


The Role of Introgression During the Radiation of Endemic Fishes Adapted to Living at Extreme Altitudes in the Tibetan Plateau

Yuting Qian,^{†1,2,3,4} Minghui Meng,^{†5} Chaowei Zhou,^{†6,7} Haiping Liu ^{*,8} Haifeng Jiang,⁹ Youwei Xu,¹⁰ Wenjun Chen,^{1,4} Zufa Ding,¹ Yang Liu,¹ Xiong Gong,^{1,4} Cheng Wang,^{1,4} Yi Lei,¹ Tai Wang,¹¹ Ying Wang,^{3,12} Xiaoni Gan,¹ Axel Meyer,¹³ Shunping He,^{*,1,3,14} and Liandong Yang^{*,1,2,3}

¹State Key Laboratory of Freshwater Ecology and Biotechnology, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan, China

²State Key Laboratory of Plateau Ecology and Agriculture, Qinghai University, Xining, China

³Academy of Plateau Science and Sustainability, Qinghai Normal University, Xining, China

⁴University of Chinese Academy of Sciences, Beijing, China

⁵Key Laboratory of Horticultural Plant Biology (MOE), College of Horticulture and Forestry Sciences, Huazhong Agricultural University, Wuhan, China

⁶College of Fisheries, Southwest University, Chongqing, China

⁷Key Laboratory of Freshwater Fish Reproduction and Development, Ministry of Education, Key Laboratory of Aquatics Science of Chongqing, Chongqing, China

⁸Integrative Science Center of Germplasm Creation in Western China (CHONGQING) Science City, Key Laboratory of Freshwater Fish Reproduction and Development (Ministry of Education), Key Laboratory of Aquatic Science of Chongqing, School of Life Sciences, Southwest University, Chongqing, China

⁹College of Animal Science and Technology, Northwest A&F University, Yangling, Shaanxi, China

¹⁰South China Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Guangzhou, China

¹¹Gansu Key Laboratory of Cold Water Fishes Germplasm Resources and Genetics Breeding, Gansu Fisheries Research Institute, Lanzhou, China

¹²Hubei Engineering Research Center for Protection and Utilization of Special Biological Resources in the Hanjiang River Basin, School of Life Sciences, Jiangnan University, Wuhan, China

¹³Department of Biology, University of Konstanz, Konstanz, Germany

¹⁴Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming, China

[†]These authors contributed equally to this work.

*Corresponding authors: E-mails: luhappy@163.com; clad@ihb.ac.cn; yangliandong1987@163.com.

Associate editor: Anne Yoder

Abstract

Recent genomic analyses of evolutionary radiations suggest that ancient introgression may facilitate rapid diversification and adaptive radiation. The loach genus *Triplophysa*, a genus with most species endemic to Tibetan Plateau, shows ecological diversity and rapid evolution and represents a potential example of adaptive radiation linked to the uplift of the Tibetan Plateau. Here, we interrogate the complex evolutionary history of *Triplophysa* fishes through the analysis of whole-genome sequences. By reconstructing the phylogeny of *Triplophysa*, quantifying introgression across this clade, and simulating speciation and migration processes, we confirm that extensive gene flow events occurred across disparate *Triplophysa* species. Our results suggest that introgression plays a more substantial role than incomplete lineage sorting in underpinning phylogenetic discordance in *Triplophysa*. The results also indicate that genomic regions affected by ancient gene flow exhibit characteristics of lower recombination rates and nucleotide diversity and may associate with selection. Simulation analysis of *Triplophysa tibetana* suggests that the species may have been affected by the Gonghe Movement in the third uplift of the Tibetan Plateau, resulting in founder effects and a subsequent reduction in N_e .

Key words: phylogenetic conflict, introgression, reticulation, recombination, *Triplophysa*.

© The Author(s) 2023. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Open Access

Introduction

Studies of phenotypic diversification and phylogenetic discordance have been foundational to our understanding of evolution. One classic example of these two phenomena is reticulate evolution during adaptive radiation (Mallet et al. 2016; Jamie and Meier 2020). Genomic analyses have shown that phylogenies of closely related species often show a high degree of reticulation, which is often associated with adaptive radiation (Fontaine et al. 2015; Meier et al. 2017; Edelman et al. 2019; Choi et al. 2021; Hu et al. 2022). Ancestral variation and subsequent hybridization generate the initial diversity that can then be sculpted by selection into an adaptive radiation (Berner and Salzburger 2015; Meier et al. 2017; Choi et al. 2021; Nelson et al. 2021). The potential for rapid adaptation to new environments and range expansion via adaptive introgression may increase with increasing amounts of genetic polymorphisms (Hedrick 2013).

Genetic exchange between lineages is generally assumed to hinder the evolution of reproductive isolation (Kulmuni et al. 2020), yet genetic exchange between lineages can also be a persistent feature of diversification of a clade and obscure the relationships among ancient branches in a phylogenetic tree (Irisarri et al. 2018; Yang et al. 2021). A phylogeny is used to visualize the evolutionary history of a set of taxa. However, phylogenetic discordance is widespread in phylogenetic analysis (Prum et al. 2015; Irisarri et al. 2017; Leebens-Mack et al. 2019), and the causes of this are manifold, difficult to characterize, and not always due to hybridization. Conflicting patterns between mitochondrial and nuclear genetic markers have been considered to be related to (sometimes adaptive), for example, introgression of mitochondrial DNA (mtDNA), demographic disparities, and sex-biased asymmetries (Toews and Brelsford 2012; Kang et al. 2013). Discordance between gene trees and species trees can also be caused by incomplete lineage sorting (ILS) or other phenomena such as horizontal gene transfers (Nichols 2001; Ruber et al. 2001; Avise and Robinson 2008; Solis-Lemus and Ane 2016). Some phylogenetic relationships are simply difficult to resolve because they involve ancient splits, with weak phylogenetic signal that occurred in short succession (Irisarri and Meyer 2016; Irisarri et al. 2017). Even the concatenation of extensive sequence data sets to analyze phylogenetic relationships in cases where lineages show frequent introgression may be insufficient for robust phylogenetic inference (Kubatko and Degnan 2007; Shen et al. 2021). Species trees can be constructed using the multispecies coalescent-based approach (Rannala and Yang 2003), but this approach might be affected by low-quality individual gene alignments to infer individual gene trees (Blom et al. 2017; Simmons and Kessenich 2020).

If introgression affects large parts of the genome, inferred phylogenies will be affected by this history of reticulate evolution in a given lineage (Mallet et al. 2016). Analyzing complete genomes in an effort to examine

adaptive radiations might allow to assess the role of ancestral genetic variation in speciation and phenotypic diversification (Meier et al. 2017; Svardal et al. 2020; Wang, Street, et al. 2020). The combination of natural selection and recombination can produce heterogeneous patterns of genomic divergence between nascent and recently evolved species (Meyer et al. 2006; Via and West 2008; Nosil et al. 2009). On one hand, genomic regions of low recombination linked with deleterious loci will be broken and purged, preventing the retention of traces of introgression suitable as a data source for reconstructing phylogenetic relationships (Schumer et al. 2018; Edelman et al. 2019; Li et al. 2019; Martin et al. 2019; Nelson et al. 2021). On the other hand, genomic regions that have undergone hybridization, introgression, and subsequent selection may make it difficult to accurately estimate species tree (Nater et al. 2015; Zhang et al. 2021).

The Tibetan Plateau is the youngest and highest plateau on Earth with the height average above 4,500 m sea level. As it uplifted, its geographical features and natural environment have experienced drastic change (Deng and Ding 2015; Su et al. 2019). *Triplophysa* is a genus of nemacheiline loaches (family Cyprinidae), the most species-rich lineage of freshwater fishes in Asia. *Triplophysa* currently includes about 140 species that can be found at altitudes up to 5,200 m. Hence, it is widely distributed across and span the entire Tibetan Plateau and adjacent regions in Central Asia (He et al. 2011; Wang et al. 2015; Li et al. 2017; Wang, Zhang, et al. 2020). This genus is variable in coloration and morphology, with different patches or spots on the body surface, and different anatomical features (Yang, Liu, et al. 2019; Yang, Wang, et al. 2019; Wang, Zhang, et al. 2020; Yuan et al. 2020; Zhou et al. 2021). They usually inhabit still waters or slow-flowing bottoms of rivers and lakes, feeding on aquatic invertebrates or algae (Bureau of fisheries of Tibet Autonomous Region 1995). Previous studies based on morphological and biogeographic data have found that changes in geography and climate associated with the uplifting of the Tibetan Plateau were, through allopatric speciation, the driving forces behind the formation of the species diversity of *Triplophysa* fishes (He et al. 2006). The main river systems in the Tibetan Plateau region—the Hexi River system, the upper reaches of the Huanghe River, and the Jialing River not only share several species of *Triplophysa* but also possess endemics (Wang and Chang 2012; Wang, Zhang, et al. 2020). Divergent populations of *Triplophysa* may come into secondary contact due to geological changes such as mountain upheaval and river diversion (Fang et al. 2005; Favre et al. 2015), resulting in genetic exchange (Feng et al. 2019).

Here, we use whole-genome sequences (WGS) of 18 representative species spanning all major lineages of the radiation of *Triplophysa* to explore the phylogenetic relationships between them and to gain deeper insights into the evolutionary processes that shaped it and characterized their genomic architectures. In this study, we estimated a comprehensive phylogenetic relationship of the

genus *Triplophysa*, then quantified signals of introgression, and simulated the history of speciation and migration events. We show the abundance of allele sharing across this clade, which may have played a crucial role during the adaptive radiation of *Triplophysa* fishes. We also find evidence of an ancient gene flow event occurred ~0.24–9.87 million years ago (Ma) between *Triplophysa tibetana* and a clade of *Triplophysa* fishes. Our analyses show that genomic regions under introgression reveal a characteristic of lower recombination rates and nucleotide diversity that have similarities with regions under selection. This is a tentatively exploration of genome-wide patterns of introgression and may help us understand the driving force behind the adaptation of *Triplophysa* fishes to the Tibetan Plateau.

Results

Summary of Sequencing, Mapping, SNP Calling, and Filtering

A total of 138 *Triplophysa* from 18 species were whole-genome resequenced on an Illumina platform, which contains the major lineages of *Triplophysa* species that have been reported in Northwest China (He et al. 2011). Most were collected from the northeastern edge of the Tibetan Plateau (Gansu Province), and some samples of *Triplophysa orientalis*, *Triplophysa bleekeri*, *Triplophysa stewarti*, and *T. tibetana* were collected from Lhasa (Xizang Autonomous Region) (supplementary table S2, Supplementary Material online). After quality control, ~3.91 Tb of whole-genome sequencing data (26.2 billion paired-end clean reads) were generated for 138 *Triplophysa* samples at an average depth of ~34.7× per sample. The mapping rate of these clean reads to the *T. tibetana* reference genome was $96.10\% \pm 1.83\%$, and the genome coverage was >80% across all chromosomes for the majority of samples (supplementary table S1, Supplementary Material online). Sequence reads were aligned to the *T. tibetana* reference genome (Yang, Liu et al. 2019) and variants called, generating a total of 11.6 million single-nucleotide variants (SNVs).

Population Structure and Genetic Differentiation Among the Species

We first described the patterns of genome-wide diversity using principal component analysis (PCA), which revealed a clear distinction among *T. tibetana* clade and grouped all individuals into four or five genetic clusters (fig. 1A). The first three principal components (PC1, PC2, and PC3; explained 21.8%, 17.3%, and 13.8%) separated *T. tibetana* (TTIB) from the other four clusters, but some *T. orientalis* (TORI) and *T. bleekeri* (TBLE) collected from the adjacent area, Lhasa, were closer to *T. tibetana*. It is worth noting that in all three principal components, *T. tibetana* were placed between to the main clusters, and many processes can lead to patterns of genetic intermediacy we observed, including hybridization (Liang et al. 2019; Liu et al. 2022).

We then estimated individual genetic assignment to detect ancient admixture or recent introgression. In the admixture plot at $K = 4$ (fig. 1B), the categories of population were consisted with major cluster in consensus with PCA analysis. At $K = 5$, *T. orientalis* and *T. tibetana* were divided into new genetic clusters. At $K \leq 16$ (supplementary fig. S2, Supplementary Material online), all clusters did not reveal evidence for strong admixture between species, except *T. tibetana* and three individuals of *T. orientalis*. This result suggests a pattern of genetic heterogeneity and could be caused by insufficient sample size in structure analysis (Dupuis and Sperling 2020; Hinojosa et al. 2022). The optimal K value is $K = 16$ (cross-validation error 0.21; supplementary fig. S3, Supplementary Material online), which grouped all individuals by species except *T. tibetana* and the cluster of *Triplophysa polyfasciata* and *T. bleekeri*. Some individuals of those two species were sampled from adjacent geographic areas, which may be the reason for this population structure.

Out-group_{f3} tests were used to measure the genetic relationships between all sampled *Triplophysa* species. The heat map of f_3 estimated values among pairs of species (supplementary fig. S4, Supplementary Material online) shows a similar pattern as the PCA results. *T. tibetana* had the highest genetic affinities to clade2, clade3, and clade4, particularly clade2 and *T. stewarti*. These results indicate that *T. tibetana* might have had genetic contact with these three clades after its initial divergence.

Phylogenetic Relationships Between the Species

Based on the nuclear genomic level data set, we used concatenation and multispecies coalescent approaches to further test the robustness of the inferred phylogenetic relationships between *Triplophysa* fishes. Five data sets from nuclear genomic sliding windows generated with partition strategies of nonoverlapping windows of 200, 100, 50, 25, and 10 kb and one data set of binned genes were used to infer local phylogenies for each window or gene. Then, we generated coalescent-based ASTRAL species trees with local trees or gene trees to reconstruct the consensus tree for each data set. We also concatenated all filtered single-nucleotide polymorphism (SNP) sites across the WGS and only protein coding sequences (CDS) and 4-fold degenerate (4D) sites to generate three concatenated data sets. Finally, we constructed the concatenation-based phylogenetic tree (supplementary fig. S5, Supplementary Material online).

All phylogenies generated from these two approaches were supported with high local posterior probability (LPP in ASTRAL) or bootstrap values (supplementary table S3, Supplementary Material online). Five species trees based on coalescent method and local tree data sets result in the identical phylogeny (fig. 1C), and the gene tree data sets supported a different phylogenetic structure. The concatenation-based approach reconstructed two different phylogenetic relationships. The phylogenies of the CDS data set and the 4D data set are consistent with

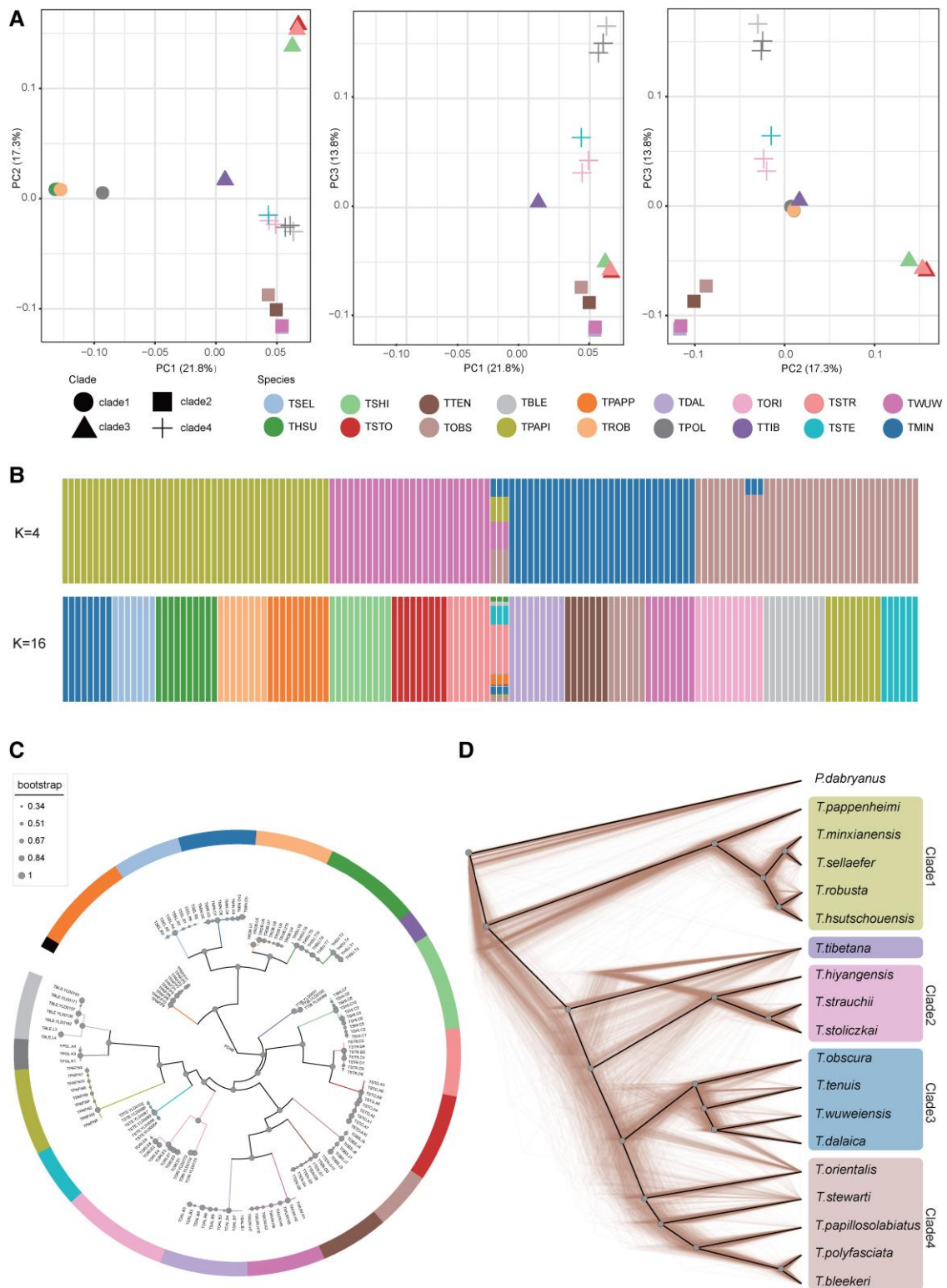


FIG. 1. Population structure and inferred phylogenetic relationships among *Triplophysa* species. (A) PCA of 138 *Triplophysa* individuals. (B) Population structure by STRUCTURE analysis assign individuals to five clusters with $K = 4$, and the optimal $K = 16$, which can assign most individuals to well-defined species. Each colored bar represents one individual, and colored segments represent proportions of genetic components. (C) The coalescent-based species tree. The node size represents ASTRAL LPP according to the scale shown on the left. The color of the branches refers to the species as in A and B (color scheme as in B, $K = 16$, except *T. tibetana* and *T. polyfasciata*). (D) Cladograms of the prevalent discordances trees among 1,000 randomly selected local trees derived from 200-kb windows, and solid black line represents the consensus tree calculated by DensiTree, which is consistent with Astral species tree (color scheme as in B, $K = 4$).

each other and the gene tree ASTRAL topology. The only discordant feature among all phylogenies was the position of *T. tibetana* (supplementary fig. S5, Supplementary Material online). Based on the distance from the outgroup, all four major clades in the phylogenetic tree could be clearly identified, except for the placement of *T. tibetana*. All five ASTRAL species trees supported a monophyletic clade formed by clade2, clade3, and clade4 as the sister to *T. tibetana*. Whereas the WGS tree supported a monophyletic clade clustered by clade1, clade2, clade3, and clade4 as sister to *T. tibetana*, the CDS/4D tree supported *T. tibetana* as sister to clade2.

Alternative Tree Topologies From Coalescent-Based Analysis

We explored the phylogenetic relationships of the *Triplophysa* species using TWISST (Martin and Van Belleghem 2017), which quantifies the frequency of alternative phylogenetic topologies in sliding windows along the genome. Our coalescent-based phylogenetic method inferred two kinds of species tree from local and gene tree sets. After merging clade3 and clade4 into one clade, the two topologies can be expressed as Topo A: (PDAB, (clade1, (TTIB, (clade2, clade3-4)))); and Topo B: (PDAB, (clade1, (clade3-, (clade2, TTIB)))); PDAB represents outgroup *Paramisgurnus dabryanus*. The most common topologies in sliding window trees and gene trees were also consistent with those two topologies, respectively (supplementary fig. S6a–f, Supplementary Material online). The ASTRAL species tree based on the normalized quartet score of 5 sliding windows reveals a positive correlation with window size (supplementary table S4, Supplementary Material online): As the window size decreased, the proportion of all the quartet trees that could be found in the species tree also decreased. Topology weighting analysis of window local maximum likelihood (ML) trees reveals that, in about 44.5–51.1% local trees, *T. tibetana* is sister to a cluster formed by clade2, clade3, and clade4 (Topo A) and, in about 37.6–43.5% local trees, *T. tibetana* and clade2 formed a cluster (Topo B) (supplementary fig. S6, Supplementary Material online). Topological discordance can also be revealed from the DensiTree plot of 1,000 random 200-kb local trees (fig. 1D). In addition to the discordant placement of *T. tibetana*, there are also different topologies within each major clade, which indicates a high degree of topological uncertainty (Bouckaert 2010). The different result was presented in topology weighting analysis of gene trees, which shows >50% of trees support Topo B and only 46% support Topo A. This discordance exemplifies that gene trees may differ from the species tree due to ILS or admixture between lineages and also reveals correlations among different data set sources used to construct phylogenetic trees.

Evidence for Rampant Past Introgression

In addition to ILS, phylogenetic conflict between genomic partitions may also be explained by hybridization (Cahill et al. 2013; Li et al. 2016). When interbreeding is not

restricted between immigrant and local species, introgression in neutral sites can take place on a wide genomic or geographic scale (Currat et al. 2008). To test whether the speciation process among *Triplophysa* was accompanied by past introgression, we calculated Patterson's *D*-statistic using ADMIXTOOLS (Patterson et al. 2012), which assesses the extent of genetic admixture among a quartet of species populations. A total of 816 quartets conformed to the species tree topology were detected (supplementary table S8, Supplementary Material online), and 100% of the quartets reached a threshold of $|Z| > 3$, which indicates a significant *D* value. In quartet topologies of the form (W, X, Y, OUT), W, X, and Y represent one species population from clade1, clade2, clade3, clade4, and *T. tibetana* in the species tree. We classified the triplets (W, X, Y) into species pairs, based on positive and negative values of the *D* value. Thus, we calculated mean *D*-values for species pairs using W and Y as a species pair if the *D*-value for the triplet was positive, and X and Y as a species pair if the *D*-value for the triplet was negative. We visualized the 136 pairs of *D*-value averages with a heat map (fig. 2A). The largest *D* value appears in the species pair of TTIB_TSTE (*T. stewarti*), which indicated a high level of genetic introgression between *T. stewarti* and *T. tibetana*. Following are the TTIB_TSTR (*Triplophysa strauchii*), TTIB_TSTO (*Triplophysa stoliczkai*), and TTIB_TSHI (*Triplophysa shiyangensis*) pairs, which feature species clustered in clade2. Similar *D* values were found between clade2 and *T. tibetana*, which suggests an episode of ancient gene flow between *T. tibetana* and the ancestral lineage of clade2. *T. stewarti* also experienced a significant gene flow event between itself and clade1/clade2 supported by high *D*-statistic values. We also detected significant gene flow between *T. orientalis* and all species in clade2/clade3, with high *D* values. Considering the chronological order of speciation (supplementary fig. S7, Supplementary Material online), we speculate that the gene flow event occurred between *T. orientalis* and the ancestral lineage of two clades. The species pairs showing incongruence across the phylogenies were supported by significant *D* values (fig. 1D), which suggests that gene flow might have contributed to the observed phylogenetic discordance.

To further distinguish whether the observed phylogenetic discordance were indeed due to introgression rather than ILS, we conducted gene tree set analysis in QuIBL (Edelman et al. 2019), which estimates observed distributions of internal branch lengths for triplet topologies. Among 816 triplets, we only found one triplet that showed a significant level of ILS (supplementary table S9, Supplementary Material online). Another 413 triplets showed significant levels of introgression ($\Delta BIC > 10$), accounting for 50.6% of the total triplet sets (supplementary table S9, Supplementary Material online). We also divided all the results into 132 species pairs and calculated the average of the total introgression proportion (totalIntroProp) for each species pair. We then visualized the 132 pairs of totalIntroProp averages in a heat map (fig. 2A) and marked introgression on the species tree with a red dotted line for the species pair of totalIntroProp >

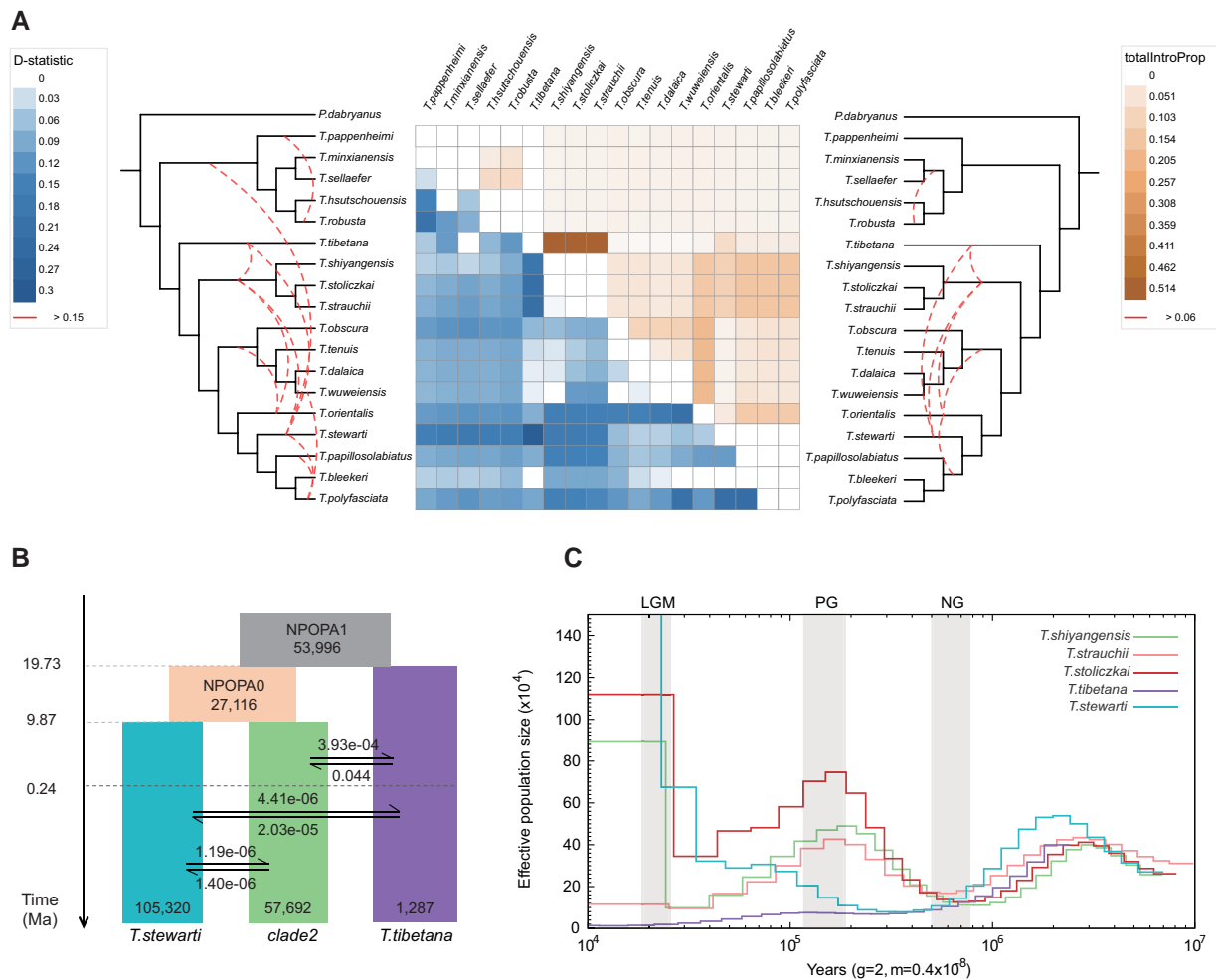


Fig. 2. Introgression between *Triplophysa* species. (A) Heat map of average pairwise D values per species pair (in left) and the average totalIntroProp per species pair inferred through the QuIBL analysis (in right). Empty squares represent zero or nonsignificant values. The species tree is shown on both sides. Red dotted lines on the species tree of two sides indicate interspecific introgression events based on summarized results with threshold of $D > 0.15$ and totalIntroProp > 0.06 , respectively. (B) Schematic of demographic history of the three *Triplophysa* species simulated by fastsimcoal2. Rectangles represent each species population and their ancestral lineages. The numbers on the bottom of rectangle represent N_e , and arrows indicate the per-generation migration rate (m) between species populations. The time node of each historical scenario is marked on the left timeline. (C) Demographic history inferred by the PSMC model. Lines represent the median estimated N_e of five *Triplophysa* species. The periods of the LGM, PG, and NG are highlighted with gray vertical bars.

0.06. Most high value results were consistent with the D -statistic analyses. Introgression signatures between *T. tibetana*–clade2, *T. tibetana*–clade4, *T. orientalis*–clade2, *T. orientalis*–*T. tibetana*, and *T. stewarti*–clade2 emerged. Hence, across the two analytical results, we found evidence for extensive introgression among several *Triplophysa* species, suggesting phylogenetic uncertainties caused by past hybridization between lineages rather than ILS.

Demographic History Infers Ancient Introgression

We performed coalescent-based demographic simulations using fastsimcoal2.7 (Excoffier et al. 2013), to complement details of introgression between *Triplophysa* species. We found that model7 (fig 2B) presented higher likelihood values and lowest Akaike information criterion (AIC) weights than other models (supplementary table S5, Supplementary Material online). The simulation results

estimated divergence events TDIV1, TDIV0 at 9,863,719 generations (equating to 19.73 Ma) and 4,934,948 generations (9.87 Ma) ago, respectively, and also suggested an ancient episode of gene flow between *T. tibetana* and clade2 at 122,301–4,934,948 generations (0.24–9.87 Ma). This species pair had a higher estimated migration rate (0.045/3.92e-04) than other species pairs, which agrees with the previous D -statistic and QuIBL analyses. Gene flow between the other two pairs of species *T. tibetana*–*T. stewarti* and *T. stewarti*–clade2 occurred after 122,301 generations (0.24 Ma). The assumptions that introgression across those three species pairs occurred after *T. stewarti* diverged and continues to the present (supplementary fig. S8b, Supplementary Material online), and all introgression events occurred in ancient ages (supplementary fig. S8h, Supplementary Material online) were not supported.

We used a pairwise sequential Markovian coalescent (PSMC) model to infer the demographic histories of the

five *Triplophysa* species that were analyzed using fastsimcoal (fig. 2C). All five species experienced a drastic rise in effective population size (N_e) ~ 10 –2.5 Ma (Middle Miocene to Pleistocene), which may reflect ancient range expansion or population growth. N_e decreases from 2.5 to 0.8 Ma to a minimum coinciding with the Naynayxungla glaciation (NG, 0.5–0.78 Ma), which is suggestive of a glaciation-driven bottleneck (Brookfield 1998; Clark et al. 2004; Zhang and Sun 2011). During the Penultimate glaciation (PG, 135–194 Ka), N_e of all three species from clade2 declined from a maximum, whereas the N_e of *T. stewarti* showed a population expansion from a minimum. *T. tibetana* showed a consistently low N_e . After this period, the N_e of all lineages declined until the Last Glacial Maximum (LGM, 26.5–19.0 Ka) except *T. stewarti*. Combined with fastsimcoal results, the increase of N_e for both clade2 species (0.2–1.0 Ma) and *T. stewartia* (26.5–250 Ka) after the NG correspond to their estimated timing of gene flow with *T. tibetana*.

Window-Based Phylogenomic Analysis of Introgressed Regions

We further analyzed the strong admixture signals between *T. tibetana* and clade2 indicated from population structure analysis, the introgression analysis, and the evidence of mtDNA introgression from mitonuclear discordance in *Triplophysa* phylogeny (Wang et al. 2016). We used the 25k window-based topology weighting analysis between *T. tibetana* and clade2 to define the topology type of each window, with clade3 as the third species and *P. dabryanus* as outgroup. We calculated the proportions of support for topo1, topo2, and topo3, at 45.7%, 1.0%, and 53.3% in all 25k window local trees, respectively (fig. 3A). The distribution of windows supports topo1 and topo3 across whole genomic regions (fig. 3B). These two tree topologies were supported by 32.3% and 37.7% of the genome, respectively, with a distribution characteristic of topo1 concentrating to the middle of the chromosome and topo3 to both ends in most chromosomes.

To confirm the relationship between introgressed loci and these two topological regions, we used a f_d statistic to calculate the fraction of introgression from *T. tibetana* to clade2 in 25k nonoverlap sliding windows (Martin et al. 2015). The results showed that the topo1 regions had significantly higher f_d values, indicating that *T. tibetana* introgression into clade2 occurred predominantly in these regions, consistent with the hypothesis that *T. tibetana* is sister to clades2 indicative of introgression between them. To evaluate whether there is an association between natural selection and the distribution of different topological regions, we examined a number of population genetic factors that would be impacted by natural selection. The 25k window-based population parameters, including F_{st} , absolute genetic divergence (Dxy), and nucleotide diversity (π), between *T. tibetana* and clade2, were significantly higher in windows supporting topo3 than those in supporting topo1, and the population recombination rate

also presented a consistent trend of significantly lower value in topo1 (Wilcoxon rank sum test; fig. 3C). In general, introgression in topo1 is mainly concentrated in genomic regions with lower Dxy, F_{st} , nucleotide diversity, and recombination.

Discussion

In this study, we used data sets from genome-wide variants to reconstruct phylogenetic trees and histories of discordance in high-altitude loaches. Despite the differences across different types of sequence data sets and analyses, most of the phylogenetic relationships inferred are generally consistent. There are four major clades comprising the 18 *Triplophysa* species included in our study, and we uncovered evidence for interspecies gene flow that occurred both in ancient as well as recent times and is consistently supported as the source of phylogenetic inconsistency.

Different from previous studies (Wang et al. 2016; Wang, Zhang et al. 2020), our phylogenetic results indicate genetic exchange among *Triplophysa* species in geographically adjacent regions that probably resulted from hybridization among microallopatric lineages (Toews and Brelsford 2012; Wang et al. 2018). Mitonuclear discordance can be an evidence of historical interspecies hybridization and mitochondrial capture between ancestors of the *T. tibetana* and *T. stewarti* lineages, consistent with fastsimcoal analyses, which were detected. However, the mitochondrial phylogeny did not reveal hybridization events between *T. tibetana* and clade2, possibly due to alternative factors such as local adaptation of mtDNA genes (Pavlova et al. 2013), asymmetric hybridization (Chan and Levin 2005), or mitochondrial introgression or fusion following secondary contact (Phuong et al. 2017; Garrick et al. 2019).

Generally, individual gene trees may differ from the species phylogeny due to gene duplication, ILS, or introgressive hybridization (Avise and Robinson 2008; Degnan and Rosenberg 2009). Phylogenetic analyses based on concatenated data do not take into account how the stochasticity of the coalescent process affects the accuracy of an obtained species tree and this bias does not diminish as the amount of data increases (Kubatko and Degnan 2007). We found another factor that contributes to inconsistency in phylogenetic construction: The source of the variation in sites used to infer phylogenetic relationships, if they are due to hybridization. In this study, the topology with the highest proportion in topology weighting analysis was different when infer from local window tree sets and gene tree sets (supplementary fig. S6, Supplementary Material online), so did the topology of ASTRAL species tree constructed from those two sets.

When using the gene trees, CDS region and 4D sites, the inferred phylogenetic relationships of *T. tibetana* and four major clades are always consistent (supplementary fig. S5g–i, Supplementary Material online). This kind of pattern was first found in a study of the radiation of *Lonchura munia* birds, in which trees inferred from

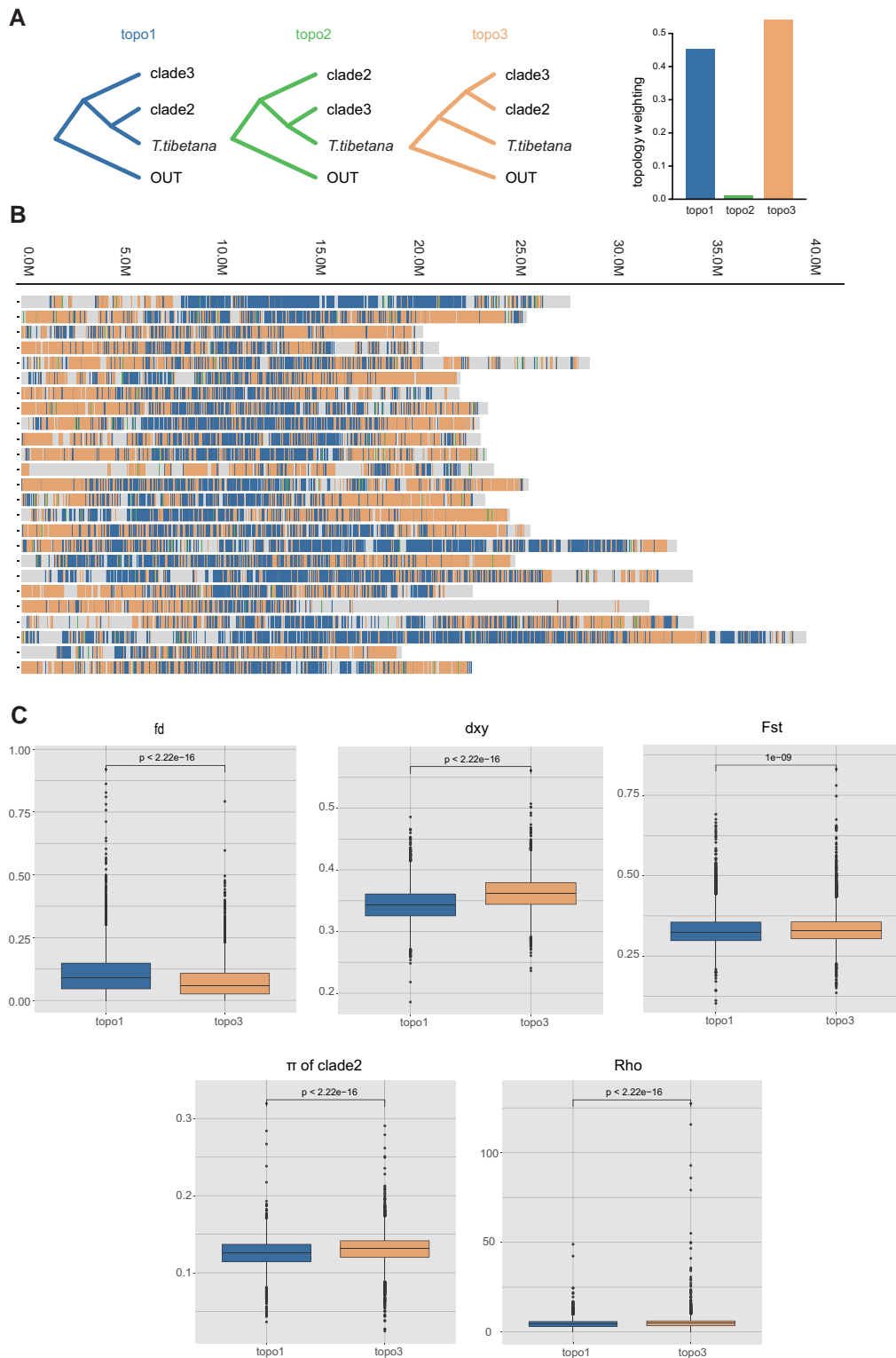


Fig. 3. Window-based population genomic analysis. (A) Three possible phylogenetic topologies inferred from 25-kb window local ML trees for *T. tibetana* clade2 and clade3, with *P. dabryanus* as outgroup species. The colored vertical bars represent the windows with complete monophyly of the three alternative topologies, where the proportion of the three topologies is the average value of three clade2 species. (B) Distribution of tree topologies across the genome. Colors are consistent with the topology in A. (C) Population genomic parameters of two different topological regions between *T. tibetana* and clade2.

autosomal regions thought to relate to color evolution differed significantly from tree inferences based on genome-wide SNPs (Stryjewski and Sorenson 2017). These results

are consistent with the hypothesis that sampling smaller sets of genomic loci may help avoid the effects of gene flow due to stochastic sampling during phylogenetic

inference, in the premise of most genomic loci misrepresent the real phylogeny (Alstrom et al. 2018; Zhang et al. 2021).

The success of introgressed adaptive traits depends on the rates of hybridization, introgression, recombination, and the strength of selection (Lenormand 2002; Ghalambor et al. 2003). A high rate of hybridization with little opportunity for recombination and/or selection may not necessarily lead to the evolution of adaptive traits. In contrast, a low rate of hybridization with an opportunity for recombination and selection will more likely lead to the incorporation of adaptive traits and simultaneous purging of maladaptive traits within progressively shorter time intervals (Feulner et al. 2013; Arnold and Kunte 2017). Our results suggested that signatures of introgression were preserved between *T. tibetana* and clade2 due to lower rates of recombination and the lower Dxy, Fst, and nucleotide diversity may be a signal of selection across this pair.

One potential explanation for the strong signals of introgression between *T. tibetana* and clade2 found in this study is that ancient admixture of ancestral variation occurred between *T. tibetana* and ancestral lineage of clade2. Previous studies have shown that genes associated with energy metabolism and hypoxic responses have undergone positive selection and rapid radiation in *Triplophysa* fishes (Wang et al. 2015). Positive selection frees the hybridized alleles from the fate of being lost due to genetic drift (Zhang et al. 2021). However, intrinsic factors alone are insufficient to explain bursts of diversification and adaptive radiation (Pease et al. 2016). Our reconstruction of the demographic history of *Triplophysa* suggests that this clade underwent an explosive speciation event corresponding to the second episode of Tibetan Plateau uplift (supplementary fig. S7, Supplementary Material online). At the late stage of the second uplift from 10 to 4 Ma, climate change caused by the monsoon cycle led to rapid major global cooling (Shi and Li 1999). Concurrently, the Ne of *Triplophysa* expanded stably during this period (fig. 2C). Due to the lack of strict reproductive isolation between early differentiated species, gene flow between ancestral lineages may also have occurred during this time (Marques et al. 2019). During the third episode of Tibetan Plateau uplift from 4 Ma to the present, the Ne of *Triplophysa* gradually decreased and presented distinct population histories in the subsequent three glaciation events.

It is thought that cold-adapted species undergo range expansions southward and into lowlands during glaciations (Sanchez-Montes et al. 2019; Dufresnes et al. 2020). After the NG ice age retreat, all species except *T. tibetana* tended to return to higher effective population size. This may be related to the extensive gene flow between *Triplophysa*. In our results, both inferred gene flows temporally coincide with an increase in the effective population size of *T. stewarti* and clade2. Thus, episodes of introgression that affect adaptive loci can promote species persistence in regions experiencing high climatic turnover through time.

However, adaptive introgression is only one potential consequence of gene flow; admixture may also result in maladaptively admixed individuals. Deleterious alleles are expected to be removed from a population by random genetic drift alone in a population of small Ne, and gene flow that results in masking of deleterious alleles is the key to sustain the growth rate of populations (Whiteley et al. 2015; Hoffmann et al. 2021; Khan et al. 2021). Stratigraphic studies show that the Tibetan Plateau has undergone three uplifting events since the late Cenozoic and that these have had a profound impact on the East Asian environment and by implication on the biodiversity of the region (Shi and Li 1999; Zhisheng et al. 2001; Li et al. 2015). A violent and uneven tectonic uplift called Gonghe Movement that occurred about 150 thousand years ago divided the river into independent closed lakes or cut off the connection between the water systems (Shi and Li 1999; Yu et al. 2022). This change may have also geographically isolated some *Triplophysa* species in the region, and it is very likely that the *T. tibetana* formed a small, isolated population during this period, as suggested by both PSMC and fastsimcoal simulations (fig. 2B and C). K-mer-based analysis showed that the *T. tibetana* has a low levels of heterozygosity (Yang, Liu, et al. 2019), and as other species that have gone through recent population size declines, *T. tibetana* also shows low genetic diversity (supplementary fig. S1, Supplementary Material online). This might be due to a possible founder effect for the *T. tibetana* population (Xue et al. 2015; Westbury et al. 2018; Liu et al. 2021).

In general, our study confirms that introgression poses certain challenges in clarifying the evolutionary relationships between *T. tibetana* species. Our work confirms the link between this introgression and selection, which is important for exploring the molecular genetic mechanisms behind the adaptive evolution of *T. tibetana*. At the same time, it is of great significance for genetic communication between species, maintenance of species diversity, and adaptive evolution. Our investigation of *T. tibetana* population history is useful for understanding changes in population dynamics. At a time when the habitat of the *Triplophysa* is increasingly fragmented, these endangered populations are in urgent need of conservation.

Materials and Methods

Sample Collection, Whole-Genome Resequencing and Genotype Calling

A total of 138 samples were collected for this study, representing 18 species of the *Triplophysa* genus. An individual of *P. dabryanus* was used as an outgroup (supplementary table S1, Supplementary Material online). Total genomic DNA was extracted from the muscle tissue, and 150-bp PE libraries were sequenced on an Illumina HiSeq X Ten platform. Before reads mapping, we used fastp (version 0.20.0) to remove adapter sequences and to trim low-quality bases with default parameters. Sequence reads

were aligned to the *T. tibetana* reference genome (https://www.ncbi.nlm.nih.gov/assembly/GCA_008369825.1/) (Yang, Liu, et al. 2019) using bwa-mem (version 0.7.15) with default parameters. SNVs and short-indel variants were called using the GATK (version 4.1.0.0) with the filter expression “QUAL < 30.0 || QD < 2.0 || FS > 60.0 || MQ < 40.0 || SOR > 4.0” to discard site types probably caused by mapping bias and following cutoffs: minimum depth > 20, minor allele frequency > 0.05, missing rate < 0.5, and only keep biallelic sites. Finally, a total of 11,683,331 sites that passed all these filtering criteria were used in further phylogenetic analyses.

Phylogenetic Relationships Analysis

To infer the phylogenetic relationship of the 18 species, we used concatenation approach and multispecies coalescent approach (Degnan and Rosenberg 2009). We inferred a ML tree based on three data sets by concatenating all filtered SNPs across the WGS, protein CDS, and 4D sites, respectively. A total of 461,486 4D sites were extracted. For the multispecies coalescent approach, we first split the whole genome into non-overlapping windows of 200, 100, 50, 25, and 10 kb. We reconstructed the ML trees for each window with <100 sites filtered out. We extracted SNPs from gene regions and kept one sample with the highest coverage in each species. From 24,138 annotated genes in a *T. tibetana* reference genome, a total of 3,600 genes were used to reconstruct gene trees in IQTree with the same manner as for our sliding window local tree analysis and a filter of >500 SNP loci. ASTRAL species trees were then inferred from the six genomic windows/gene tree data sets (Mirarab et al. 2014). All ML trees were inferred using IQTree with a GTR + ASC substitution model and 1,000 bootstrap replicates. We plotted a subset of 1,000 randomly selected 200-kb window ML trees in the program DensiTree (Bouckaert et al. 2019). The topology weighting analysis between each phylogenetic clade and *T. tibetana* species, which shows different sizes of sliding windows supporting all possible topologies, was conducted from sliding window-based local ML trees using Twisst (Martin and Van Belleghem 2017).

Population Genetic Structure Analysis

We used a linkage disequilibrium filter (Zhou et al. 2020) implemented in PLINK (v1.90b6.13 64-bit) (Purcell et al. 2007) with the parameter “indep-pairwise 100 kb 1 0.8” to generate a data set of 1,746,277 SNPs for the population genetic analysis. PCA was performed with the Genome-wide Complex Trait Analysis (GCTA, version: 1.25.3) (Yang et al. 2011). We used the software ADMIXTURE (Alexander et al. 2009) with a random seed to infer population structure of 18 species. The number of genetic clusters (K) was set from 2 to 20. We estimated nucleotide diversity (π) among 18 species of *Triplophysa* using VCFtools (0.1.15).

Analyses of Introgression

The outgroup f_3 -statistic and D -statistics were calculated using the qp3pop and qpDstat programs in ADMIXTOOLS (Patterson et al. 2012) to detect gene flow between all *Triplophysa* fishes. Outgroup f_3 statistics were performed in the form of (PDAB; A and B), where A and B represent a pair of species of *Triplophysa* and PDAB represents outgroup *P. dabryanus*. The expected value was proportional to the shared genetic history between A and B: The larger the f_3 value, the greater the genetic relatedness between the two populations. For the D -statistics, we used the quartet topology form (W, X, Y, and OUT), where W, X, and Y represent populations from clade1, clade2, clade3, clade4, and the TTIB population in a species tree. All 816 detected quartets conformed to the species tree. We considered a Z-score of >3 and a positive value of D -value indicative that a W population shared genetic admixture with a Y population in a test form and identified a species pair W_Y. A D -value less than -3 indicated that an X population shared genetic admixture with a Y population in a test form and identified as a species pair X_Y. All D -statistics were classified as 136 species pairs.

We then employed QuIBL (Edelman et al. 2019), a method based on internal branch length distribution, to quantify the proportions of ILS and introgression contributions to topology incongruence. As a tree-based method, the results of QuIBL are trusted to the assumption that input trees are inferred from loci with limited internal recombination. We therefore chose the gene tree set for analysis. We used the species tree topology to assign the outgroup to each triplet in QuIBL. For each triplet form W_X_Y with outgroup Y, we also identified a species pair as W_X. After filtering with the significance determination value $|d\text{BIC}| > 10$, we found 132 discordant species pairs and calculated the average of the total IntroProp for each species pair.

Demographic History Reconstruction

We used PSMC (Li and Durbin 2011) to estimate effective population size changes through time based on whole-genome variants of six individuals: TTIB.YLD0088, TSTE.YLD0087, TSHI.C10, TSTR.D7, and TSTO.A10, representing *T. tibetana*, *T. stewarti*, *T. shiyangensis*, *T. strauchii*, *T. stoliczkai*, respectively. The consensus sequences were generated using vcfutils.pl (vcf2fq -d 10 -D 80). The fq2psmcf tool was used to create the input file for PSMC modeling. Sequences were used as the input for the PSMC estimates with the options -N25 -t15 -r5.

Given the high gene introgression levels exhibited in D -statistic analyses between *T. tibetana* and clade2, as well as exhibited between *T. tibetana* and *T. stewarti*, we used a coalescent simulation-based method employed in fastsimcoal2.7 (Excoffier et al. 2013) to infer the demographic histories for those three population. We selected two samples from each species in clade2 (*T. shiyangensis*, *T. strauchii*, and *T. stoliczkai*) and joined them together

as a whole population named clade2. Populations of *T. tibetana* and *T. stewarti* contained three and six individuals, respectively. We used the two-dimensional joint SFS (2D site frequency spectrum) constructed from sample allele frequencies using *angsd* (v0.934) (Nielsen et al. 2012). To minimize the bias in demographic inferences due to selection, only neutral sites (4D) were used for this analysis. A total of 325,487 segregating sites were filtered from all 473,182 4D sites. The setting of the model is mainly based on the assumption that introgression and migration events between occurred three *Triplophysa* groups. The eight models differed in 1) whether gene flow was present between all three species; 2) whether before-divergence gene flow between *T. tibetana* and clade2 was present; and 3) the time, pattern, and migration rate size of gene flow between the three *Triplophysa* species. Three demographic models (supplementary fig. S8a–c, Supplementary Material online) assumed no gene flow, gene flow between three populations after TDIV0, or gene flow between A0 and *T. tibetana*. In another five demographic models, we add TS as time split point, where one or more gene flows between three populations occurred before or after TS (supplementary fig. S8d–h, Supplementary Material online). Several parameters were used to define evolutionary scenarios. TDIV1 was used as the divergence time between *T. tibetana* and ancestral lineage of *T. stewarti*/clade2 (named A0), and TDIV0 was used as the divergence time between *T. stewarti* and clade2. The ML estimates for all demographic parameters under each model were obtained from 50 independent runs, with 100,000 simulations applied for parameter estimation (-n 100,000) and 40 loops (-L 40) to perform during likelihood maximization. Model comparison was based on the maximum value of the likelihood using the AIC and Akaike weight. The model with the maximum Akaike weight value was chosen as the optimal one. A generation time of 2 years (Zhou et al. 2021) and a mutation rate of 4.13×10^{-9} (Fu et al. 2010) per million years were applied in simulations.

Divergence Time Estimation

We estimated the lineage divergence times based on 25-kb window data set. Only SNVs in the windows concordant with the consensus phylogeny (5.13 million SNVs) were used (Yang et al. 2021). We then used MCMCtree from the package PAML (Yang 2007) to estimate divergence times with the independent rate model (clock = 2). The MCMC run was executed with 2,000 generations as burn-in and sampled every 10 generations until a total of 20,000 samples were collected.

Window-Based Genealogy Difference Analysis and Population Genomic Analysis

We used *Twisst* to assess and quantify the phylogenetic discordance among *T. tibetana* (TTIB), clade3, and clade2 along the genome. The genealogical relationships of these

species can be defined by three possible topologies: ((TTIB, clade2), clade3); ((TTIB, clade3), clade2); ((clade2, clade3), TTIB). *Paramisgurnus dabryanus* was used as an outgroup species, and local phylogenetic subtrees were inferred from 25-kb window local ML trees. The genealogy type with the highest percentage in topology weightings was determined as the type of the window. We estimated window-based population parameters including *F_{st}*, absolute genetic divergence (*D_{xy}*), nucleotide diversity (π) using the *genomics_general* package with 25-kb nonoverlapping sliding windows (https://github.com/simonhmartin/genomics_general). The package is also used to calculate *f_d* and estimate the proportion of introgression in the same window (Martin et al. 2015). Under a given four-taxon topology ((clade3, clade2), TTIB, OUT), a positive *f_d* statistic value indicates the introgression proportion from population TTIB to clade2. The minimum sites used in each window were set to 3 through -m flag. For windows of *D* < 0, or of *D* > 0, but *f_d* > 1, the *f_d* statistic value becomes meaningless, and we filtered it out. We estimated the population recombination rate (*Rho*) for species of clade2 in *FastEPRR* 2.0 (Gao et al. 2016) in the same window size.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (31972866), the Strategic Priority Research Program of Chinese Academy of Sciences (Grant No. XDB31000000), research grant for talent introduction from the Southwest University (20230207), the National Talent Research Grant for 2023 (No. 5330500953), the National Natural Science Foundation of China (32170480, 31601858, 32072980), the Third Xinjiang Scientific Expedition Program (Grant No. 2021xjkk0500), Chinese Academy of Sciences (Youth Innovation Promotion Association), Chinese Academy of Sciences (<http://www.yicas.cn>), the Open Project of State Key Laboratory of Plateau Ecology and Agriculture, Qinghai University (2021-KF-06), and the Young Top-notch Talent Cultivation Program of Hubei Province. This research was supported by the Wuhan Branch, Supercomputing Center, Chinese Academy of Sciences, China. We are grateful to Chase Doran Brownstein and Thomas J. Near from Yale University for help with English editing.

Author Contributions

S.H. led the research. L.Y. designed the research. Y.Q., M.M., and L.Y. analyzed the data. C.Z., H.L., H.J., Y.X., W.C., Z.D., Y.L., X.G., C.W., Y.L. T.W., Y.W., X.G., and L.Y. collected the samples. Y.Q., A.M., and L.Y. wrote the paper.

Data Availability

Illumina whole genome resequencing reads for 18 species of *Triplophysa* are available at Genome Sequence Archive (GSA <https://ngdc.cnbc.ac.cn/gsa/>) under BioProject PRJCA015528. Results of D-statistic and QuIBL analyses are available as [Supplementary Material](#) online. All templates used in the fastsimcoal analysis are available on GitHub at https://github.com/rainstop2019/Project_2022.

Conflict of interest statement. The authors declare no competing interests.

References

- Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**:1655–1664.
- Alstrom P, Rheindt FE, Zhang R, Zhao M, Wang J, Zhu X, Gwee CY, Hao Y, Ohlson J, Jia C, et al. 2018. Complete species-level phylogeny of the leaf warbler (Aves: Phylloscopidae) radiation. *Mol Phylogenet Evol.* **126**:141–152.
- Arnold ML, Kunte K. 2017. Adaptive genetic exchange: a tangled history of admixture and evolutionary innovation. *Trends Ecol Evol (Amst)*. **32**:601–611.
- Avise JC, Robinson TJ. 2008. Hemiplasy: a new term in the lexicon of phylogenetics. *Syst Biol.* **57**:503–507.
- Berner D, Salzburger W. 2015. The genomics of organismal diversification illuminated by adaptive radiations. *Trends Genet.* **31**:491–499.
- Blom MPK, Bragg JG, Potter S, Moritz C. 2017. Accounting for uncertainty in gene tree estimation: summary-coalescent species tree inference in a challenging radiation of Australian lizards. *Syst Biol.* **66**:352–366.
- Bouckaert R, Vaughan TG, Barido-Sottani J, Duchene S, Fourment M, Gavryushkina A, Heled J, Jones G, Kuehnert D, De Maio N, et al. 2019. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLoS Comput Biol.* **15**:e1006650.
- Bouckaert RR. 2010. DensiTree: making sense of sets of phylogenetic trees. *Bioinformatics* **26**:1372–1373.
- Brookfield M. 1998. The evolution of the great river systems of southern Asia during the Cenozoic India-Asia collision: rivers draining southwards. *Geomorphology.* **22**:285–312.
- Bureau of fisheries of Tibet Autonomous Region. 1995. *The fishes and fish resources in Xizang, China*. Beijing: China Agriculture Press.
- Cahill JA, Green RE, Fulton TL, Stiller M, Jay F, Ovsvyanikov N, Salamzade R, St John J, Stirling I, Slatkin M, et al. 2013. Genomic evidence for island population conversion resolves conflicting theories of polar bear evolution. *PLoS Genet.* **9**:e1003345.
- Chan KMA, Levin SA. 2005. Leaky prezygotic isolation and porous genomes: rapid introgression of maternally inherited DNA. *Evolution* **59**:720–729.
- Choi JY, Dai X, Alam O, Peng JZ, Rughani P, Hickey S, Harrington E, Juul S, Ayroles JF, Purugganan MD, et al. 2021. Ancestral polymorphisms shape the adaptive radiation of *Metrosideros* across the Hawaiian Islands. *Proc Natl Acad Sci U S A.* **118**:e2023801118.
- Clark M, Schoenbohm L, Royden L, Whipple K, Burchfiel B, Zhang X, Tang W, Wang E, Chen L. 2004. Surface uplift, tectonics, and erosion of eastern Tibet from large-scale drainage patterns. *Tectonics* **23**:n/a–n/a.
- Currat M, Ruedi M, Petit RJ, Excoffier L. 2008. The hidden side of invasions: massive introgression by local genes. *Evolution* **62**:1908–1920.
- Degnan JH, Rosenberg NA. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol Evol (Amst)*. **24**:332–340.
- Deng T, Ding L. 2015. Paleoaltimetry reconstructions of the Tibetan Plateau: progress and contradictions. *Natl Sci Rev.* **2**:417–437.
- Dufresnes C, Nicieza AG, Litvinchuk SN, Rodrigues N, Jeffries DL, Vences M, Perrin N, Martinez-Solano I. 2020. Are glacial refugia hotspots of speciation and cytonuclear discordances? Answers from the genomic phylogeography of Spanish common frogs. *Mol Ecol.* **29**:986–1000.
- Dupuis JR, Sperling FAH. 2020. Phylogenomic test of mitochondrial clues to archaic ancestors in a group of hybridizing swallowtail butterflies. *Mol Phylogenet Evol.* **152**:106921.
- Edelman NB, Frandsen PB, Miyagi M, Clavijo B, Davey J, Dikow RB, Garcia-Accinelli G, Van Belleghem SM, Patterson N, Neafsey DE, et al. 2019. Genomic architecture and introgression shape a butterfly radiation. *Science* **366**:594–599.
- Excoffier L, Dupanloup I, Huerta-Sanchez E, Sousa VC, Foll M. 2013. Robust demographic inference from genomic and SNP data. *PLoS Genet.* **9**:e1003905.
- Fang XM, Zhao ZJ, Li JJ, Yan MD, Pan BT, Song CH, Dai SA. 2005. Magnetostratigraphy of the late Cenozoic Laojunmiao anticline in the northern Qilian Mountains and its implications for the northern Tibetan Plateau uplift. *Science in China Series D-Earth Sciences.* **48**:1040–1051.
- Favre A, Packert M, Pauls SU, Jahnig SC, Uhl D, Michalak I, Muellner-Riehl AN. 2015. The role of the uplift of the Qinghai–Tibetan Plateau for the evolution of Tibetan biotas. *Biological Reviews.* **90**:236–253.
- Feng C, Zhou W, Tang Y, Gao Y, Chen J, Tong C, Liu S, Wanghe K, Zhao K. 2019. Molecular systematics of the *Triplophysa robusta* (Cobitoidea) complex: extensive gene flow in a depauperate lineage. *Mol Phylogenet Evol.* **132**:275–283.
- Feulner PGD, Gratten J, Kijas JW, Visscher PM, Pemberton JM, Slate J. 2013. Introgression and the fate of domesticated genes in a wild mammal population. *Mol Ecol.* **22**:4210–4221.
- Fontaine MC, Pease JB, Steele A, Waterhouse RM, Neafsey DE, Sharakhov IV, Jiang X, Hall AB, Catteruccia F, Kakani E, et al. 2015. Mosquito genomics. Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science* **347**:1258524.
- Fu B, Chen M, Zou M, Long M, He S. 2010. The rapid generation of chemical genes expanding protein diversity in zebrafish. *BMC Genomics* **11**:657.
- Gao F, Ming C, Hu W, Li H. 2016. New software for the fast estimation of population recombination rates (FastEPRR) in the genomic era. *G3 Genes|Genomes|Genetics* **6**:1563–1571.
- Garrick RC, Banusiewicz JD, Burgess S, Hyseni C, Symula RE. 2019. Extending phylogeography to account for lineage fusion. *J Biogeogr.* **46**:268–278.
- Ghalambor CK, Walker JA, Reznick DN. 2003. Multi-trait selection, adaptation, and constraints on the evolution of burst swimming performance. *Integr Comp Biol.* **43**:431–438.
- He C, Song Z, Zhang E. 2011. *Triplophysa* fishes in China and the status of its taxonomic studies. *Sichuan Journal of Zoology* **30**:150–155.
- He DK, Chen YX, Chen YF. 2006. Molecular phylogeny and biogeography of the genus *Triplophysa* (Osteichthyes:Nemacheilinae) in the Tibetan Plateau inferred from cytochrome b DNA sequences. *Progress in Natural Science* **16**:1395–1404.
- Hedrick PW. 2013. Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. *Mol Ecol.* **22**:4606–4618.
- Hinojosa JC, Dapporto L, Pitteloud C, Koubínová D, Hernández-Roldán J, Vicente JC, Alvarez N, Vila R. 2022. Hybridization fuelled diversification in *Spialia* butterflies. *Mol Ecol.* **31**:2951–2967.
- Hoffmann AA, Miller AD, Weeks AR. 2021. Genetic mixing for population management: from genetic rescue to provenancing. *Evol Appl.* **14**:634–652.
- Hu X, Li H, Lin Y, Wang Z, Feng H, Zhou M, Shi L, Cao H, Ren Y. 2022. Genomic deciphering of sex determination and unique immune

- system of a potential model species rare minnow (*Gobiocypris rarus*). *Sci Adv*. **8**:eabl7253.
- Irisarri I, Baurain D, Brinkmann H, Delsuc F, Sire J-Y, Kupfer A, Petersen J, Jarek M, Meyer A, Vences M, et al. 2017. Phylotranscriptomic consolidation of the jawed vertebrate time-tree. *Nat Ecol Evol*. **1**:1370–1378.
- Irisarri I, Meyer A. 2016. The identification of the closest living relative(s) of tetrapods: phylogenomic lessons for resolving short ancient internodes. *Syst Biol*. **65**:1057–1075.
- Irisarri I, Singh P, Koblmüller S, Torres-Dowdall J, Henning F, Franchini P, Fischer C, Lemmon AR, Lemmon EM, Thallinger GG, et al. 2018. Phylogenomics uncovers early hybridization and adaptive loci shaping the radiation of Lake Tanganyika cichlid fishes. *Nat Commun*. **9**:3159.
- Jamie GA, Meier JI. 2020. The persistence of polymorphisms across species radiations. *Trends Ecol Evol (Amst)*. **35**:795–808.
- Kang JH, Scharl M, Walter RB, Meyer A. 2013. Comprehensive phylogenetic analysis of all species of swordtails and platies (Pisces: Genus *Xiphophorus*) uncovers a hybrid origin of a swordtail fish, *Xiphophorus monticolus*, and demonstrates that the sexually selected sword originated in the ancestral lineage of the genus, but was lost again secondarily. *BMC Evol Biol*. **13**:1–19.
- Khan A, Patel K, Shukla H, Viswanathan A, van der Valk T, Borthakur U, Nigam P, Zachariah A, Jhala YV, Kardos M, et al. 2021. Genomic evidence for inbreeding depression and purging of deleterious genetic variation in Indian tigers. *Proc Natl Acad Sci U S A*. **118**:e2023018118.
- Kubatko LS, Degnan JH. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Syst Biol*. **56**:17–24.
- Kulmuni J, Butlin RK, Lucek K, Savolainen V, Westram AM. 2020. Towards the completion of speciation: the evolution of reproductive isolation beyond the first barriers. *Philos Trans R Soc B Biol Sci*. **375**:20190528.
- Leebens-Mack JH, Barker MS, Carpenter EJ, Deyholos MK, Gitzendanner MA, Graham SW, Grosse I, Li Z, Melkonian M, Mirarab S, et al. 2019. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature*. **574**:679–685.
- Lenormand T. 2002. Gene flow and the limits to natural selection. *Trends Ecol Evol (Amst)*. **17**:183–189.
- Li G, Davis BW, Eizirik E, Murphy WJ. 2016. Phylogenomic evidence for ancient hybridization in the genomes of living cats (Felidae). *Genome Res*. **26**:1–11.
- Li H, Durbin R. 2011. Inference of human population history from individual whole-genome sequences. *Nature*. **475**:493–496.
- Li J, Wang Y, Jin H, Li W, Yan C, Yan P, Zhang X, He S, Song Z. 2017. Identification of *Triplophysa* species from the Qinghai–Tibetan Plateau (QTP) and its adjacent regions through DNA barcodes. *Gene*. **605**:12–19.
- Li J, Zhou S, Zhao Z, Zhang J. 2015. The Qingzang Movement: the major uplift of the Qinghai–Tibetan Plateau. *Sci China Earth Sci*. **58**:2113–2122.
- Liang Z, Duan S, Sheng J, Zhu S, Ni X, Shao J, Liu C, Nick P, Du F, Fan P, et al. 2019. Whole-genome resequencing of 472 *Vitis* accessions for grapevine diversity and demographic history analyses. *Nat Commun*. **10**(1):1190.
- Liu S, Westbury MV, Dussex N, Mitchell KJ, Sinding M-HS, Heintzman PD, Duchêne DA, Kapp JD, von Seth J, Heiniger H, et al. 2021. Ancient and modern genomes unravel the evolutionary history of the rhinoceros family. *Cell*. **184**:4874–4885.e16.
- Liu S, Zhang L, Sang Y, Lai Q, Zhang X, Jia C, Long Z, Wu J, Ma T, Mao K, et al. 2022. Demographic history and natural selection shape patterns of deleterious mutation load and barriers to introgression across *Populus* genome. *Mol Biol Evol*. **39**:msac008.
- Mallet J, Besansky N, Hahn MW. 2016. How reticulated are species? *Bioessays*. **38**:140–149.
- Marques DA, Meier JI, Seehausen O. 2019. A combinatorial view on speciation and adaptive radiation. *Trends Ecol Evol (Amst)*. **34**:531–544.
- Martin SH, Davey JW, Jiggins CD. 2015. Evaluating the use of ABBA–BABA statistics to locate introgressed loci. *Mol Biol Evol*. **32**:244–257.
- Martin SH, Davey JW, Salazar C, Jiggins CD. 2019. Recombination rate variation shapes barriers to introgression across butterfly genomes. *PLoS Biol*. **17**:e2006288.
- Martin SH, Van Belleghem SM. 2017. Exploring evolutionary relationships across the genome using topology weighting. *Genetics*. **206**:429–438.
- Meier JI, Marques DA, Mwaiko S, Wagner CE, Excoffier L, Seehausen O. 2017. Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nat Commun*. **8**:14363.
- Meyer A, Salzburger W, Scharl M. 2006. Hybrid origin of a swordtail species (Teleostei: *Xiphophorus clemenciae*) driven by sexual selection. *Mol Ecol*. **15**:721–730.
- Mirarab S, Reaz R, Bayzid MS, Zimmermann T, Swenson MS, Warnow T. 2014. ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics*. **30**:i541–i548.
- Nater A, Burri R, Kawakami T, Smeds L, Ellegren H. 2015. Resolving evolutionary relationships in closely related species with whole-genome sequencing data. *Syst Biol*. **64**:1000–1017.
- Nelson TC, Stathos AM, Vanderpool DD, Finseth FR, Yuan Y, Fishman L. 2021. Ancient and recent introgression shape the evolutionary history of pollinator adaptation and speciation in a model monkeyflower radiation (*Mimulus* section *Erythranthe*). *PLoS Genet*. **17**:e1009095.
- Nichols R. 2001. Gene trees and species trees are not the same. *Trends Ecol Evol*. **16**:358–364.
- Nielsen R, Korneliussen T, Albrechtsen A, Li YR, Wang J. 2012. SNP calling, genotype calling, and sample allele frequency estimation from new-generation sequencing data. *Plos One*. **7**:e37558.
- Nosil P, Funk DJ, Ortiz-Barrientos D. 2009. Divergent selection and heterogeneous genomic divergence. *Mol Ecol*. **18**:375–402.
- Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D. 2012. Ancient admixture in human history. *Genetics*. **192**:1065–1093.
- Pavlova A, Amos JN, Joseph L, Loynes K, Austin JJ, Keogh JS, Stone GN, Nicholls JA, Sunnucks P. 2013. Perched at the mito-nuclear crossroads: divergent mitochondrial lineages correlate with environment in the face of ongoing nuclear gene flow in an Australian bird. *Evolution*. **67**:3412–3428.
- Pease JB, Haak DC, Hahn MW, Moyle LC. 2016. Phylogenomics reveals three sources of adaptive variation during a rapid radiation. *PLoS Biol*. **14**:e1002379.
- Phuong MA, Bi K, Moritz C. 2017. Range instability leads to cytonuclear discordance in a morphologically cryptic ground squirrel species complex. *Mol Ecol*. **26**:4743–4755.
- Prum RO, Berv JS, Dornburg A, Field DJ, Townsend JP, Lemmon EM, Lemmon AR. 2015. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature*. **526**:569–573.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. **81**:559–575.
- Rannala B, Yang ZH. 2003. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics*. **164**:1645–1656.
- Ruber L, Meyer A, Sturmbauer C, Verheyen E. 2001. Population structure in two sympatric species of the Lake Tanganyika cichlid tribe Eretmodini: evidence for introgression. *Mol Ecol*. **10**:1207–1225.
- Sanchez-Montes G, Recuero E, Marcia Barbosa A, Martinez-Solano I. 2019. Complementing the Pleistocene biogeography of European amphibians: testimony from a southern Atlantic species. *J Biogeogr*. **46**:568–583.
- Schumer M, Xu C, Powell DL, Durvasula A, Skov L, Holland C, Blazier JC, Sankararaman S, Andolfatto P, Rosenthal GG, et al. 2018. Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science*. **360**:656–660.

- Shen X-X, Steenwyk JL, Rokas A. 2021. Dissecting incongruence between concatenation- and quartet-based approaches in phylogenomic data. *Syst Biol*. **70**:997–1014.
- Shi Y, Li J. 1999. Uplift of the Qinghai-Xizang (Tibetan) plateau and east Asia environmental change during late Cenozoic. *Acta Geographica Sinica* **01**:10–21.
- Simmons MP, Kessenich J. 2020. Divergence and support among slightly suboptimal likelihood gene trees. *Cladistics* **36**:322–340.
- Solis-Lemus C, Ane C. 2016. Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. *PLoS Genet*. **12**:e1005896.
- Stryjewski KF, Sorenson MD. 2017. Mosaic genome evolution in a recent and rapid avian radiation. *Nat Ecol Evol*. **1**:1912–1922.
- Su T, Spicer RA, Li S-H, Xu H, Huang J, Sherlock S, Huang Y-J, Li S-F, Wang L, Jia L-B, et al. 2019. Uplift, climate and biotic changes at the Eocene–Oligocene transition in south-eastern Tibet. *Natl Sci Rev*. **6**:495–504.
- Svardal H, Quah FX, Malinsky M, Ngatunga BP, Miska EA, Salzburger W, Genner MJ, Turner GF, Durbin R. 2020. Ancestral hybridization facilitated species diversification in the Lake Malawi cichlid fish adaptive radiation. *Mol Biol Evol*. **37**:1100–1113.
- Toews DPL, Brelsford A. 2012. The biogeography of mitochondrial and nuclear discordance in animals. *Mol Ecol*. **21**:3907–3930.
- Via S, West J. 2008. The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Mol Ecol*. **17**:4334–4345.
- Wang J, Street NR, Park E-J, Liu J, Ingvarsson PK. 2020. Evidence for widespread selection in shaping the genomic landscape during speciation of *Populus*. *Mol Ecol*. **29**:1120–1136.
- Wang K, Lenstra JA, Liu L, Hu Q, Ma T, Qiu Q, Liu J. 2018. Incomplete lineage sorting rather than hybridization explains the inconsistent phylogeny of the wisent. *Commun Biol*. **1**:169.
- Wang N, Chang M. 2012. Discovery of fossil Nemacheilids (Cypriniformes, Teleostei, Pisces) from the Tibetan Plateau, China. *Science China-Earth Sciences* **55**:714–727.
- Wang T, Zhang Y-P, Yang Z-Y, Liu Z, Du Y-Y. 2020. DNA barcoding reveals cryptic diversity in the underestimated genus *Triplophysa* (Cypriniformes: Cobitidae, Nemacheilinae) from the northeastern Qinghai–Tibet Plateau. *BMC Evol Biol*. **20**:151.
- Wang Y, Shen Y, Feng C, Zhao K, Song Z, Zhang Y, Yang L, He S. 2016. Mitogenomic perspectives on the origin of Tibetan loaches and their adaptation to high altitude. *Sci Rep*. **6**:29690.
- Wang Y, Yang L, Zhou K, Zhang Y, Song Z, He S. 2015. Evidence for adaptation to the Tibetan Plateau Inferred from Tibetan loach transcriptomes. *Genome Biol Evol*. **7**:2970–2982.
- Westbury MV, Hartmann S, Barlow A, Wiesel I, Leo V, Welch R, Parker DM, Sicks F, Ludwig A, Dalén L, et al. 2018. Extended and continuous decline in effective population size results in low genomic diversity in the world's rarest hyena species, the brown hyena. *Mol Biol Evol*. **35**:1225–1237.
- Whiteley AR, Fitzpatrick SW, Funk WC, Tallmon DA. 2015. Genetic rescue to the rescue. *Trends Ecol Evol (Amst)*. **30**:42–49.
- Xue Y, Prado-Martinez J, Sudmant PH, Narasimhan V, Ayub Q, Szpak M, Frandsen P, Chen Y, Yngvadottir B, Cooper DN, et al. 2015. Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. *Science* **348**:242–245.
- Yang J, Lee SH, Goddard ME, Visscher PM. 2011. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet*. **88**:76–82.
- Yang L, Wang Y, Wang T, Duan S, Dong Y, Zhang Y, He S. 2019. A chromosome-scale reference assembly of a Tibetan loach, *Triplophysa siluroides*. *Front Genet*. **10**:991.
- Yang W, Feiner N, Pinho C, While GM, Kaliontzopoulou A, Harris DJ, Salvi D, Uller T. 2021. Extensive introgression and mosaic genomes of Mediterranean endemic lizards. *Nat Commun*. **12**:2762.
- Yang X, Liu H, Ma Z, Zou Y, Zou M, Mao Y, Li X, Wang H, Chen T, Wang W, et al. 2019. Chromosome-level genome assembly of *Triplophysa tibetana*, a fish adapted to the harsh high-altitude environment of the Tibetan Plateau. *Mol Ecol Resour*. **19**:1027–1036.
- Yang ZH. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. **24**:1586–1591.
- Yu Y, Wang X, Li Z, Li Y, Lu H. 2022. Late quaternary drainage evolution and its response to fault activity and climatic change in the source region of the Yangtze River. *Quaternary Sciences* **19**:880–898.
- Yuan D, Chen X, Gu H, Zou M, Zou Y, Fang J, Tao W, Dai X, Xiao S, Wang Z. 2020. Chromosomal genome of *Triplophysa bleekeri* provides insights into its evolution and environmental adaptation. *Gigascience* **9**(11):giaa132.
- Zhang D, Rheindt FE, She H, Cheng Y, Song G, Jia C, Qu Y, Alström P, Lei F. 2021. Most genomic loci misrepresent the phylogeny of an avian radiation because of ancient gene flow. *Syst Biol*. **70**:961–975.
- Zhang T, Sun H. 2011. Phylogeographic structure of *Terminalia franchetii* (Combretaceae) in southwest China and its implications for drainage geological history. *J Plant Res*. **124**:63–73.
- Zhisheng A, Kutzbach JE, Prell WL, Porter SC. 2001. Evolution of Asian monsoons and phased uplift of the Himalaya–Tibetan Plateau since Late Miocene times. *Nature* **411**:62–66.
- Zhou C, Hu B, Tang Y, Yang C, Ma W, Xi W, Liu R, Yan X, Dong J, Xianfeng W, et al. 2021. The chromosome-level genome of *Triplophysa dalaica* (Cypriniformes: Cobitidae) provides insights into its survival in extremely alkaline environment. *Genome Biol Evol*. **13**:evab153.
- Zhou Y, Zhao X, Li Y, Xu J, Bi A, Kang L, Xu D, Chen H, Wang Y, Wang YG, et al. 2020. Triticum population sequencing provides insights into wheat adaptation. *Nat Genet*. **52**:1412–1422.