BMC Genomics

RESEARCH

Open Access

Whole-genome relaxed selection and molecular constraints in *Triplophysa* under adapted Qinghai-Tibetan Plateau

Qingyuan Zhao[1*†], Fei Xie[1†], Qiuyue He[1], Lulu Wang[1], Kenan Guo[1], Cong Zhang[1] and Yong Wang[1*]

Abstract

High-altitude environments are inhospitable, but *Triplophysa*, the largest taxon among the three major fish groups in the Qinghai-Tibetan Plateau (QTP), is an exception. However, the evolutionary profiling of the common ancestor and its contribution to the adaptation of existing QTP native species is unclear.We researched the comparative genomics of *Triplophysa* species and found that the genome-wide genes of *Triplophysa* and its ancestry have the characteristics of rapid evolution.Moreover, the rapid evolution of the ancestral genes was caused by relaxed selection. Natural selection analysis showed that more ancestral relaxed selection genes were under strongly purifying selection and showed higher expression in QTP endemic *Triplophysa* species.The change in natural selection might be associated with the adaptation to QTP. It should be noted that SPT5 homolog, DSIF elongation factor subunit (supt5h) experienced relaxed selection in common ancestral populations of *Triplophysa* but under purifying selection in extant species, which might be related to hypoxia adaptation of QTP. In summary, the extant species in different environments were used to infer the evolutionary profile of the common ancestor and to identify candidate genes based on changes in natural selection. Our work might provide new clues for understanding adaptation to extreme environments.

Keywords Qinghai-Tibetan Plateau, Rapid evolution, Relaxed selection, Adaptation, Comparative genomics

Introduction

High-altitude environments are inhospitable [1]. The extreme environment of low oxygen, low temperature and high ultraviolet radiation on the Qinghai-Tibet Plateau (QTP) has brought great challenges to the survival of organisms [1, 2]. Animals living on the QTP have evolved convergent phenotypes (such as hypoxic resistance and cold tolerance) to overcome the intense selection pressures of the plateau environment [1, 3–6]. These studies typically compare one or several extant low-altitude species/populations with high-altitude species/populations and find positive selective variation [7] to understand the genetic basis of adaptation to extreme plateau environments.However, recent research on snowfinches in the Qinghai-Tibet Plateau has suggested that the initial genetic foundations for adaptations to extreme environments may emerge from ancestral lineages [7–9]. Species in a shared environment tend to evolve similar phenotypes under the influence of their phylogenetic context. However, it is interesting that some other QTP species share molecular evolutionary characteristics with their close low-altitude species despite living in different

†Qingyuan Zhao and Fei Xie contributed equally to this work and should be considered co-first authors.

*Correspondence:
Qingyuan Zhao
zqingyuan@tmmu.edu.cn
Yong Wang
yongwanglas@tmmu.edu.cn
[1] Department of Laboratory Animal Science, College of Basic Medical Sciences, Army Medical University (Third Military Medical University), Chongqing, China

environments. For instance, the rate of gene evolution in species native to the Tibetan Plateau (such as *Glyptosternon maculatum*, *Triplophysa siluroides* and *Phrynocephalus erythrurus*) and their closely related counterparts at lower altitudes outpaces that of other distantly related species at lower altitudes [10–12]. The occurrence of similar evolutionary profiling in different taxa is hardly a coincidence. Thus, it is intriguing whether the common ancestor of the Tibetan Plateau native species and its close relatives at low altitude shares evolutionary characteristics with existing species and provides the genetic basis for the adaptive evolution of existing species.

Triplophysa belongs to the family Nemacheilidae of Cypriniformes. *Triplophysa* is the largest genus of freshwater fishes in Asia, and more than 200 species have been discovered so far [13, 14]. *Triplophysa* species are distributed all over the world but mainly inhabit the Qinghai-Tibet Plateau and its surrounding areas [15, 16], including the underground karst caves in southwest China, the highly alkaline environment, the high-altitude areas of the Qinghai-Tibet Plateau and other extreme environmental regions [17, 18]. The species of *Triplophysa* can adapt to various extreme environments, and species in different environments have the same characteristics of rapid genome evolution, making it an excellent research material for studying the relationship between high-altitude extreme environment adaptation and ancestral genome evolution. At present, the whole genome data of six species of *Triplophysa* have been published, including *T. tibetana* (Habitat elevation:~4,000–5,000 m) [18], *T. siluroides* (Habitat elevation: ~3,000–4,000 m) [16], *T. bleekeri* (Habitat elevation: ~200-3,000 m) [19], *T. dalaica* (genome sampling Site: 43°220′4300″N, 116°390′2400″E; Habitat elevation: 1226 m) [20], *T. rosa* (Habitat elevation: 314 m) [21], *T. bombifrons* (Habitat elevation: ~1300–1450 m) [22]. Therefore, only *T. tibetana* and *T. siluroides* are endemic animals in high-altitude extreme habitats of the Qinghai-Tibet Plateau. *T. yarkandensis*, also known as *Hedinichthys yarkandensis*, does not cluster with *Triplophysa* into one taxon based on phylogenetic analysis of mitochondrial genomes [23]. Although several genomes of *Triplophysa* have been published, including species endemic to the Qinghai-Tibet Plateau, systematic studies on the evolution of adaptability to the plateau environment have not been carried out.

This study aimed to explore the mechanism for the extensive adaptability of *Triplophysa* species, especially in high-altitude environments, and the role of the ancestral genome in it. We conducted comparative genomic studies using all published genomes of *Triplophysa* and its closely related species. The results showed that the common ancestral genome of *Triplophysa* was under relaxed selection, leading to rapid evolution. Rapid evolution may have provided the genetic basis for the adaptation of the species to QTP. We found that genes that evolved rapidly in ancestral genomes and were subject to strongly purifying selection in descendant species are likely to be involved in high altitude hypoxia adaptation, which provides new insights into the study of QTP environmental adaptation.

Results

The analysis of phylogenesis and demographic history

To better perform the research of adaptive evolution, a phylogenetic analysis was conducted on 15 species using 1781 high-confidence one-to-one orthologous genes from a genome-wide set. All of the 15 species belong to Cypriniformes, except for the electric eel, which serves as the outgroup. The analysis showed that zebrafish and loach diverged 80 million years ago (Fig. 1a), consistent with previous research [21, 24]. The taxonomic relationships of 13 species of loaches were also consistent with previous studies [25]. In addition, with the exception of *T. yarkandensis*, the other *Triplophysa* species were grouped into one taxon, indicating that *T. yarkandensis* is not genetically a *Triplophysa* species and is consistent with previous studies [23]. Among *the six Triplophysa* species, *T. rosa* was identified as the basal species within *Triplophysa*, consistent with the previous study that *T. rosa* is the basal species [26]. The divergence time of the Qinghai-Tibet native *Triplophysa* species was later than that of the *T. rosa* species but earlier than the other three *Triplophysa* species. The origins of *T. rosa*, *T.tibetica* and *T.siluroides* were 17.27 (95% CI: 16.425–18.1523), 12.67 (95% CI: 12.0416–13.309) and 16.34 (95% CI: 15.5356–17.1701) million years ago, respectively. The origin, evolutionary relationship, and differentiation time of *Triplophysa* species are consistent with previous studies of the mitochondrial genome but earlier than other genomics studies [21, 26, 27].

To understand the historical population dynamics of *Triplophysa* species and other loaches, we utilized psmc software to estimate the effective population size. The results indicate that *Oreonectes daqikongensis* had the largest effective population, while *Beaufortia kweichowensis* had the smallest effective population two to three million years ago (Fig. S1). Since that time, the effective population sizes of all species, except for *T. tibetica*, have experienced significant expansions and contractions.

Rapid evolution in *Triplophysa* and its ancestor

Previous research showed that QTP endemic species (such as species from *Triplophysa*) and their close relatives have relatively rapid evolutionary rates. Thus, we suspect that all *Triplophysa* species have a relatively high

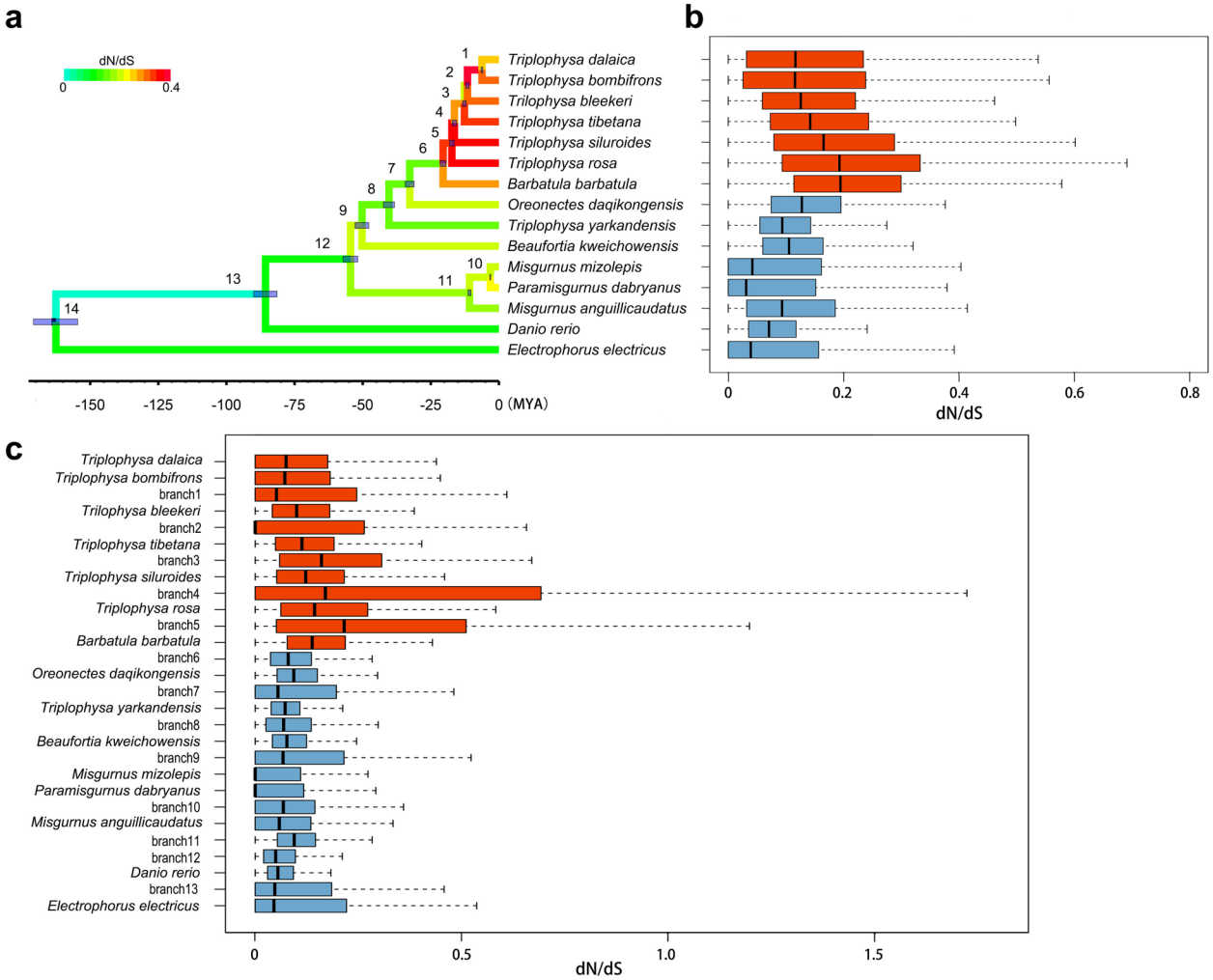

Fig. 1 Phylogenetic and evolutionary analysis of *Triplophysa*. **a** Divergence time (95% confidence interval) and evolutionary rate estimates in *Triplophysa* and other fishes. The dN/dS values were calculated by tandem CDS from 1781 single-copy orthologous genes. **b** Evolutionary rate statistics of protein-coding genes for each species based on estimates of 14,982 orthologous genes. **c** Evolutionary rate distributions of extant species and their ancestral lineage based on estimates of 1781 single-copy orthologous genes. The horizontal line in the centre of the box chart represents the median value

evolutionary rate, so we first tested the ratio of nonsynonymous (dN) to synonymous (dS) substitutions (dN/dS) for all genes in each species. The results showed that, as we suspected, all *Triplophysa* species had relatively high dN/dS values (Fig. 1b).

We suspect that the similar rapid evolution of all species is associated with the rapid evolution of the ancestral sequence. Therefore, the CDS of 1781 high-confidence single-copy orthologous genes were concatenated, dN/dS values were evaluated, and dN/dS values of each species and ancestral branch were obtained.The results showed that all species of *Triplophysa* had faster evolutionary rates than other species, and it was more noteworthy that each ancestral branch of *Triplophysa* also had higher dN/dS values, indicating that the ancestral branches were similar to existing species and had faster evolutionary rates (Fig. 1a). In order to exclude the interference of a small number of gene evolution rates on the overall evolution rate, we again evaluated the dN/dS value of each high-confidence single-copy orthologous gene. We obtained dN/dS values for all species and ancestral branches. Consistent with the above results, most of the species of *Triplophysa* and the ancestral branches (except for *T. dalaica*, *T. bombifrons*, branch1 and branch2) had higher dN/dS values (Fig. 1c). In particular, the dN/dS value from the common ancestor of *Triplophysa* is the largest, which may be related to the robust adaptability of *Triplophysa* fish (Fig. 1c). Intriguingly, it appears that

the earlier the species and the ancestral branch occurred within the genus *Triplophysa*, the higher the dN/dS value of the ancestral branch and species.

Genome-wide relaxed selection in *Triplophysa* ancestor

The rapid evolution of the common ancestor of *Triplophysa* fish is probably associated with the relaxed selection or positive selection. Therefore, a test of relaxed selection for genome-wide protein-coding genes was conducted. We used the RELAX software [28] to calculate an intensity parameter k for each gene in every test branch (k > 1 indicates intensified selection, and k < 1 indicates relaxed selection). Firstly, we compared all branches (including all internal branches) of *Triplophysa* and all branches of non-*Triplophysa* loach. The results showed that most of the genes from *Triplophysa* were under relaxed selection (Fig. S2).And the number of relaxed selection genes in *Triplophysa* was significantly higher than that in non-*Triplophysa* (Fig. S2).We then calculate the free k values of all the internal branches, and the result shows that branch 5 has the lowest k value. These results suggest that the rapid evolution of the common ancestor of *Triplophysa* is associated with relaxed selection.

So we compared the common ancestor of *Triplophysa* (branch5) with the ancestor of *Triplophysa* and *Barbatula barbatula* (branch6) to determine whether the common ancestor of *Triplophysa* underwent relaxed selection compared to the earlier ancestor (Fig. 1a). Compared to branch6, branch5 had more relaxed selection genes (k < 1) than intensified selection genes (k > 1) (Fig. 2a). Although there are still many intensified selection genes in both branch5 and branch6, branch5 genomes tend to be under relaxed selection compared to branch6. Hence, the relaxation of natural selection dominates genome evolution in the common ancestor of *Triplophysa*. The

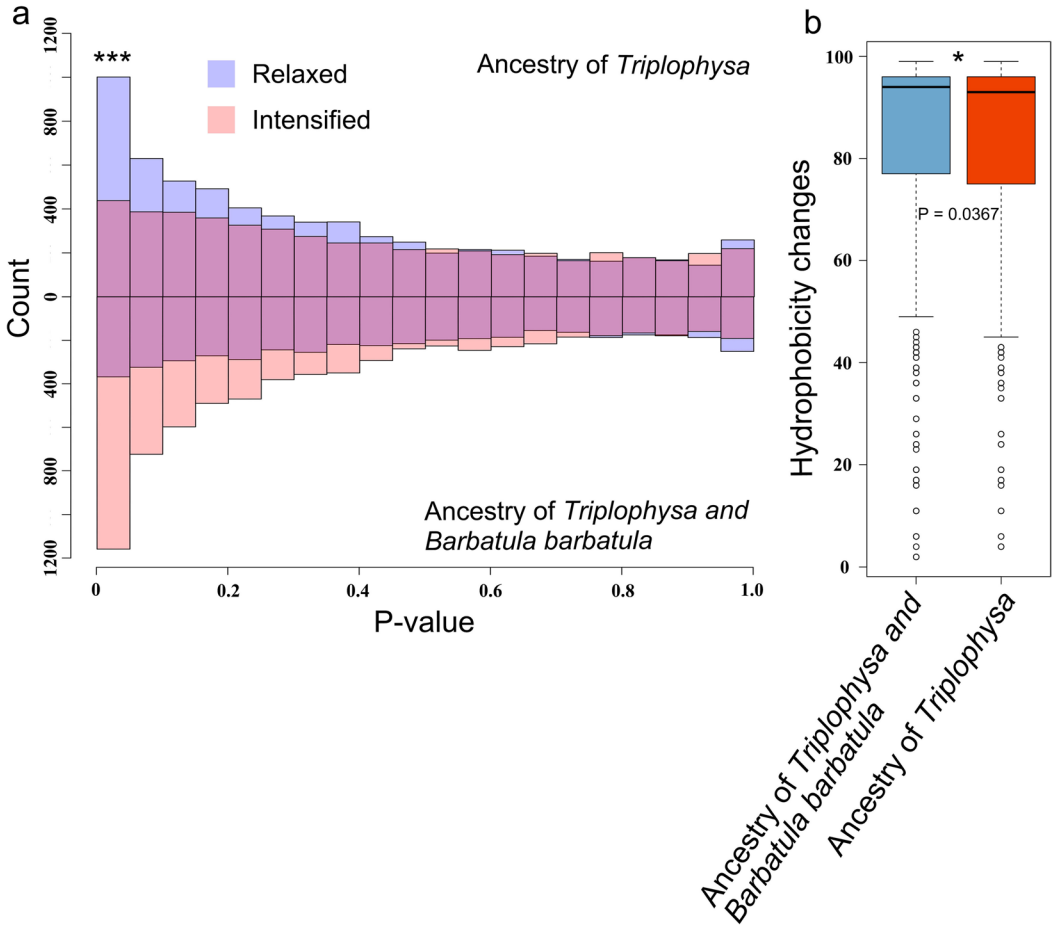

Fig. 2 Genome-wide relaxed selection of common ancestor of *Triplophysa*. **a** *P*-value distributions of the RELAX tests of the common ancestor of *Triplophysa* (top half) and the common ancestor of *Triplophysa* and *Barbatula barbatula* (bottom half). Intensified: k > 1; Relaxed: k < 1. *** $p < 0.001$, Fisher's exact test; The ancestry of *Triplophysa* contained more relaxed selection genes. **b** Hydrophilicity scores (HSs) of the common ancestor of *Triplophysa* and the common ancestor of *Triplophysa* and *B. barbatula* branches. Significant differences: * $p < 0.05$, generalised linear mixed model; The analysis was based on 1781 high-confident single-copy orthologous genes

consequence of relaxed selection is that the genome will accumulate more slightly deleterious mutations [29–31].

To further confirm that genome-wide relaxed selection on the ancestry of *Triplophysa*. We analysed the potential "harmfulness" of protein amino acid substitution in the common ancestor of *Triplophysa* and the ancestor of *Triplophysa* and *Barbatula barbatula*. We use the hydrophobicity score (HS) as the "harmfulness" of nonsynonymous substitutes. The basic assumption is that the structure and function of protein is affected by changes in the hydrophilic (hydrophobic) properties of amino acid in the mutation site [32], greater hydrophilic (or hydrophobic) differences in amino acid changes may be more harmful [33, 34]. A lower HS score indicates a tremendous difference in hydrophilicity, while 100 indicates the same properties [35]. Consistent with the ancestor of *Triplophysa* accumulating more nonsynonymous mutations than the ancestor of *Triplophysa* and *Barbatula barbatula*, the HS score in the ancestor of *Triplophysa* was significantly lower than the ancestor of *Triplophysa* and *Barbatula barbatula* (universal linear mixing model, $p=0.0367$; Fig. 2b). It suggests that the ancestor of *Triplophysa* proteins has more slightly harmful mutations than those of *Triplophysa* and *Barbatula barbatula*. These results further support that the *Triplophysa* ancestor might under genome-wide relaxed selection.

More ancestral relaxed selection genes were under natural selection in *Triplophysa*, which is endemic to QTP

The increase in mutation load and the accumulation of slightly harmful mutations caused by relaxed selection may be the prerequisite for the adaptability of the ancestral population to various extreme environments [30, 36–48]. In this research, the mutated sites of proteins favorable to the QTP adaptation, which accumulated through the rapid evolution of the common ancestor population of *Triplophysa*, might have two fates in their offspring: (1) They only exist in the native species of QTP and are lost in the genomes of other species; (2) They are critical variations for species to adapt to QTP, but they exist in all genomes (or most genomes) of extant species. The analysis of this study identified positive selection as the former and strongly purifying selection as the latter. Therefore, we performed a positive selection analysis on *T. siluroides* and *T. tibetana*, resulting in 209 and 231 positive selection genes in *T. siluroides* and *T. tibetana*, respectively (branch site model, $p<0.05$). These extant species do not have more positive selection genes derived from ancestral relaxed selection genes (Fig. S4a). Next, we analyzed whether there are more relaxed selection genes of the common ancestor of *Triplophysa* experiencing strong purifying selection in *T. siluroides* and *T. tibetana*. Therefore, we obtained 1553 and 978 genes with dN/dS values less than 0.1, *P*-values less than 0.01, and no positive selection sites, which were determined by the codeml branch model and branch-site model from *T. Tibetana* and *T. siluroides*, assumed as strongly purifying selection genes. More of these strongly purifying selection genes were derived from significantly relaxed selection genes in the ancestors of *Triplophysa* (Fig. 3a). Such results were unexpected because both methods are based on dN/dS to determine whether genes are under relaxed selection (intensified selection) or purifying selection (positive selection), so the strongly purifying selection is likely to have more overlap with the intensified selection.

Next, we analysed the relationship between relaxed selection genes (intensified selection) and positive selection genes (or strongly purifying selection genes) in *T. tibetana* and *T. siluroides*. We found that only a small number of positive selection genes are also relaxed selection genes in extant species (Fig.S4b). So, relatively more ancestral genes under relaxed selection were positively selected in extant species (compare to Fig.S4a). In contrast to the number and proportion of strongly purifying selection genes in significantly intensified selection genes, the number and proportion of strongly purifying selection genes in significantly relaxed selection genes in the two QTP species are minimal and very low (Fig. 3b), as shown in the above results (Fig. 3a). These results demonstrate that more ancestral relaxed selection genes were under natural selection (specifically strongly purifying selection) in *Triplophysa* species endemic to QTP. The results and the fact that *Triplophysa* species have a strong ability to adapt to various extreme environments make it difficult to be convinced that strongly purifying selection was only to eliminate the massive, slightly deleterious mutations caused by genome-wide relaxed selection in the ancestral species of *Triplophysa* to make them have stronger and faster adaptability. Therefore, the relaxed selection of the common ancestral population of *Triplophysa* might have provided genetic advantages for the extreme environmental adaptation of *T. Tibetana* and *T. siluroides*. Finally, we measured the gene expression in the brain and liver of *T. siluroides* using published transcriptome data. We found that the strongly purifying selection genes that experienced significant relaxed selection from common ancestry had higher expression levels than background genes in *T. siluroides* (Fig. S5). Although the transcriptome data for *T. siluroides* were derived from domesticated individuals in low-altitude laboratories. The expression of these screened genes is higher in the liver than in the brain (Fig. S5) in similar elevation environment. This indicates that these genes may play an active role in adapting to the new environment.

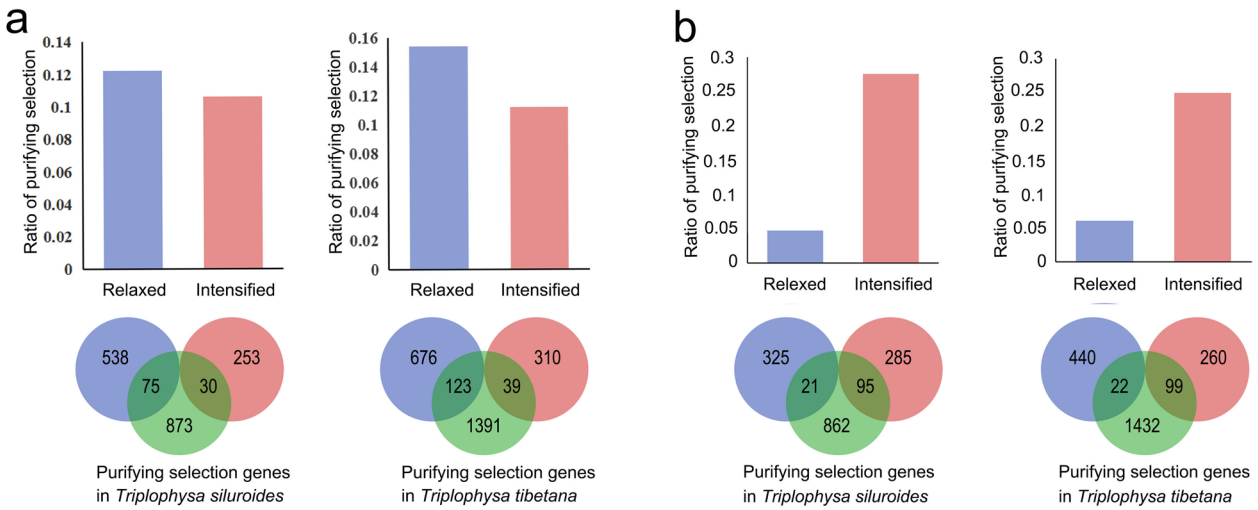

Fig. 3 More ancestral relaxed selection genes were under strongly purifying selection in *Triplophysa* that endemic to QTP. **a** More strongly purifying selection genes in *T. siluroides* and *T. tibetana* were derived from significantly relaxed selection genes of the common ancestors of *Triplophysa*; Top half: the ratio of significantly relaxed selection genes and Intensified selection genes ($P < 0.05$) from the common ancestors of *Triplophysa* in strongly purifying selection (dN/dS < 0.1; $P < 0.01$; no positive selection sites) of extant QTP species; Bottom half: the Venn diagram of relaxed selection genes, intensified selection genes in common ancestry of *Triplophysa* and strongly purifying selection in extant QTP species. **b** More strongly purifying selection genes overlapped with significantly Intensified selection genes in *T. siluroides and T. tibetana*; Top half: the ratio of significantly relaxed selection genes and Intensified selection genes ($P < 0.05$) in strongly purifying selection (dN/dS < 0.1; $P < 0.01$; no positive selection sites) of extant QTP species; Bottom half: the Venn diagram of relaxed selection genes, intensified selection genes and strongly purifying selection in extant QTP species

Functional analysis of candidate genes in *Triplophysa*

We conducted GO enrichment analysis of the genes under strongly purifying selection and the QTP-specific predominant expression genes in *T. siluroides* and *T. tibetana*, which experienced significant relaxed selection in the common ancestry of *Triplophysa*. The results showed that the genes with predominant expression in the brain were not significantly enriched. In contrast, the dominant expression genes in the liver had 14 significant enrichment GO terms, most of which were related to metabolism (Fig. 4a). In addition, it is worth noting that the "oxidation-reduction process" and "response to oxygen-containing compound" are also significantly enriched (Table S1), and these results are consistent with the previous transcriptome analysis of QTP species [1, 49]. Among the target evolutionary genes of *T. tibetana*, we found that 47 GO terms were significantly enriched (Table S2). The functions of these genes were related to development, such as nervous system development, kidney development, and multi-organ processes (Fig. 4b). Notably, "positive regulation of hematopoietic stem cell differentiation" and "positive regulation of hematopoietic progenitor cell differentiation" were also significantly enriched, possibly related to the hypoxic environment at high altitude.

While screening the function of genes in the GO term "positive regulation of hematopoietic stem cell differentiation", we discovered an interesting gene called supt5h. Supt5h, also known as the "SPT5 Homolog, DSIF Elongation Factor Subunit", is a transcription factor that regulates mRNA processing and transcription elongation by RNA polymerase II. The gene was under significantly relaxed selection in the common ancestry of *Triplophysa* (k value: 0.87; *P* value: 0.0059) but was under significantly purifying selection in *T. siluroides* and *T. tibetana* (dN/dS for *T. siluroides* and *T. tibetana*: 0.025 and 0.01; *P* value: 2.71E-15). Many cases and studies have shown that in humans, mutations or inactivation of this gene can lead to abnormal levels of Hb A2 and thalassaemia [50–56]. It means that the gene can affect the formation of hemoglobin, which affects oxygen transport. The evolution of supt5h may be related to the adaptability of *Triplophysa* fish to the extreme hypoxic environment at QTP.

Discussion

Our study shows that the QTP endemic species and other species involved in *Triplophysa* have similar characteristics of rapid evolution in protein-coding genes, and the same is true for ancestral branches. We detected a strong tendency of relaxed selection in the common ancestor of *Triplophysa*, which is consistent with the rapid evolution. In endemic species of the QTP environment, genes that have experienced natural selection (specifically, strongly purifying selection)

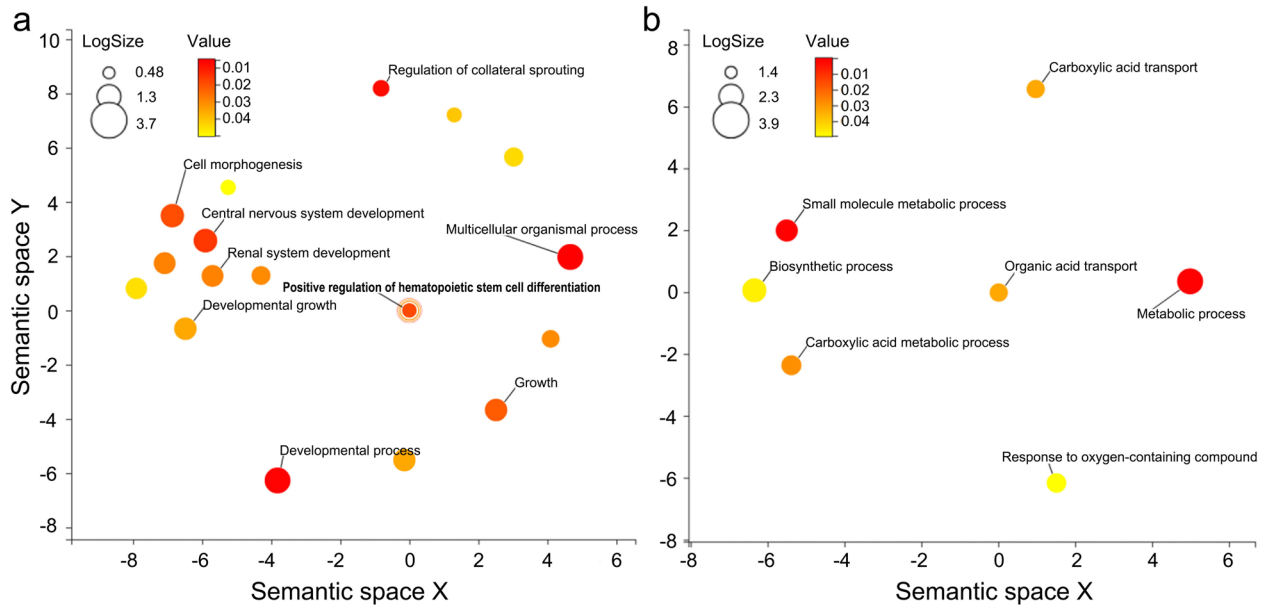

Fig. 4 GO analyses of candidate genes. **a** The genes experienced strongly purifying selection in extant QTP species but under significantly relaxed selection in the common ancestor of *Triplophysa*. **b** The function of dominant expression gene in *T. siluroides* liver

show a higher level of relaxed selection in ancestral genes compared to intensified selection. These genes are more commonly associated with development and liver metabolism. What is more noteworthy is that the GO term "positive regulation of hematopoietic stem cell differentiation" has been significantly enriched. The gene supt5h, contained in this GO term, is associated with hemoglobin production and red blood cell development.

In this research, we used extant species in different habitats to infer ancestral evolutionary profiling. The rapid evolution of *Triplopyhsa* ancestral genomes (or genes) might contribute to adaptation in QTP of extant species. Recent research has shown that the CDS, introns, and potential regulatory regions from the common ancestor genome of different species of snowfinches in the QTP were characterized by rapid evolution. The rapid evolution of developmental and metabolic-related genes provides the genetic basis for the snowfinches adapted to the extreme plateau environment [7, 8]. However, it should be noted that the earlier ancestor of snowfinch, namely, the snowfinch-tree-sparrow ancestor, also has the characteristics of rapid evolution of CDS, introns, and potential regulatory regions, consistent with our results. This suggests that the rapid evolution of earlier ancestors may have provided genetic resources for the adaptation of the QTP endemic *Triplophysa* species. This further supports the idea that genetic adaptations from ancestral species played a crucial role in the adaptation of extant species to their environments.

The rapid evolution caused by the relaxed selection of ancestral genomes provides possible genetic resources for rapidly forming new adaptive phenotypes. Many recent studies have demonstrated the role of relaxed selection in adapting populations to extreme environments [36–38] and the rapid formation of new phenotypes [39, 40].Relaxed selection will increase population diversity (or mutation load) [36, 41, 42]. The accumulation of slightly deleterious mutations provides the possibility for the generation of new phenotypes [30, 43]. Moreover, relaxed selection proves advantageous for phenotypic plasticity and the inherent diversity it contains [44–46]. Phenotypic plasticity can play a vital role in initial adaptation to a new extreme environment [47, 48]. Therefore, relaxed selection provides a probability for species to quickly adapt to new environments, whether from the perspective of slightly deleterious mutations or phenotypic plasticity. The adaptation of cavefish to extreme environments is associated with pervasive relaxation of natural selection in the genome [21], and the marks of relaxed selection tend to have existed before the cavefish entered the underground cave [57, 58]. The species of *Triplophysa* have shown the ability to adapt to a variety of extreme environments, including underground caves, high-salt lakes, and the Qinghai-Tibet Plateau. This adaptability may be linked to the high diversity and rapid evolution of the ancestral population of *Triplophysa*, possibly due to relaxed selection in the genome. The rapid evolution in the genome of extant *Triplophysa* species is likely connected to a reduced effective population size.

We cannot estimate the effective population size of the common ancestor of *Triplophysa*, which originated more than 20 million years ago based on current methods. However, our results show that the size of effective populations in the ancestors of *Triplophysa* fish 2–3 million years ago was not low. Moreover, the effective population size of the common ancestor of *Triplophysa*, which has evolved into more than two hundred species that can adapt to various complex environments, might not be small. Therefore, the increase in population diversity and rapid evolution caused by relaxed selection might be an essential factor in the adaptability of *Triplophysa* species to various extreme environments and the Qinghai-Tibet Plateau (QTP) environment.

In this study, we analyzed candidate genes for QTP adaptation based on changes in the selection pressure of genes in ancestral populations and extant species. The change in natural selection accompanies the change in environment. After entering the new environment, the mutations that are favorable for survival will persist and exhibit positive selection. Conversely, the unfavorable mutations for survival will be eliminated and demonstrate purifying selection [41, 59, 60]. However, the genes undergoing positive or purifying selection are crucial for adapting to new environments. Despite this, limited research has demonstrated the significant impact of purifying selection in generating adaptive traits, with the exception of studies on the extended lifespan and increased body size of mammals [61]. Many studies have demonstrated the role of positive selection in environmental adaptation, but the role of purifying selection has received less attention. One possible reason is that many housekeeping genes are also in a state of purifying selection. In this study, we investigate the potential molecular evolution underlying QTP adaptation by comparing gene selection pressures between ancestral genome and extant species. More relaxed selection genes from the common ancestor of *Triplophysa* experienced strongly purifying selection in extant *Triplophysa* species. The changes in the rate of evolution (or selection pressure) suggest that these genes may play a role in the adaptation of *Triplophysa* to the Qinghai-Tibet Plateau (QTP).

The extreme anoxic environment of QTP brings great challenges to the survival of organisms. The fish in QTP also need to overcome the extremely anoxic environment through mutations of some genes favourable to the QTP environment [11, 49, 62, 63]. Many studies have shown that the adaptation of QTP endemic mammals to the QTP environment with low oxygen is related to the mutation of several genes [4, 64–69]. However, the mutations of these genes still cannot explain the complex physiological phenotype of QTP animals or humans that adapted to low-oxygen environments completely [6, 69, 70]. Moreover, the molecular mechanisms of each QTP species or population adapted to the hypoxic environment were differed [4, 69, 70]. It is still meaningful to fully understand the adaptation mechanism to high-altitude environments from different species and research perspectives. In this study, we obtained some candidate genes related to QTP adaptation whose function (such as development and metabolism) was consistent with those of candidate genes related to high-altitude adaptation obtained from other species [7, 8, 71, 72]. The discovery of the setp5h candidate gene may supplement the explanation of vertebrate adaptation to the low oxygen environment in the plateau from another perspective. Our work might provide a new perspective for studying plateau extreme adaptive evolution.

Our study has limitations. We can only indirectly infer that the effective population size of the common ancestor of *Triplophysa* is likely not low. This inference is based on existing population genetic analyses and the presence of numerous species within the *Triplophysa* genus. Therefore, there is a lack of direct evidence.Secondly, further functional validation is required to confirm that the genes under relaxed selection in the ancestor indeed harbor loci that allow extant native QTP species to adapt to the QTP environment. Lastly, our analyses were based solely on published transcriptome data, which may limit the robustness of our findings.Nonetheless, our results may offer new insights into the evolution of QTP adaptation.

Methods

Data acquisition, phylogenomic analysis and demographic history analysis

We selected 15 species, including six *Triplophysa*, seven other loaches, a cyprinoid species, zebrafish, and electric eel as outgroup. These genomes were obtained from public databases (Table S3).The identification of orthologs was described in previous work [21]. The longest transcript and protein sequence were used for analysis for each gene.These protein sequences were used for gene family clustering analysis by OrthoMCL v1.4 [73] with the following parameters: orthomclFilterFaster: 10 20, blastp: Evalue 1e-5, coverage 50%, and mcl: -l 1.5. The gene family tree was constructed by RAxML (8.2.12 raxmlHPC-PTHREADS-AVX) [74].The UPhO software was used to find orthologs without inparalogs from input gene family trees [75]. After filtering the ortholog groups that deviate from the phylogeny in previous studies [23, 25].We obtained 14,982 orthologous groups (containing at least eight species) from the 15 genomes. Next, all protein-coding sequences in each orthogroup were extracted and aligned using TranslatorX 1.1 [76], which internally runs PRANK v.140,110 [77] for alignments and Gblocks 0.91b [78] to eliminate unreliable regions.

Based on the above results, we obtained multiple sequence alignment files of single-copy orthologous genes that include all species. Genes with CDS shorter than 270 bp or containing erroneous alignment regions were excluded.Finally, we obtained 1781 high-confidence single-copy orthologous genes. These protein-coding sequence alignments were concatenated to construct a phylogenetic tree with RAxML [74] using the GTRGAMMA model and the maximum likelihood method.We used one soft-bound calibration time point (zebrafish-electric eel: ~130–174 million years ago (Ma)) acquired from the TimeTree website (http://www.timetree.org/). The MCMCTree in PAML (version 4.9e) package [79] was used to estimate the divergence time between 15 species.

We downloaded genome sequencing data from five species of *Triplophysa* and four other loches from the NCBI database (*T. tibetana*: SRR8118711; *T. dalaica*: SRR11526794; *T. siluroides*: SRR9089996; *O. daqikongensis*: SRR8204517; *B. kweichowensis*: SRR12366070; *M. mizolepis*: SRR15343681; *M. anguillicaudatus*: SRR30107141; *T. bleekeri*: SRR9179728). Subsequently, the sequencing reads were filtered by using FastQC and Trimmomatic [80].BWA [81] was used to align the filtered data to the reference genome. Samtools [82] was then employed for filtering and sorting the results, with the filtering parameters set as follows: -b -h -F 4 -q 20 -F 256. Following this, Bcftools [83] was utilized to analyze SNPs and generate diploid consensus sequences.Finally, PSMC [84] was applied to estimate the historical dynamics of the population of each species, using the following parameters: psmc -N25 -T15 -R5, and psmc_plot.pl -u 4.13e-09 -g 2.

Evolutionary rate analysis

To understand the evolution rate of genes in each *Triplophysa* species and its common ancestry, dN/dS values in each branch were estimated using codeml in the PAML packages (version 4.9e) with default parameters and the free ratio model based on the phylogenetic tree with all 15 genomes. Next, we performed selection pressure analysis for each gene. For the subsequent PAML analysis, each gene's phylogenetic tree was pruned from the 15-genome phylogenetic tree, which was generated by 1781 high-confidence single-copy orthologous genes. We first compared the two-ratio model (model = 2 NSsites = 0) and the same model with fix_omega = 1 and omega = 1. We also use the branch site model (model = 2 NSsites = 2) to test whether the gene has positive selection sites. Finally, we used a chi-square distribution with 1 degree of freedom (or the degree of freedom from the results) to estimate each LRT analytical probability (*p*-value). In order to facilitate subsequent analysis, genes with dN/dS < 0.1 and $P < 0.01$ calculated by the branch model and no positive selection site detected in the branch site model were screened out and named as strongly purifying selection genes.

Analysis of relaxed selection

To further understand the role of relaxed selection in the accelerated evolution of common *Triplophysa* ancestry, we used the parameter k to measure whether genes were under relaxed selection. The parameter k was calculated using the hyphy RELAX program [28], and its relationship with the value of ω (dN/dS) was (ω background) k = ω foreground.If k < 1 indicates that the foreground branch was under relaxed selection, and if k > 1 indicates that the foreground branch was under intensified selection. Firstly, we used HyPhy RELAX to obtain K values for each species and its internal nodes based on 1781 single copy orthologous genes with default parameters, which could infer a free relaxation parameter k value for every branch.

Next, all of 14,982 were used for RELAX analysis. We used RELAX for LRT analysis by comparing the model with k = 1 to the model with k < 1 (or k > 1).If k < 1, $p < 0.05$ indicated that the test branch was under significantly relaxed selection, and if k > 1, $p < 0.05$ indicated that the test branch was under significantly intensified selection. We conducted multiple rounds of test analysis to test the relaxed selection of target species, branches groups or internal branches as follows: First, all branches of *Triplophysa* (including all internal branches and species) and all other loaches (including all internal branches and species) were alternatively set as the foreground branch, while zebrafish and electric eel were set as the unclassified branch.Then, we alternately set branch 5 and branch 6 as foreground branches (or background branches) respectively, while all other branches and their internal branches were classified as unclassified to analyze the significant relaxation selection of the ancestral *Triplophysa*. Finally, the two QTP native species of *Triplophysa* were alternatively set as foreground branches, other species of *Triplophysa* were set as background branches, and non-*Triplophysa* species were classified as unclassified for analysis.During the analysis of one extant QTP *Triplophysa* species, another QTP *Triplophysa* species was removed.

We evaluated the potential harmfulness of nonsynonymous mutations (or amino acid mutations) by analyzing changes in hydrophilicity or hydrophobicity of amino acid site mutations based on 1781 high-confidence single-copy orthologs as described in previous studies [33, 34]. We used codeml (aaml) to predict the ancestor sequence of each internal branch, with the RateAncestor parameter set to 1.We then score the change in hydrophobicity from

ancestral amino acids to replacement amino acids based on a score matrix of hydrophobicity (or hydrophilicity) [35]. Finally, the differences in scores between the ancestry of Triploplysa and the ancestry of *Triplophysa* and Barbatula barbatula were analyzed and compared. The significant difference test was conducted using a generalized linear mixed model from the lme4 v1.1–23 package in R (version 3.63) [85].

Transcriptome analysis

The transcriptome sequencing data were obtained from public databases, including the reads from brain and liver tissues of three species, namely *T. dalaica* (SRR11526792, SRR11526793), *T. bleekeri* (SRR6296205, SRR19866132), and *T. siluroides* (SRR10248769, SRR10248766). The *T. dalaica* transcriptome was obtained directly from frozen wild specimens [49]. The *T. bleekeri* specimen was briefly maintained in the laboratory after collection from its natural habitat, and then dissected. This was done due to the close proximity and similarity of the *T. bleekeri* habitat to the laboratory [26, 86]. The *T. siluroides* specimen, which was lab-domesticated [87]. We only screened the longest-length CDS of each gene for expression analysis based on the whole genome and annotations of the species mentioned above. The Bowtie v1.3.1 [88] was used for reads alignment, and then the RSEM v1.3.1 [89] was used to quantify gene expression to obtain the FPKM and reads count. We used edgR [90] for differentially expressed genes analysis.The gene was identified as the dominant expression gene in the QTP environment, as the expression level of this gene in *Triplophysa* siluroides was significantly higher than in *Triplophysa* dalaica and *Triplophysa bleekeri*. Subsequently, GO enrichment analysis and visualization were conducted using OmicShare web software (https://www.omicshare.com/tools) on the dominant expression genes. These genes were found to be under significantly relaxed selection in the common ancestry of *Triplophysa*, but under strongly purifying selection in QTP extant species.A q-value of ≤ 0.05 (FDR adjusted *p*-value) indicates that the GO category was significantly enriched. The REVIGO [91] was used to remove redundant terms in the significantly enriched GO categories.

Ethics

This work did not require ethical approval from a human subject or animal welfare committee.

Declaration of AI use

We have not used AI-assisted technologies in creating this article.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-025-11290-w.

Supplementary Material 1.

Supplementary Material 2.

Supplementary Material 3.

Supplementary Material 4.

Authors' contributions

W.Y. funding acquisition; W.Y and Q.Y. conceptualisation, supervision, writing-review and editing; Z.Q.Y. and X.F. formal analysis, methodology, visualisation, revision; H.Q.Y. and W.L.L. data curation, resources, validation; G.K.N. and Z.C. project administration validation and visualisation. All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Funding

This work was supported by the National Science and Technology Major Project for Transgenic Organisms (2016ZX08009-003).

Data availability

The genomic data used in this research were downloaded from public databases; for details, see Table S3. The scripts can be downloaded from https://github.com/zqingyuan/Comparative-genomics-on-the-adaptive-evolution-of-cavefish.git.

Declarations

Competing interests

The authors declare no competing interests.

Received: 16 July 2024 Accepted: 24 January 2025
Published online: 10 February 2025

References

1. Yang L, Wang Y, Sun N, Chen J, He S. Genomic and functional evidence reveals convergent evolution in fishes on the Tibetan Plateau. Mol Ecol. 2021;30:5752–64.
2. Qu Y, et al. Ground tit genome reveals avian adaptation to living at high altitudes in the tibetan plateau. Nat Commun. 2013;4:2071.
3. Zhang Z, et al. Convergent evolution of rumen microbiomes in high-altitude mammals. Curr Biol. 2016;26:1873–9.
4. Witt KE, Huerta-Sánchez E. Convergent evolution in human and domesticate adaptation to high-altitude environments. Philos Trans R Soc Lond B Biol Sci. 2019;374:20180235.
5. Xiong Y, Fan L, Hao Y, Cheng Y, Chang Y, Wang J, Lin H, Song G, Qu Y, Lei F. Physiological and genetic convergence supports hypoxia resistance in high-altitude songbirds. PLoS Genet. 2020;16:e1009270.
6. Sharma V, Varshney R, Sethy NK. Human adaptation to high altitude: a review of convergence between genomic and proteomic signatures. Hum Genomics. 2022;16:21.
7. Hao Y, Song G, Zhang YE, Zhai W, Jia C, Ji Y, Tang S, Lv H, Qu Y, Lei F. Divergent contributions of coding and noncoding sequences to initial high-altitude adaptation in passerine birds endemic to the Qinghai-Tibet Plateau. Mol Ecol. 2023;32:3524–40.
8. Qu Y, et al. The evolution of ancestral and species-specific adaptations in snowfinches at the Qinghai-Tibet Plateau. Proc Natl Acad Sci U S A. 2021;118:e2012398118.
9. Toda Y, et al. Early origin of sweet perception in the songbird radiation. Science. 2021;373:226–31.
10. Ma X, Dai W, Kang J, Yang L, He S. Comprehensive transcriptome analysis of six catfish species from an altitude gradient reveals adaptive evolution in Tibetan fishes. G3 (Bethesda). 2015;G3(6):141–8.

11. Wang Y, Yang L, Zhou K, Zhang Y, Song Z, He S. Evidence for adaptation to the Tibetan Plateau inferred from tibetan Loach Transcriptomes. Genome Biol Evol. 2015;7:2970–82.
12. Yang Y, Wang L, Han J, Tang X, Ma M, Wang K, Zhang X, Ren Q, Chen Q, Qiu Q. Comparative transcriptomic analysis revealed adaptation mechanism of *Phrynocephalus erythrurus*, the highest altitude lizard living in the Qinghai-Tibet Plateau. BMC Evol Biol. 2015;15:101.
13. Fricke R, Eschmeyer WN, Van der Laan R, editors. Eschmeyer's catalog of fishes: genera, species, references. 2024. (http://researcharchive.calacademy.org/research/ichthyology/catalog/fishcatmain.asp). Electronic version accessed dd mmm.
14. Qian Y, et al. The role of Introgression during the radiation of endemic fishes adapted to living at extreme altitudes in the Tibetan Plateau. Mol Biol Evol. 2023;40:msad129.
15. Chen YY, Chen YF, Liu HZ. Studies on the position of the Qinghai-Tibet plateau region in zoogeographic divisions and its eastern demarcation line. Acta Hydrobiol Sin. 1996;20:97–103.
16. Zhu S. The loaches of the Subfamily Nemacheilinae in China (Cypriniformes: Cobitidae). Nanjing: Jiangsu Science and Technology Publishing House; 1989.
17. Wu YF, Tan QJ. Characteristics of the fish-fauna of the characteristics of Qinghai-Xizang plateau and its geological distribution and formation. Acta Zool Sinica. 1991;37:135–52.
18. Yang X, et al. Chromosome-level genome assembly of *Triplophysa tibetana*, a fish adapted to the harsh high-altitude environment of the Tibetan Plateau. Mol Ecol Resour. 2019;19:1027–36.
19. Wang ZJ, Huang J, Zhang YG. The reproductive traits of *Triplophysa bleekeri* in the Daning River. Freshw Fisheries. 2013;43:8–12.
20. Zhou C, et al. The chromosome-level genome of *Triplophysa dalaica* (Cypriniformes: Cobitidae) provides insights into its survival in extremely alkaline environment. Genome Biol Evol. 2021;13:evab153.
21. Zhao Q, Shao F, Li Y, Yi SV, Peng Z. Novel genome sequence of Chinese cavefish (*Triplophysa rosa*) reveals pervasive relaxation of natural selection in cavefish genomes. Mol Ecol. 2022;31:5831–45.
22. Wang X, Song Y, Xie H, Zi F, Chen S, Luo S. Complete mitogenome of the *Triplophysa bombifrons*: comparative analysis and phylogenetic relationships among the members of *Triplophysa*. Genes (Basel). 2023;14:128.
23. Chen S, Ya N, Xie C, Wang S, Ren D. Complete mitochondrial genome of the *Triplophysa (Hedinichthys) yarkandensis (day)*. Mitochondrial DNA B Resour. 2016;1:235–6.
24. Hughes LC, et al. Comprehensive phylogeny of ray-finned fishes (Actinopterygii) based on transcriptomic and genomic data. Proc Natl Acad Sci U S A. 2018;115:6249–54.
25. Yu P, Zhou L, Yang WT, Miao LJ, Li Z, Zhang XJ, Wang Y, Gui JF. Comparative mitogenome analyses uncover mitogenome features and phylogenetic implications of the subfamily Cobitinae. BMC Genomics. 2021;22:50.
26. Wang Y, Shen Y, Feng C, Zhao K, Song Z, Zhang Y, Yang L, He S. Mitogenomic perspectives on the origin of tibetan loaches and their adaptation to high altitude. Sci Rep. 2016;6:29690.
27. Yuan D, Chen X, Gu H, Zou M, Zou Y, Fang J, Tao W, Dai X, Xiao S, Wang Z. Chromosomal genome of *Triplophysa bleekeri* provides insights into its evolution and environmental adaptation. Gigascience. 2020;9:giaa132.
28. Wertheim JO, Murrell B, Smith MD, Kosakovsky Pond SL, Scheffler K. RELAX: detecting relaxed selection in a phylogenetic framework. Mol Biol Evol. 2015;32:820–32.
29. Kimura M. The neutral theory of molecular evolution. Cambridge: Cambridge University Press; 1983. https://doi.org/10.1017/CBO9780511623486.
30. Ohta T. Slightly deleterious mutant substitutions in evolution. Nature. 1973;246:96–8.
31. Ohta T. The nearly neutral theory of molecular evolution. Annu Rev Ecol Syst. 1992;23:263–86.
32. Pace CN, et al. Contribution of hydrophobic interactions to protein stability. J Mol Biol. 2011;408:514–28.
33. Henry L, Schwander T, Crespi BJ. Deleterious mutation accumulation in asexual Timema stick insects. Mol Biol Evol. 2012;29:401–8.
34. Brandt A, Schaefer I, Glanz J, Schwander T, Maraun M, Scheu S, Bast J. Effective purifying selection in ancient asexual oribatid mites. Nat Commun. 2017;8:873.
35. Riek RP, Handschumacher MD, Sung SS, Tan M, Glynias MJ, Schluchter MD, Novotny J, Graham RM. Evolutionary conservation of both the hydrophilic and hydrophobic nature of transmembrane residues. J Theor Biol. 1995;172:245–58.
36. Cui R, Medeiros T, Willemsen D, Iasi LNM, Collier GE, Graef M, Reichard M, Valenzano DR. Relaxed selection limits lifespan by increasing mutation load. Cell. 2020;180:1272–9.
37. Li SJ, Hua ZS, Huang LN, Li J, Shi SH, Chen LX, Kuang JL, Liu J, Hu M, Shu WS. Microbial communities evolve faster in extreme environments. Sci Rep. 2014;4:6205.
38. Iverson ENK, Criswell A, Havird JC. Stronger evidence for relaxed selection than adaptive evolution in high-elevation animal mtDNA. bioRxiv. 2024;23:2024.01.20.576402. https://doi.org/10.1101/2024.01.20.576402.
39. Lahti DC, Johnson NA, Ajie BC, Otto SP, Hendry AP, Blumstein DT, Coss RG, Donohue K, Foster SA. Relaxed selection in the wild. Trends Ecol Evol. 2009;24:487–96.
40. Dapper AL, Wade MJ. Relaxed selection and the rapid evolution of reproductive genes. Trends Genet. 2020;36:640–9.
41. Cvijović I, Good BH, Desai MM. The effect of strong purifying selection on genetic diversity. Genetics. 2018;209:1235–78.
42. Peischl S, et al. Relaxed selection during a recent human expansion. Genetics. 2018;208:763–77.
43. Charlesworth J, Eyre-Walker A. The other side of the nearly neutral theory, evidence of slightly advantageous back-mutations. Proc Natl Acad Sci U S A. 2007;104:16992–7.
44. Hunt BG, Ometto L, Wurm Y, Shoemaker D, Yi SV, Keller L, Goodisman MA. Relaxed selection is a precursor to the evolution of phenotypic plasticity. Proc Natl Acad Sci U S A. 2011;108:15936–41.
45. Leichty AR, Pfennig DW, Jones CD, Pfennig KS. Relaxed genetic constraint is ancestral to the evolution of phenotypic plasticity. Integr Comp Biol. 2012;52:16–30.
46. Schrader L, et al. Relaxed selection underlies genome erosion in socially parasitic ant species. Nat Commun. 2021;12:2918.
47. Chevin LM, Hoffmann AA. Evolution of phenotypic plasticity in extreme environments. Philos Trans R Soc Lond B Biol Sci. 2017;372:20160138.
48. Ho WC, Zhang J. Evolutionary adaptations to new environments generally reverse plastic phenotypic changes. Nat Commun. 2018;9:350.
49. Wang Y, Yang L, Wu B, Song Z, He S. Transcriptome analysis of the plateau fish (*Triplophysa dalaica*): implications for adaptation to hypoxia in fishes. Gene. 2015;565:211–20.
50. Achour A, et al. A new gene associated with a β-thalassemia phenotype: the observation of variants in SUPT5H. Blood. 2020;136:1789–93.
51. Charnay T, Cerino M, Gonnet K, Bonello-Palot N, Bréchard MP, Badens C. A novel SUPT5H variant associated with a beta-thalassaemia trait. Br J Haematol. 2022;196:e70–1.
52. Harteveld CL, et al. The hemoglobinopathies, molecular disease mechanisms and diagnostics. Int J Lab Hematol. 2022;44:28–36.
53. Martell DJ, et al. RNA polymerase II pausing temporally coordinates cell cycle progression and erythroid differentiation. Dev Cell. 2023;58:2112–e21274.
54. Lou J, Ye Y, Sun M, Zhao Y, Fu Y, Liu Y. A stepwise haematological screening and whole-exome sequencing reveal multiple mutations from SUPT5H causing an elevation of Hb A2 from a cohort of 47336 individuals. Int J Lab Hematol. 2023;45:90–5.
55. Xiao ZQ, Jiang F, Li DZ. β-Thalassemia trait caused by SUPT5H defects: another case report. Hemoglobin. 2023;47:145–6.
56. Lin Z, Liang X, Wei X, Liang G, Zhu D, Xie H, Yan T, Shang X. SUPT5H mutations associated with elevation of hb A2 level: identification of two novel variants and literature review. Gene. 2024;908:148294.
57. Liu S, Ludwig A, Peng Z. Nine novel microsatellites for the cavefish (*Triplophysa rosa* Chen & Yang, 2005). J Appl Ichthyol. 2017;33:119–20.
58. Moran RL, et al. Selection-driven trait loss in independently evolved cavefish populations. Nat Commun. 2023;14:2557.
59. Hurst L. Genetics and the understanding of selection. Nat Rev Genet. 2009;10:83–93.
60. Vermeij GJ. When does natural selection take place? Evolution. 2023;77:1739–43.
61. Kowalczyk A, Partha R, Clark NL, Chikina M. Pan-mammalian analysis of molecular constraints underlying extended lifespan. Elife. 2020;9:e51089.
62. Yang L, Wang Y, Zhang Z, He S. Comprehensive transcriptome analysis reveals accelerated genic evolution in a Tibet fish, *Gymnodiptychus pachycheilus*. Genome Biol Evol. 2014;7:251–61.

63. Chen J, Shen Y, Wang J, Ouyang G, Kang J, Lv W, Yang L, He S. Analysis of multiplicity of Hypoxia-Inducible factors in the evolution of *Triplophysa* Fish (Osteichthyes: Nemacheilinae) reveals hypoxic environments adaptation to Tibetan Plateau. Front Genet. 2020;11:433.
64. Wu T, et al. Hemoglobin levels in Qinghai-Tibet: different effects of gender for tibetans vs. Han. J Appl Physiol (1985). 2005;98:598–604.
65. Yi X, et al. Sequencing of 50 human exomes reveals adaptation to high altitude. Science. 2010;329:75–8.
66. Huerta-Sánchez E, Casey FP. Archaic inheritance: supporting high-altitude life in Tibet. J Appl Physiol (1985). 2015;119:1129–34.
67. Huerta-Sánchez E, et al. Altitude adaptation in tibetans caused by introgression of Denisovan-like DNA. Nature. 2014;512:194–7.
68. Zhang X, Witt KE, Bañuelos MM, Ko A, Yuan K, Xu S, Nielsen R, Huerta-Sanchez E. The history and evolution of the Denisovan-EPAS1 haplotype in tibetans. Proc Natl Acad Sci U S A. 2021;118:e2020803118.
69. Lee FS. Hypoxia Inducible factor pathway proteins in high-altitude mammals. Trends Biochem Sci. 2024;49:79–92.
70. Pamenter ME, Hall JE, Tanabe Y, Simonson TS. Cross-species insights into genomic adaptations to Hypoxia. Front Genet. 2020;11:743.
71. Palacios C, et al. Genomic variation, population history, and long-term genetic adaptation to high altitudes in Tibetan Partridge (*Perdix hodgsoniae*). Mol Biol Evol. 2023;40:msad214.
72. Zhang T, et al. Comparative transcriptomic analysis delineates adaptation strategies of Rana kukunoris toward cold stress on the Qinghai-Tibet Plateau. BMC Genomics. 2024;25:363.
73. Li L, Stoeckert CJ Jr, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 2003;13:2178–89.
74. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30:1312–3.
75. Ballesteros JA, Hormiga G. A new orthology assessment method for phylogenomic data: unrooted phylogenetic orthology. Mol Biol Evol. 2016;33:2117–34.
76. Abascal F, Zardoya R, Telford MJ. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acids Res. 2010;38:W7–13.
77. Löytynoja A. Phylogeny-aware alignment with PRANK. Methods Mol Biol. 2014;1079:155–70.
78. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol. 2000;17:540–52.
79. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24:1586–91.
80. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20.
81. Jung Y, Han D, BWA-MEME. BWA-MEM emulated with a machine learning approach. Bioinformatics. 2022;38:2404–13.
82. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 1000 genome project data processing subgroup. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25:2078–9.
83. Danecek P, McCarthy SA. BCFtools/csq: haplotype-aware variant consequences. Bioinformatics. 2017;33:2037–9.
84. Nadachowska-Brzyska K, Burri R, Smeds L, Ellegren H. PSMC analysis of effective population sizes in molecular ecology and its application to black-and-white Ficedula flycatchers. Mol Ecol. 2016;25:1058–72.
85. Bates D, et al. Fitting Linear mixed-effects models using lme4. J Stat Softw. 2015;67:1–48.
86. Zhao Q, Zhang R, Xiao Y, Niu Y, Shao F, Li Y, Peng Z. Comparative transcriptome profiling of the loaches Triplophysa bleekeri and Triplophysa rosa reveals potential mechanisms of eye degeneration. Front Genet. 2020;10:1334.
87. Chen Y, Gong Q, Lai J, Song M, Liu Y, Wu Y, Ai J, Long Z. Transcriptome analysis identifies candidate genes associated with skin color variation in Triplophysa siluroides. Comp Biochem Physiol Part D Genomics Proteom. 2020;35:100682.
88. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9.
89. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12:323.
90. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26:139–40.
91. Supek F, Bošnjak M, Škunca N, Šmuc T. REVIGO summarises and visualises long lists of gene ontology terms. PLoS ONE. 2011;6:e21800.

Publisher's Note