	LTR	SINE	Unknown	Total
Salmoniformes	<i>S. salar</i>	23.08	12.08	5.72	1.20	2.97	46.61
Salmoniformes	<i>O. mykiss</i>	24.91	13.19	5.09	1.61	4.65	48.36
Cypriniformes	<i>D. rerio</i>	45.87	4.42	5.94	1.67	0.56	57.61
Cypriniformes	<i>C. idella</i>	26.21	1.33	2.20	0.18	5.50	35.55
Cypriniformes	<i>M. amblycephala</i>	28.35	2.26	4.45	0.19	9.66	45.05
Cypriniformes	<i>S. grahami</i>	21.69	2.74	4.02	0.36	5.71	35.26
Cypriniformes	<i>C. auratus</i>	22.40	5.71	5.56	0.85	4.42	38.98
Cypriniformes	<i>C. carpio</i>	20.14	4.37	3.63	0.60	5.73	34.99
Cypriniformes	<i>S. o'connori</i>	28.04	7.36	4.21	0.62	3.15	43.17

Table S18 HSP similarity detected by LASTZ self-genome alignment for *Schizothorax o'connori* and *Salmo salar*. Related to Figure 3.

>=Similarity(%)	<i>S. o'connori</i>			<i>S. salar</i>		
	HSP	Bases	% (Aligned)	HSP	Bases	% (Aligned)
20	2,546,344	785,064,602	100.00%	7,087,842	1,162,403,509	100.00%
50	2,541,414	784,753,816	99.96%	7,018,452	1,159,904,473	99.79%
60	2,431,560	772,967,794	98.46%	6,425,960	1,119,035,522	96.27%
70	2,256,276	755,186,316	96.19%	4,735,508	955,544,663	82.20%
90	1,865,654	717,796,843	91.43%	793,106	282,555,870	24.31%
95	1,595,226	671,310,201	85.51%	392,872	175,177,305	15.07%
96	1,495,270	645,611,039	82.24%	331,564	157,259,638	13.53%
97	1,354,402	603,855,740	76.92%	267,196	139,908,672	12.04%
98	1,143,210	526,095,462	67.01%	206,324	121,209,138	10.43%
99	772,774	343,972,720	43.81%	144,814	96,745,885	8.32%
100	278,296	47,559,101	6.06%	64,936	17,818,418	1.53%

Table S19 Insertion and deletion variants identification for the R1 and R2. Indel means insert and deletion between in R1 and R2. SNP means substitution between WGD fragments. Small indel means deletion/insertions in WGD aligned blocks ranged from 1 to 159 bp. Related to Figure 3.

Variation Type	Num	Bases	Percentage	Average happen size (bp)
Indel	271,251	179,996,077	8.71%	7,621
SNP	10,628,445	10,628,445	0.51%	165
Small indel	5,385,400	18,275,005	0.88%	338
Total	--	208,899,527	10.11%	--

Table S21 KEGG pathway enrichment analysis with species-specific in *Schizothorax o'connori* gill expression module. The enrichment was made by David 6.8. Related to Figure 4.

Term	Count	P-value	FDR
Cytokine-cytokine receptor interaction	18	1.37×10^{-4}	0.153829
Proteasome	10	6.91×10^{-4}	0.774936
Adherens junction	12	0.005101	5.595715
Tight junction	12	0.00968	10.37445
Endocytosis	23	0.018771	19.2141
Cell adhesion molecules (CAMs)	13	0.019247	19.65448
Jak-STAT signaling pathway	11	0.026025	25.68969
Mucin type O-Glycan biosynthesis	5	0.031393	30.17304
VEGF signaling pathway	9	0.032722	31.24428
Protein processing in endoplasmic reticulum	15	0.033383	31.77132
Apoptosis	8	0.051323	44.74722
ECM-receptor interaction	8	0.071888	56.83032
TGF-beta signaling pathway	9	0.077451	59.65602

Table S23 Gene ontology enrichment of genes for natural positively selected genes. Stars in the table represented that corresponding GO terms were significantly enriched among natural positively selected genes identified by various methods. Related to Figure 5.

GO term	ortholog	ohnolog	population
cell-cell adhesion	*	*	
mannosyl-oligosaccharide 1,2-alpha-mannosidase activity	*		
phospholipid catabolic process"	*	*	
phospholipase activity	*	*	
mismatched DNA binding	*	*	
intermediate filament cytoskeleton organization	*		*
intermediate filament	*		*
troponin complex	*		*
integrin complex	*		*
MHC class II protein complex	*		*
transferase activity transferring hexosyl groups	*		*
antigen processing and presentation	*		*
keratin filament	*		
alpha-mannosidase activity	*		
mannose metabolic process	*		
proton transport	*		
protein-L-isoaspartate (D-aspartate) O-methyltransferase activity	*		
actin cytoskeleton reorganization	*		
transferase activity transferring glycosyl groups	*		
microtubule organizing center	*		
phosphoric diester hydrolase activity	*		
glycerophosphodiester phosphodiesterase activity	*		
DNA repair	*		
cellular response to DNA damage stimulus	*		
immune response	*		

structural molecule activity	*		
nucleosome assembly		*	
chemokine activity		*	
protein dimerization activity		*	
multicellular organism development		*	
receptor binding		*	
mismatch repair		*	
DNA integration			*
voltage-gated potassium channel complex			*
nuclear envelope			*
guanylate cyclase activity			*
insulin-like growth factor binding			*
ion channel activity			*
ubiquitin-protein transferase activity			*
sensory perception of sound			*
lipoprotein metabolic process			*
neuron projection development			*
cGMP biosynthetic process			*

Table S24 KEGG biological pathway enrichment of genes for natural positively selected genes that identified from orthologs, ohnologs and population method. Stars in the table represented that corresponding KEGG pathways were significantly enriched among natural positively selected genes identified by various methods. Related to Figure 5.

KEGG pathway	ortholog	ohnolog	population
Cell adhesion molecules (CAMs)	*	*	*
Intestinal immune network for IgA production	*	*	*
Antigen processing and presentation	*	*	*
Linoleic acid metabolism	*	*	
Arachidonic acid metabolism	*	*	
Th1 and Th2 cell differentiation	*	*	
Apoptosis - multiple species	*	*	
Leukocyte transendothelial migration	*	*	
alpha-Linolenic acid metabolism	*	*	
NF-kappa B signaling pathway	*	*	
p53 signaling pathway	*	*	
Fanconi anemia pathway	*	*	
Hematopoietic cell lineage	*		*
Phagosome	*		*
Necroptosis	*		
MAPK signaling pathway - fly	*		
Collecting duct acid secretion	*		
Pentose and glucuronate interconversions		*	
Ether lipid metabolism		*	
Cytokine-cytokine receptor interaction		*	
Porphyrin and chlorophyll metabolism		*	
Osteoclast differentiation		*	
Homologous recombination		*	

Jak-STAT signaling pathway	*	
Protein export	*	
Steroid hormone biosynthesis	*	
Metabolism of xenobiotics by cytochrome P450	*	
Ferroptosis	*	
Serotonergic synapse		*
Fatty acid biosynthesis		*
Glycosphingolipid biosynthesis - globo and isoglobo series		*
Mucin type O-glycan biosynthesis		*
Glycosphingolipid biosynthesis - lacto and neolacto series		*
Renin secretion		*
mRNA surveillance pathway		*

Table S26 Summary of genome resequencing for the *Schizothorax o'connori* genome. Related to Figure 5.

Sample	Raw Base (bp)	Clean Base (bp)	Q20 (%)	Q30 (%)	GC Content (%)
Soco4700	100,884,532,500	100,555,179,600	97.00	95.46	38.80
Soco4000	99,544,687,800	99,066,715,500	96.77	94.86	38.38
Soco3100	113,226,539,400	112,599,624,300	96.88	95.26	38.59
Soco3600	109,571,165,700	109,310,314,800	97.02	95.56	38.39
Soco4200	97,375,310,700	97,218,252,300	96.93	95.33	38.42

Table S27 Mapping statistics summary of DNA resequencing clean data to the *Schizothorax o'connori* assembly. Related to Figure 5.

Sample	Total reads	Mapped reads	Mapping rate (%)	Average depth (X)	Coverage at least 1X (%)	Coverage at least 4X (%)
Soco4700	670,367,864	659,679,714	98.41	41.50	96.02	92.00
Soco4000	660,444,770	650,241,721	98.46	41.67	96.05	92.14
Soco3100	750,664,162	737,648,959	98.27	46.44	96.35	92.80
Soco3600	728,735,432	716,949,522	98.38	44.79	96.21	92.59
Soco4200	648,121,682	635,996,294	98.13	39.84	96.11	92.08

Table S28 Summaries of SNP annotated by ANNOVAR program for the *Schizothorax o'connori*. Related to Figure 5.

Sample	Exonic	Upstream	Downstream	Intergenic	Het rate(‰)	Total
Soco4700	181,403	201,721	192,191	5,338,764	3.88	8,608,643
Soco4200	180,202	201,451	191,563	5,315,069	3.88	8,567,019
Soco4000	180,618	202,510	193,237	5,365,074	3.91	8,649,913
Soco3600	180,934	205,314	194,417	5,401,142	3.95	8,702,258
Soco3100	183,343	206,028	195,309	5,452,834	3.97	8,775,788

Table S30 Statistics of variants calling. The short sequence reads generated by Illumina platform were mapped to the *Schizothorax o'connori* genome with BWA and performed variant calling with SAMtools to get the SNP result. Related to Figure 5.

	Number	Percentage (%)
All SNP	3,980,508	0.24
Heterozygosis SNP	3,911,059	0.24
Homology SNP	69,449	0.01

Table S31 *Schizothorax o'connori* assembly completeness evaluated using CEGMA. Related to Figure 1.

Species	Complete		Complete + Partial	
	# Prots	%completeness	# Prots	%completeness
<i>S. o'connori</i>	233	93.95	242	97.58

Table S32 Shared transposon identification in the *Schizothorax o'connori* and *Carassius auratus* genome. Related to Figure 2.

Species	Ratio of shared transposon length to whole genome length	Ratio of shared transposon length to total transposon length
<i>S. o'connori</i>	0.22	0.53
<i>C. auratus</i>	0.08	0.25

Table S33 Summary of gene lost in homologous regions of *Schizothorax o'connori*. Related to Figure 2.

	<i>C. auratus</i>	<i>S. grahami</i>	<i>S. o'connori</i>	<i>S. salar</i>
Subgenome divergence time (Ma)	26.62	17.12	1.23	80.00
Synteny pair	21,593	14,750	14,882	20,476
Genes of synteny	30,065	24,885	22,940	27,176
Synteny Genes On same sequence	1,872	0	542	565
Multiple pair lost	531	253	267	267
Singleton	2,246	2,695	1,546	999
Singleton (%)	11.70	18.59	10.99	5.09
Singleton/Ma	84	157	1,257	12
Exonerate+tblastn	1,074	1,300	760	470
Gene lost	1,172	1,395	786	529
Lost/Ma	44	81	639	7

Supplementary Figures

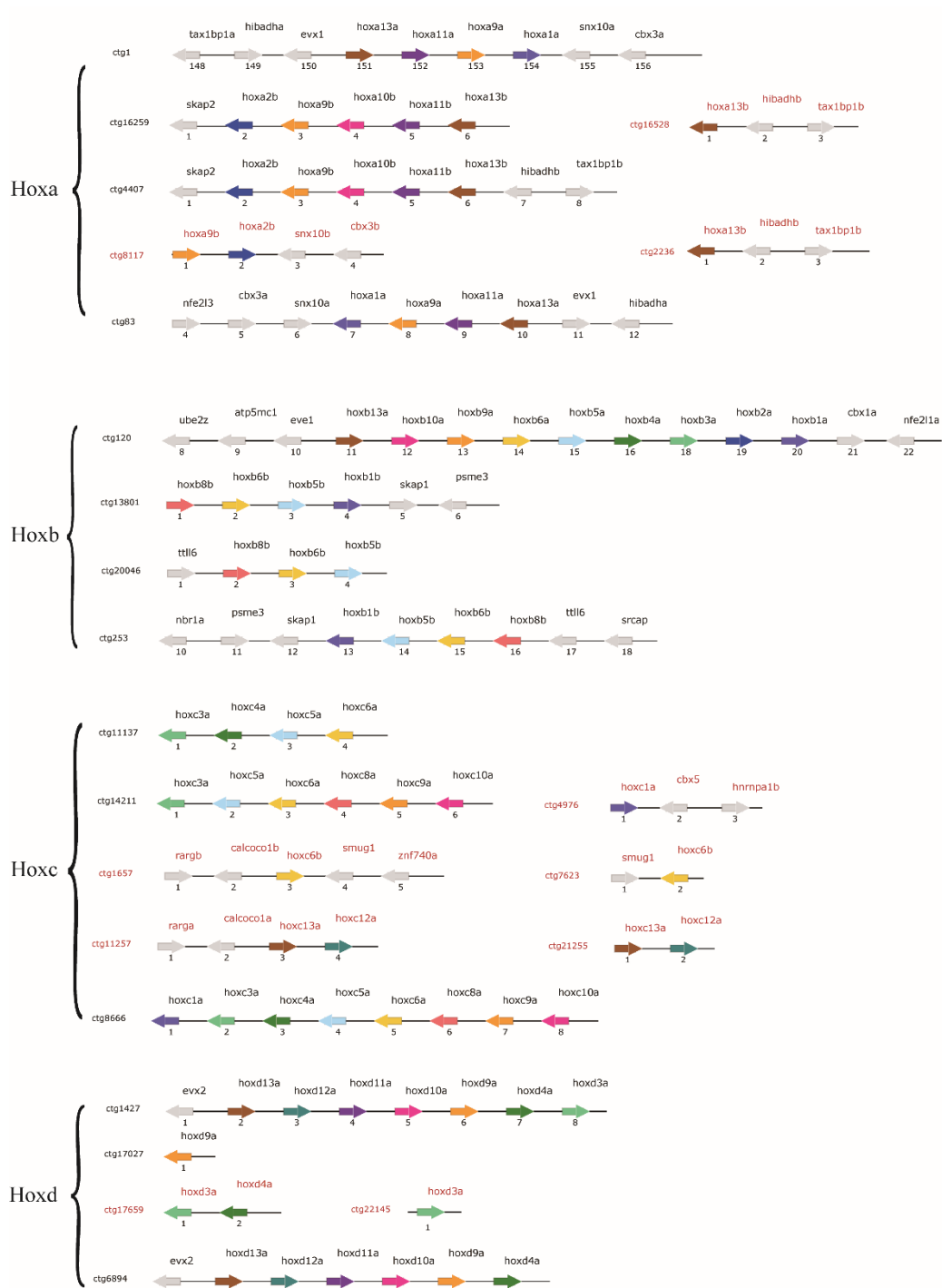


Figure S1 Hox gene clusters in *Schizothorax o'connori* genome. Related to Figure 1.

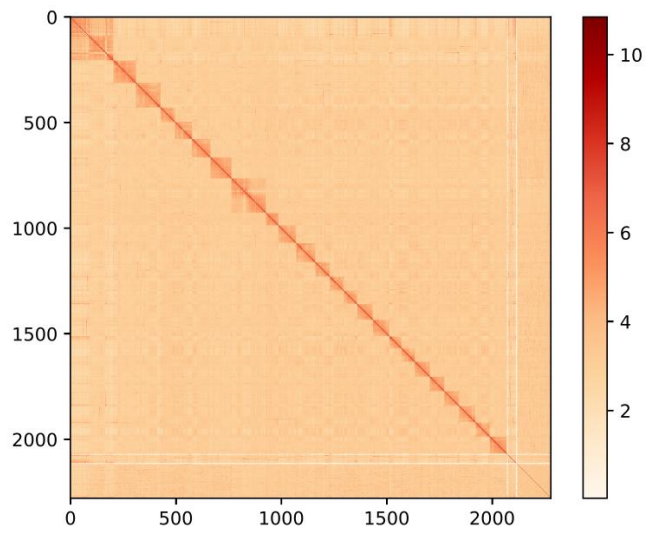


Figure S2 The whole genome Hic interaction map generated using a sliding window of 500 kb. Related to Figure 1.

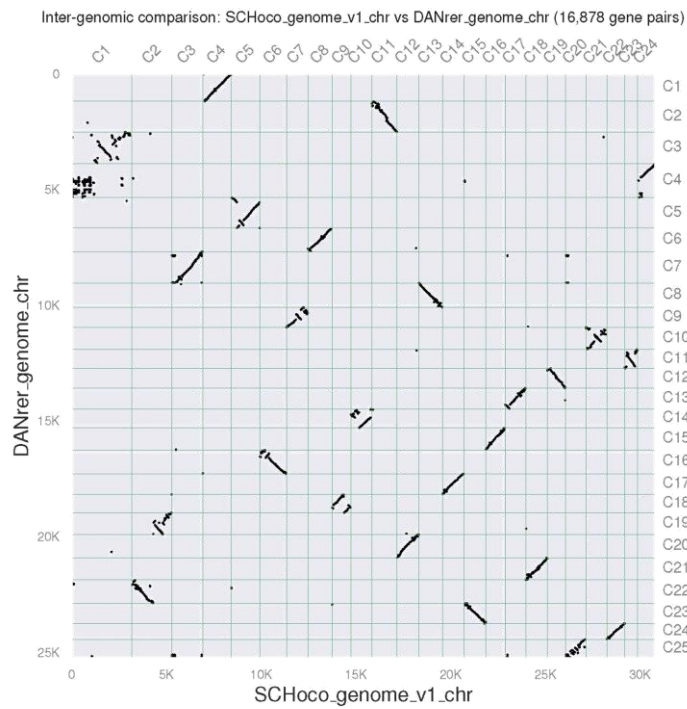


Figure S3. The whole-genome conservation synteny of *Schizothorax o'connori* and *Danio rerio*. Related to Figure 1.

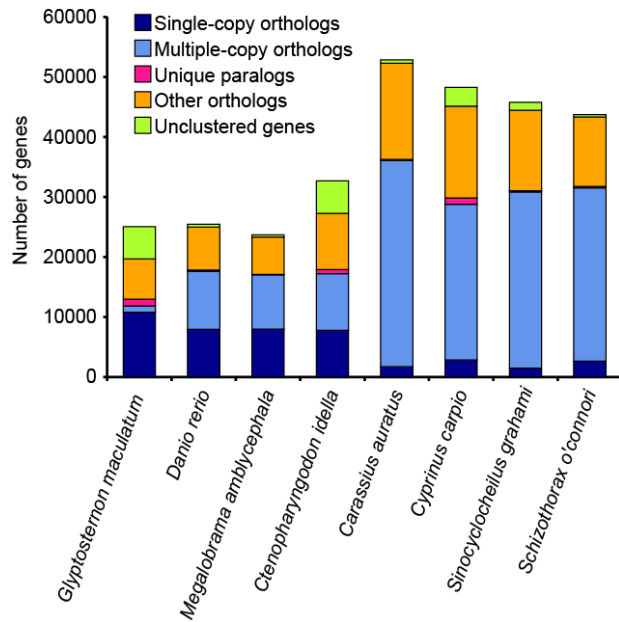


Figure S4 The orthologous relationships of *Schizothorax o'connori* and related species.

Single-copy orthologs means that gene families have copy numbers of 1 in 8 species; Multiple-copy orthologs means gene families with copy numbers larger than 1 in 8 species; Unique paralogs means gene families with no homologous genes in other 7 species, but existed in one species; Other orthologs means gene families presented in the several species, but not existed in all 8 species. Unclustered genes means gene families that genes were not clustered together. Related to Figure 2.

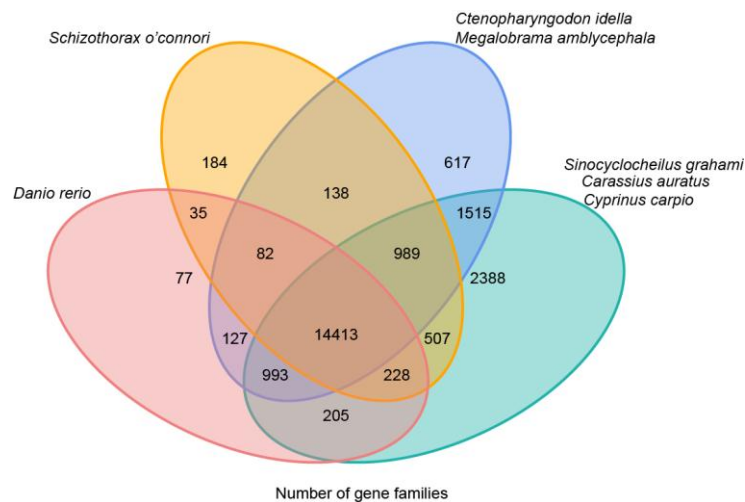


Figure S5 Venn diagram of the shared gene families in *Schizothorax o'connori* and related species. Related to Figure 2.

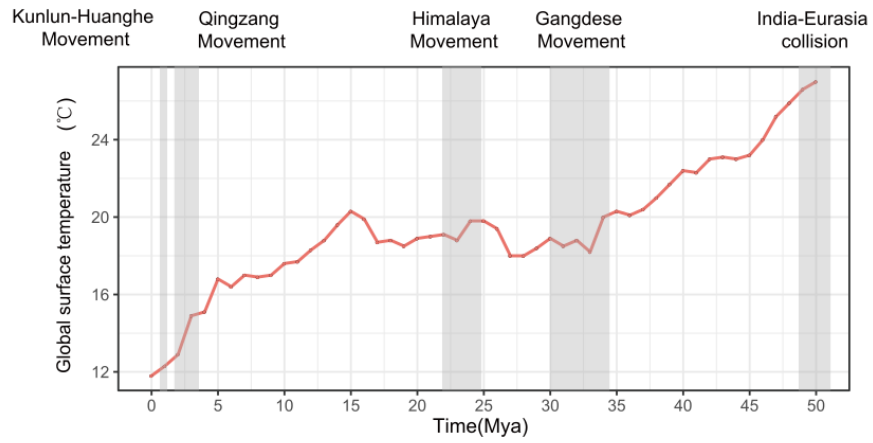


Figure S6 the dramatic tectonic movements in the Qinghai-Tibet Plateau and global surface temperature fluctuations after the India-Eurasia collision, shown by the gray bars. Related to Figure 2.

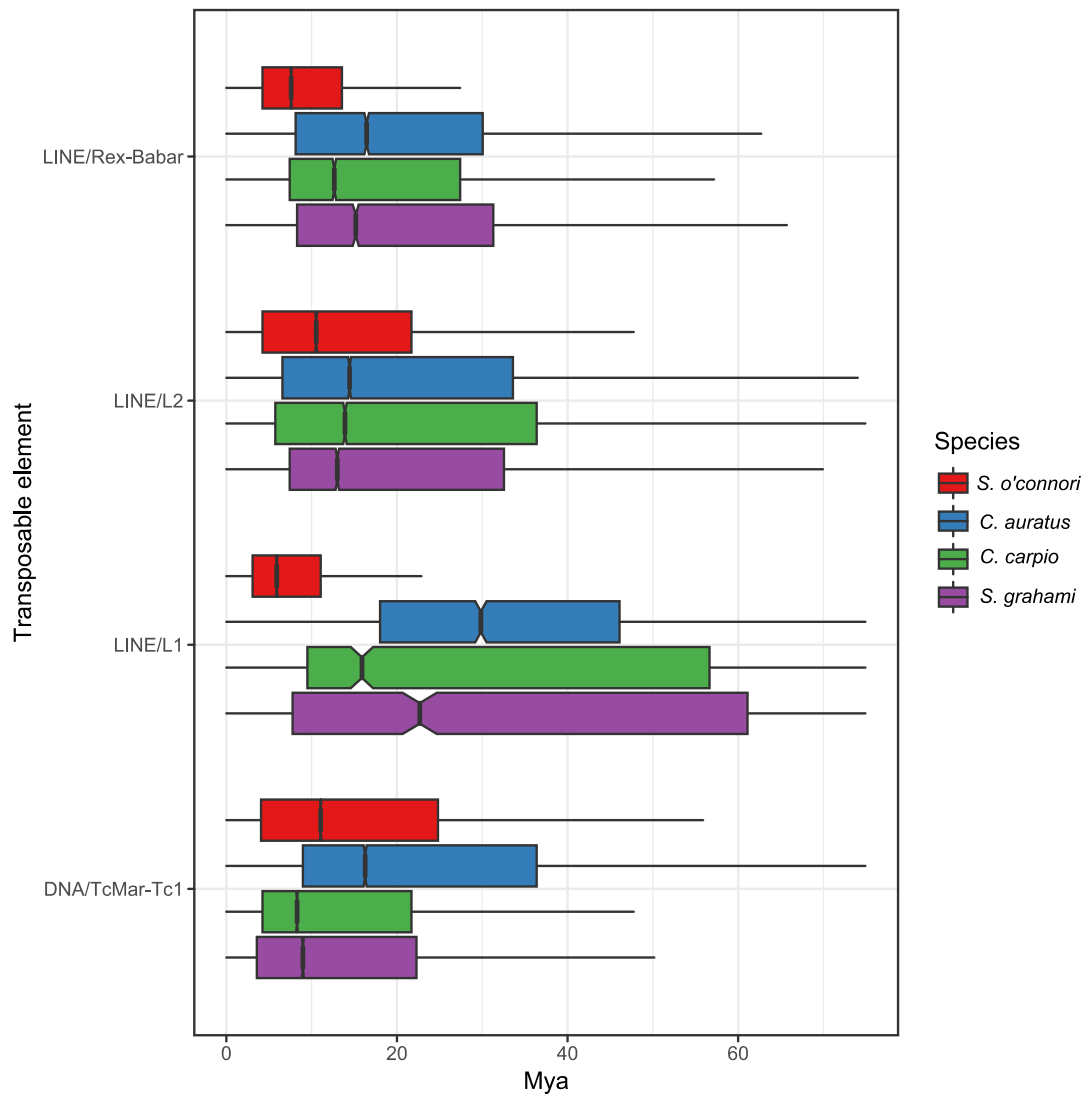


Figure S7 Boxplot display the Transposable elements divergence time calculated by using the synonymous substitution site per million years estimated for cyprinids (0.003121781). Related to Figure 2.

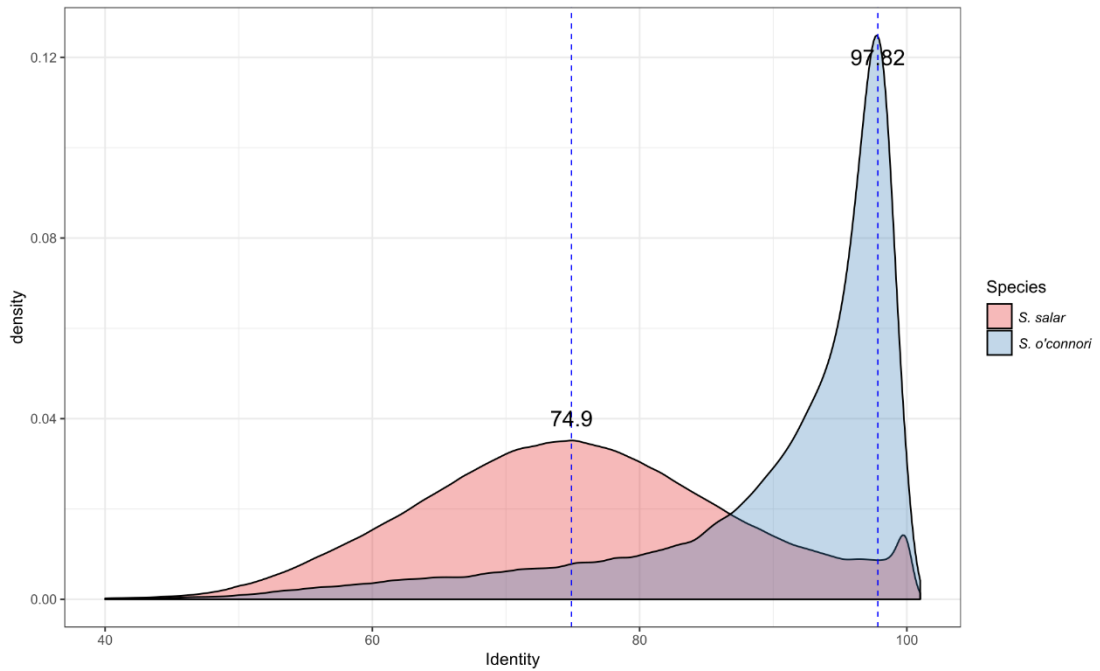


Figure S8 High score pair (HSP) similarity distribution for *Schizothorax o'connori* and Salmon homologous regions. Related to Figure 2.

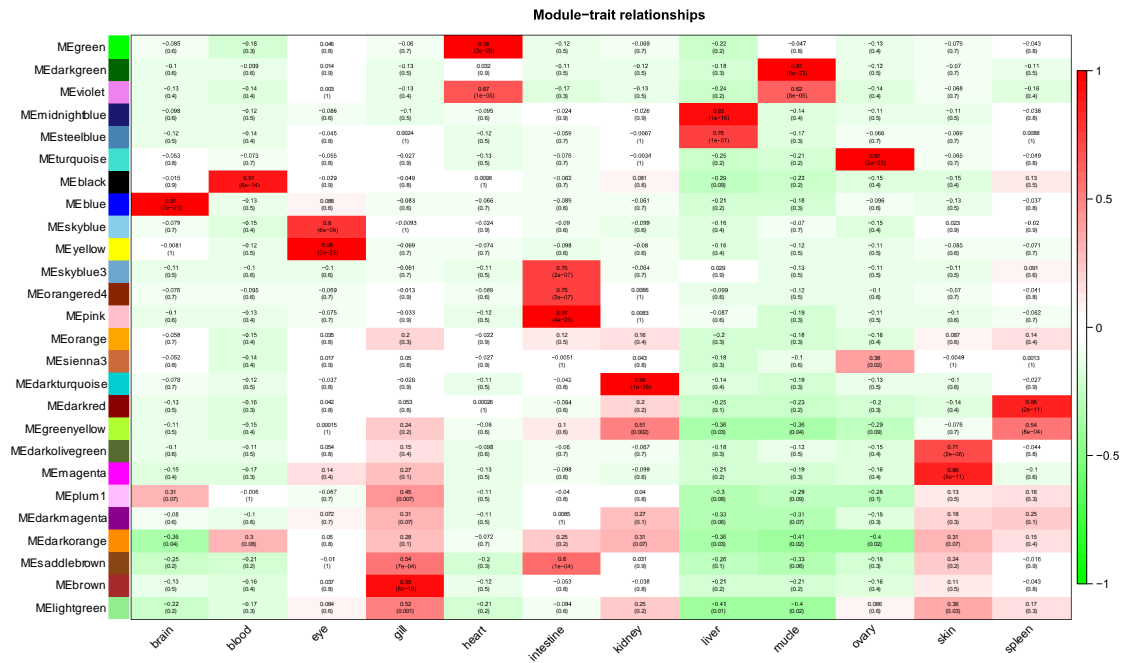


Figure S9 Model and tissue correlation and p-value for *Schizothorax o'connori* from WGCNA analysis. In each cell, the top value means Pearson correlation value and bottom value means p-value. Related to Figure 4.

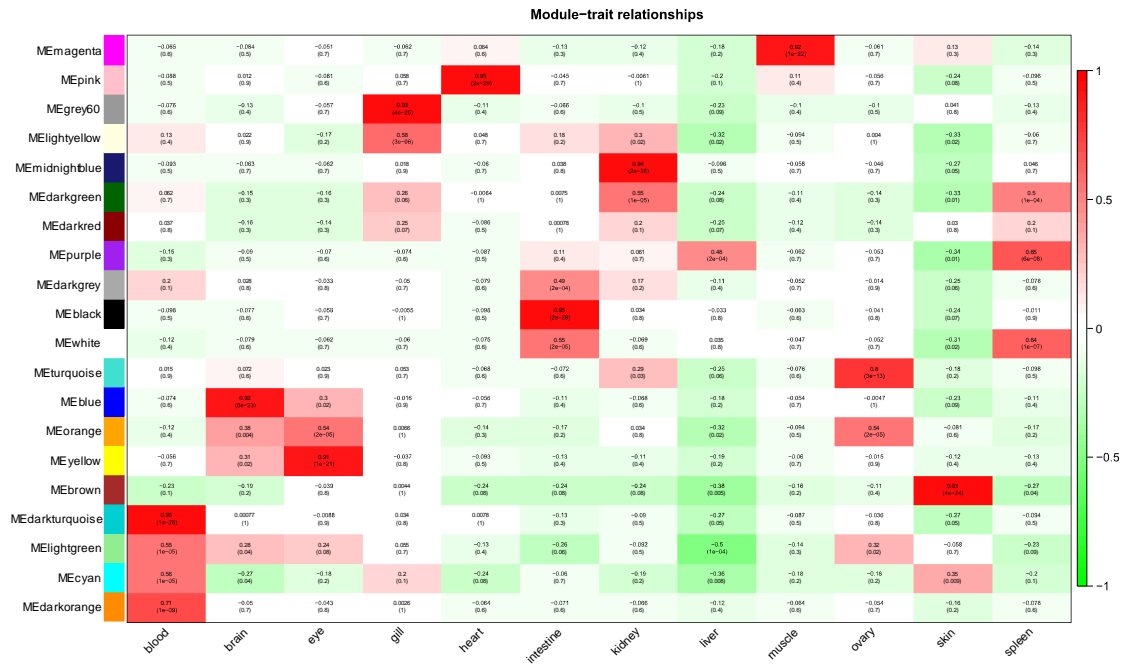


Figure S10 Model and tissue correlation and p-value for *Danio rerio* from WGCNA analysis. In each cell, the top value means Pearson correlation value and bottom value means p-value. Related to Figure 4.

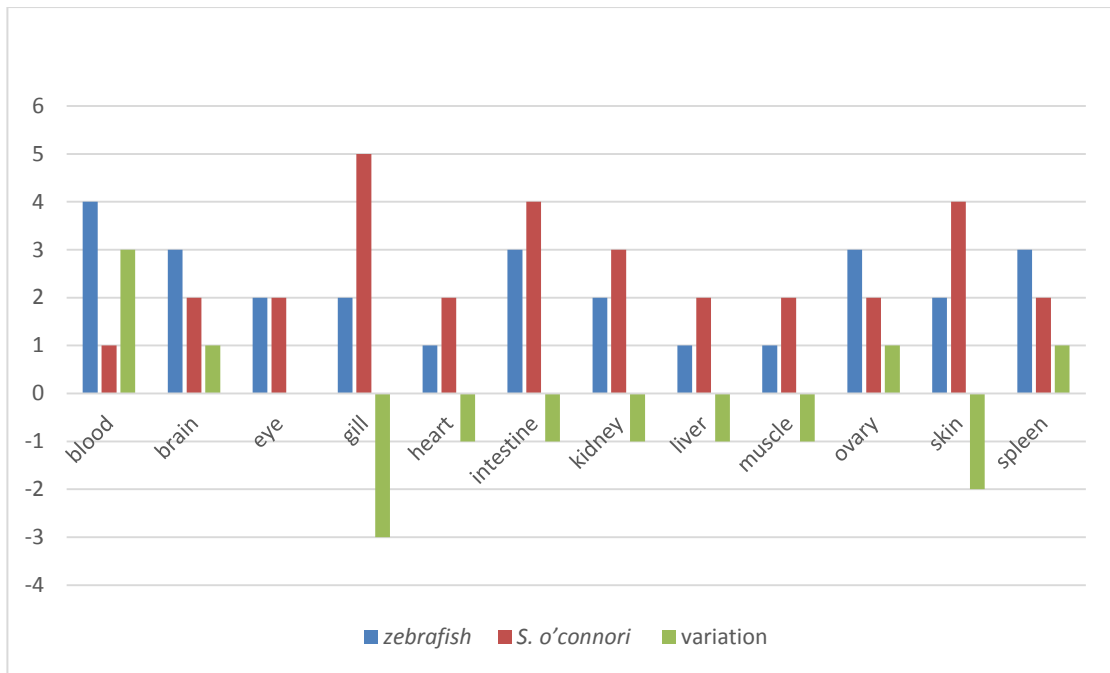


Figure S11 Model number for each tissue with correlation great than 0.3. Variation means the model number of *Danio rerio* minus that of *Schizothorax o'connori*. Related to Figure 4.

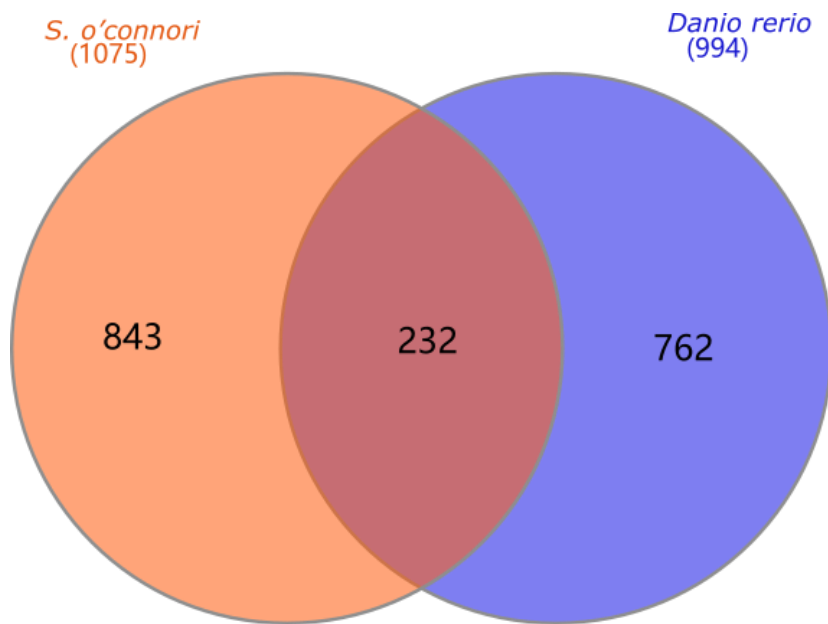


Figure S12 The Venn plot display the share genes in the top correlation modules of gill. We chose 1,699 *Schizothorax o'connori* homology genes (single best hit) to *D. rerio* to make the plot. Related to Figure 4.

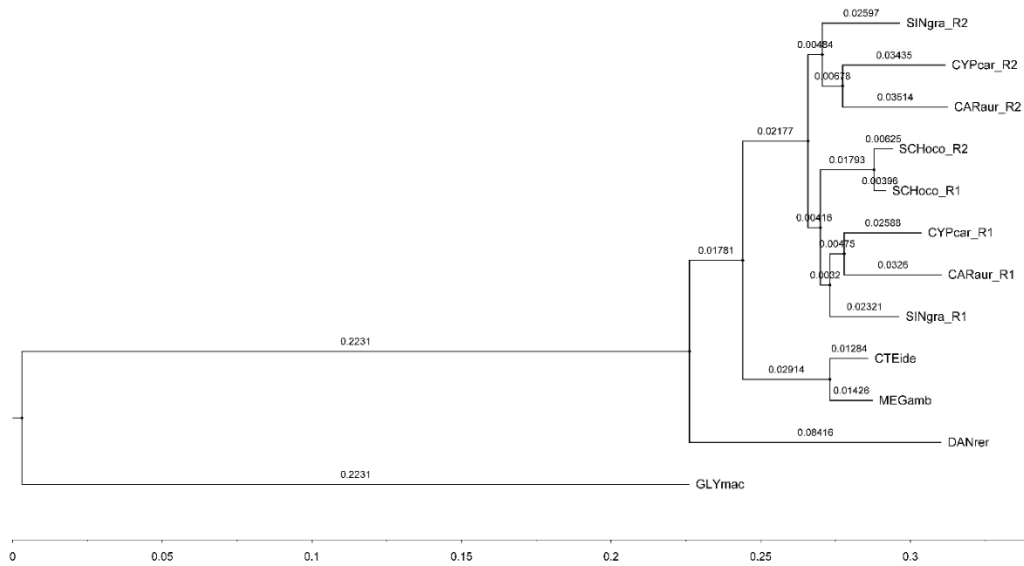


Figure S13 Phylogeny of the (sub)genomes of 8 species based on 297 genes totaling 566,916 nucleic acid sites. It is reconstructed using RAXML (-v 8.2.12) with GTRGAMMA model and 1,000 fast bootstrap replicates were also performed. The name of species were listed in Table S1. Related to Figure 2.

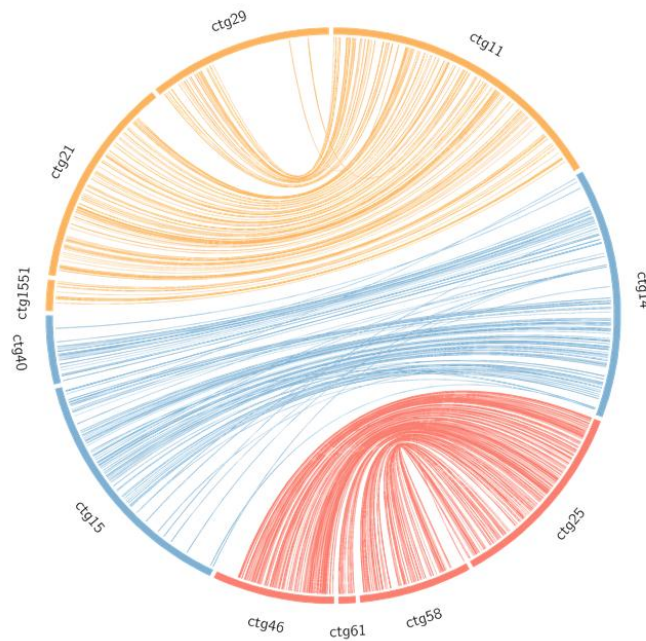


Figure S14. The synteny of shared transposon between duplicated genome regions for *Schizothorax o'connori*. Related to Figure 2.

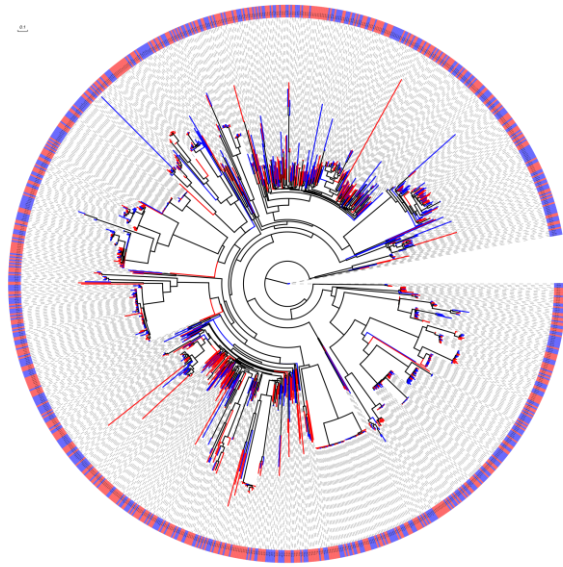


Figure S15. The phylogeny of 1,417 DNA/TcMar-Tc1 of ctg1 and reciprocal sequence in the *Schizothorax o'connori* genome. The red color represented the transposon from R1 and blue for R2. TEs with length <200bp were removed and the topology were inferred with Raxml (f a -m GTRGAMMA -# 100). Related to Figure 2.

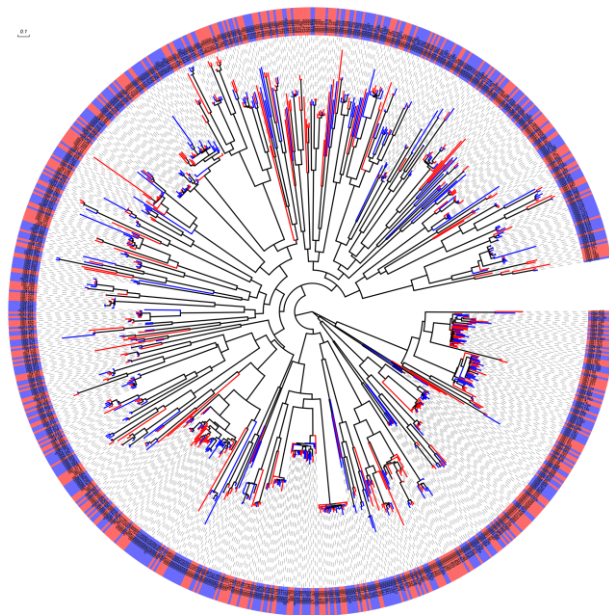


Figure S16. The phylogeny of 992 DNA/hAT-Ac of ctg1 and reciprocal sequence in the *Schizothorax o'connori* genome. The red color represented the transposon from R1 and blue for R2. TEs with length <200 bp were removed and the topology were inferred with Raxml (f a -m GTRGAMMA -# 100). Related to Figure 2.

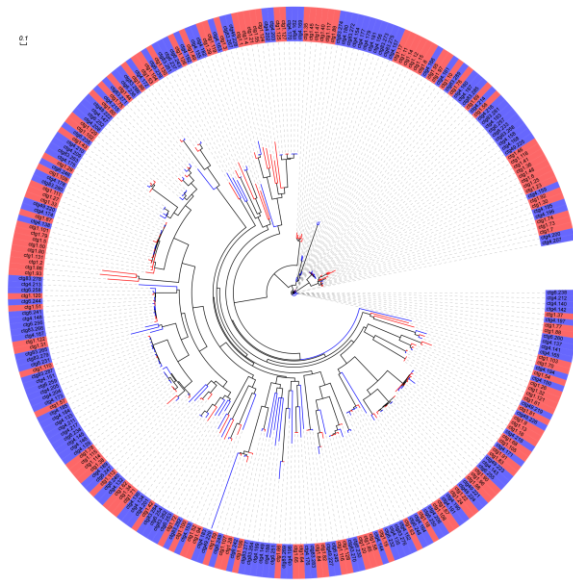


Figure S17. The phylogeny of 285 LINE/L2 of ctg1 and reciprocal sequence in the *Schizothorax o'connori* genome. The red color represented the transposon from R1 and blue for R2. TEs with length < 200 bp were removed and the topology were inferred with Raxml (f a -m GTRGAMMA -# 100). Related to Figure 2.

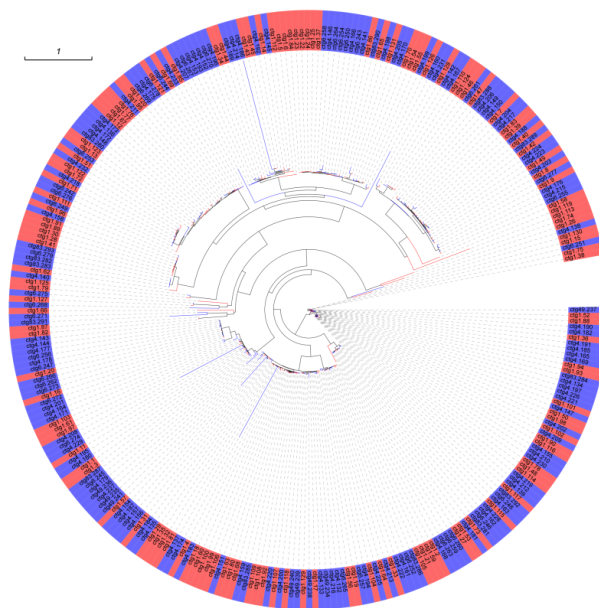


Figure S18. The phylogeny of 292 LINE/Rex-Babar of ctg1 and reciprocal sequence in the *Schizothorax o'connori* genome. The red color represented the transposon from R1 and

blue for R2. TEs with length < 200 bp were removed and the topology were inferred with Raxml (f a -m GTRGAMMA -# 100). Related to Figure 2.

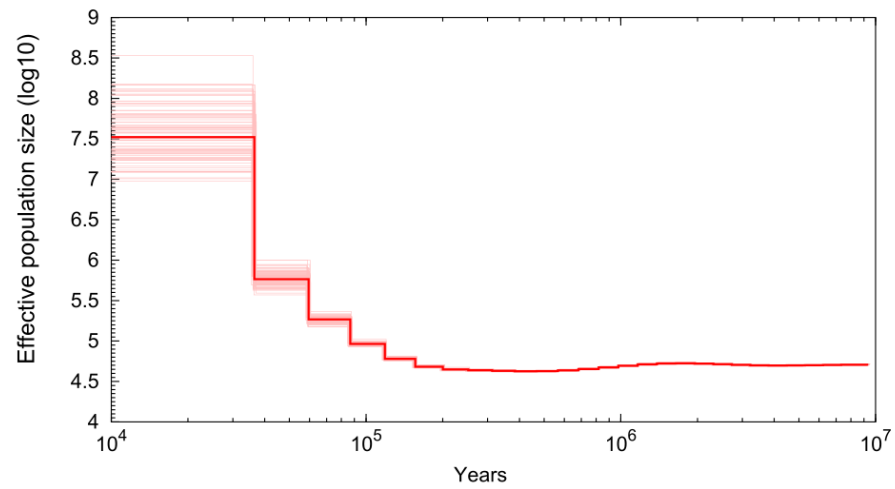


Figure S19 The deduced historical effective population for *Schizothorax o'connori*. Related to Figure 2.

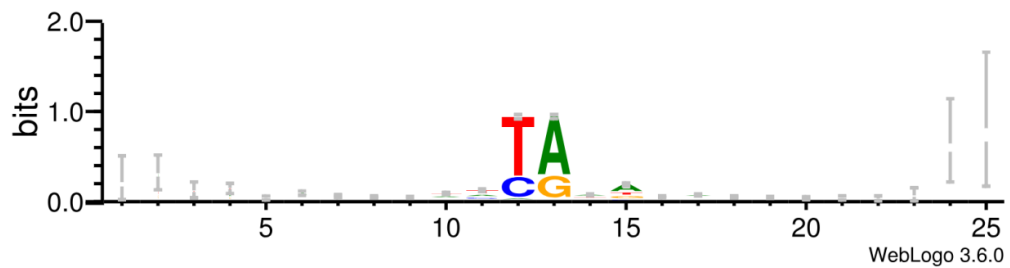


Figure S20 The webLogo show that "TA enrichment" around insertion sites between duplicated genomic regions. TA was the insert site of DNA transposons. Related to Figure 2.

Duplication depth	Reference chromosome	Collinear blocks
1	SCOC_01949	SCOC_06137
1	SCOC_01951	
1	SCOC_01953	SCOC_06140
1	SCOC_01954	SCOC_06142
1	SCOC_01955	SCOC_06143
1	SCOC_01956	SCOC_06144
1	SCOC_01957	
1	SCOC_01958	
1	SCOC_01959	SCOC_06145
1	SCOC_01960	SCOC_06148
1	SCOC_01961	SCOC_06149
1	SCOC_01963	SCOC_06150
1	SCOC_01964	SCOC_06151
1	SCOC_01966	SCOC_06153
3	SCOC_01967	SCOC_06154 SCOC_37643 SCOC_11835
3	SCOC_01968	SCOC_06155
3	SCOC_01969	
3	SCOC_01970	SCOC_06156 SCOC_37642 SCOC_11834
3	SCOC_01971	SCOC_06158 SCOC_37641 SCOC_11833
3	SCOC_01972	SCOC_06159
3	SCOC_01973	SCOC_06160 SCOC_40209 SCOC_11832
4	SCOC_01975	SCOC_06161 SCOC_40208 SCOC_11831 SCOC_42056
4	SCOC_01976	SCOC_06162 SCOC_40207 SCOC_11830 SCOC_42057
3	SCOC_01977	SCOC_06163 SCOC_42059
3	SCOC_01978	SCOC_47928
3	SCOC_01979	SCOC_47929
3	SCOC_01980	SCOC_20700 SCOC_40204 SCOC_42060
1	SCOC_01981	SCOC_20703
1	SCOC_01982	SCOC_20704

Figure S21 Gene lost events displayed on synteny blocks. The blue rectangle means gene lost, for the red rectangle block, it may be caused by other whole genome duplication, so we don't consider about it. Related to Figure 2.